

2 Le Web

Le *World Wide Web*, ou plus simplement le Web, 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. Ces données, textuelles notamment, sont codés dans un langage, le **HTML**, et sont transmises sur le réseau via le **protocole HTTP**.

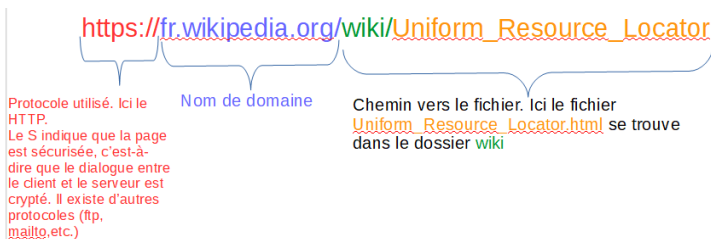
1 Accéder à un contenu ou l'histoire d'un dialogue entre un client et un serveur.

L'accès à un contenu, que ce soit une page Web, une vidéo ou autre, se fait à travers un navigateur, un logiciel qui s'exécute sur votre ordinateur : **le client**.

Les pages Web sont des fichiers qui sont stockés sur un ordinateur distant, qu'on appelle **le serveur**.

Pour accéder à ces fichiers, on indique l'adresse du serveur où elles se trouvent en saisissant une **adresse URL (Uniform Resource Locator)** dans le navigateur.

L'**URL** est composé de plusieurs parties :



Lors d'une recherche sur un moteur de recherche ou de la saisie d'une URL, le client (l'ordinateur) envoie une requête HTTP au serveur qui va lui répondre. Celui-ci renvoie un fichier nommé **index.html** qui correspond au site que l'on souhaite visiter.

Repères historiques :

- 1989/90 : Tim Berners-Lee avec Roger Caillau pose les bases du Web : l'url, le HTML et le protocole HTTP
- 1990 : Premier navigateur Web : WorldWideWeb
- 1993 : Mise dans le domaine public des technologies du Web.
- 1995 : Création de Yahoo et d'Internet Explorer.
- 1998 : Naissance de Google.
- 2004 : Lancement de Facebook.
- 2005 : Lancement de Youtube, avènement du Web 2.0.

2 Les fondements du Web : l'hypertexte, l'URL et le HTML

Les premières pages Web ne contenaient que du texte mais s'appuyaient déjà sur ce qui fait le fondement d'une page, l'**hypertexte** : la possibilité d'accéder à d'autres pages à l'aide de lien. Aujourd'hui, les pages Web sont un assemblage de contenus, de textes, d'images ou de vidéos. L'affichage et la structure de ces contenus est régie par le langage HTML, que seul un navigateur peut interpréter.

Le **HTML (HyperText Markup Language)** est un langage à **balises**, qui structure une page de manière brute. Il est accompagné d'une feuille de style **CSS (Cascading Style Sheet)** qui permet la mise en forme (couleur, police).

3 Les moteurs de recherche

Le Web contient plus de 100 milliards de pages. Pour s'y retrouver et accéder à la page souhaitée, on fait appel à un **moteur de recherche (Google, Bing, Qwant)**. Son rôle est d'indexer les sites, page par page, selon certains critères, le plus simple d'entre eux étant les **mots-clés**.

Toutes les pages ne sont pas indexées par les moteurs de recherche. En effet, pour indexer ces pages, un robot, un ordinateur avec un algorithme dédié, parcourt l'ensemble des sites en suivant les liens présents dans celles-ci. Si aucun lien ne pointe vers une page ou si un propriétaire de site ne souhaite pas son référencement, la page n'est pas indexée (référéncée dans le moteur de recherche). L'ensemble de ces sites non indexés constitue le **Web profond**. Pour y accéder, il faut en connaître l'adresse ou alors utiliser certains moteurs de recherche spécialisés.

En plus de l'indexation, les moteurs de recherche utilisent l'architecture du Web pour extraire les pages les plus intéressantes, celles que tout le monde finira par voir parce que les liens mènent à elles. On schématise le Web sous forme d'une toile avec les liens entre les sites. Plus un site a de lien pointant vers lui, mieux il est référencé sur le moteur de recherche.

4 Paramètres de sécurité d'un navigateur

Les sites Internet utilisent des petits morceaux de codes, les cookies, afin de recueillir diverses informations sur les visiteurs de leur site (données personnelles, etc.) Certains logiciels (*CookieViz* ou l'extension *LightBeam* sous *Firefox*) permettent de visualiser ces traces.

u Pour sécuriser un navigateur, on peut effectuer quelques actions :

- a. **Interdire l'exécution du Javascript sur le navigateur.** avec l'extension *NoScript*.
- b. Dans les options Internet, **ne pas conserver l'historique et bloquer les cookies.**
- c. **Mettre à jour** très régulièrement les plugins Java et Flash ;
- d. **Installer sur l'ordinateur des extensions** comme *Ghostery* (bloque les traceurs, les publicités), *DoNotTrackMe* (bloque les traceurs) et *HTTPSeverywhere* (force la connexion sécurisée pour tous les sites).

Lexique :

Navigateur : logiciel permettant d'afficher à l'écran le site visité (Microsoft Edge, Firefox, Chrome, ...)

Code : ensemble d'instructions constituant un programme.

Javascript : langage de programmation spécifique au Web qui permet de coder des effets particuliers sur une page.

plugins : module installé sur un navigateur pour apporter une fonctionnalité supplémentaire.

5 Impact sur les pratiques humaines

Le web est devenu omniprésent aujourd'hui tant pour s'informer que pour effectuer des démarches en ligne. Cette pratique amène à s'interroger sur la maîtrise des données personnelles (autorisation des sites internet à conserver des données telles que l'âge, le nombre d'enfants, etc.) le droit à l'oubli (possibilité de supprimer définitivement des contenus publiés, par exemple dans les réseaux sociaux) ainsi que sur la véracité des informations lues (tout le monde peut publier du contenu, mais personne n'en vérifiant la nature, ils peuvent être totalement faux).

L'accès à la connaissance a été profondément bouleversé, tout un chacun peut accéder instantanément à une quantité importante d'informations. Pour autant, ce n'est pas parce qu'un contenu est accessible qu'il n'est pas soumis au droit d'auteur. Ce qui signifie que son utilisation est protégée par le droit d'auteur, le droits d'usage ou par certaines licences comme les Creative Commons qui permettent de le redistribuer librement.