

Sciences du Numérique et Technologie



Le Web

TD n° 1 Les moteurs de recherche



Objectif de la séance :

- Connaître les étapes du développement du Web.
- Décomposer l'URL d'une page
- Comprendre les enjeux de la publication d'informations.



Mots-clés :

Client : logiciel envoyant des demandes à un serveur.

CSS (Cascading Style Sheets) : feuilles de style décrivant la présentation des documents HTML.

HTML (HyperText Markup Language) : langage de balisage utilisé pour la création de pages Web.

HTTP ou HTTPS : protocole de transmission permettant d'accéder à des pages Web via un navigateur. Le HTTPS est sécurisé.

Indexation : analyse du contenu des pages Web pour détecter les mots-clés.

Lien hypertexte : élément d'une page Web (texte, image, etc.) qui, lorsque l'on clique dessus, renvoie vers une autre zone de la même page, une autre page, ou un autre site Web.

Moteur de recherche : application informatique permettant de rechercher une ressource (pages Web, images, vidéos, fichiers, etc.) à partir d'une requête sous forme de mots.

Navigateur : logiciel permettant d'afficher des sites Internet (Chrome, Firefox, Edge, Safari, etc.).

Serveur : ordinateur exécutant des opérations suivant les requêtes effectuées par un ordinateur appelé « client ».

URL (Uniform Resource Locator) : adresse d'une page d'un site.

Web (World Wide Web) : système hypertexte, utilisant le protocole HTTP, permettant de visiter des pages sur Internet.



A la fin de cette séance j'aurai acquis ...

Des connaissances	Des compétences
- URL - Web et noms de domaines - Moteur de recherche - HTTP	- Comparer les paramétrages de différents navigateurs. - Paramétrer un navigateur de manière qu'il interdise l'exécution d'un programme sur le client. - Régler les paramètres de confidentialité dans son navigateur ou dans un service en ligne.

Vous écrirez les réponses aux questions sur le document

Présentation qu'est-ce que le Web ?



Le Web (ou Toile) est un ensemble d'informations numériques reliées et accessibles par des liens cliquables, sur internet. À partir des années 2000, le Web « social » avec l'apparition de sites sociaux comme Facebook, Twitter, YouTube ou Flickr, permettant le partage en tout genre.

Aujourd'hui le Web continue de se développer en devenant immersif et en interagissant avec le réel. Ainsi, les accès au Web se diversifient et se joignent à tout type de terminal (lecteur mp3, tablette, smartphone, montre connectée, etc.). Ces données collectées alimentent des bases de données dont l'exploitation et le traitement présentent un enjeu économique majeur.

Le Web démocratise les voix de l'information en se calquant sur le phénomène de mondialisation. Il gomme les frontières, mais reste tout de même façonné selon les lois et particularités des pays. Ceux qui créent le Web, le font à leur image.

1. Le Web : Une toile mondiale

Originaire du mot anglais « Web », qui se traduit littéralement par « toile », **le concept du Web** a pris naissance dans un cadre technologique et numérique spécifique. Sir Tim Berners-Lee, un scientifique britannique, a inventé le **World Wide Web (WWW) en 1989**. Cette appellation, « World Wide Web », exprime l'idée d'une toile qui s'étend à l'échelle mondiale, englobant la planète entière.

Ainsi, le Web est, en réalité, une immense constellation de documents et d'autres ressources web, reliés entre eux par des hyperliens. Ces éléments sont interconnectés et accessibles au public via le réseau Internet. En somme, **le Web est comparable à un vaste océan d'informations** dans lequel nous pouvons plonger à tout moment pour explorer, apprendre et interagir.

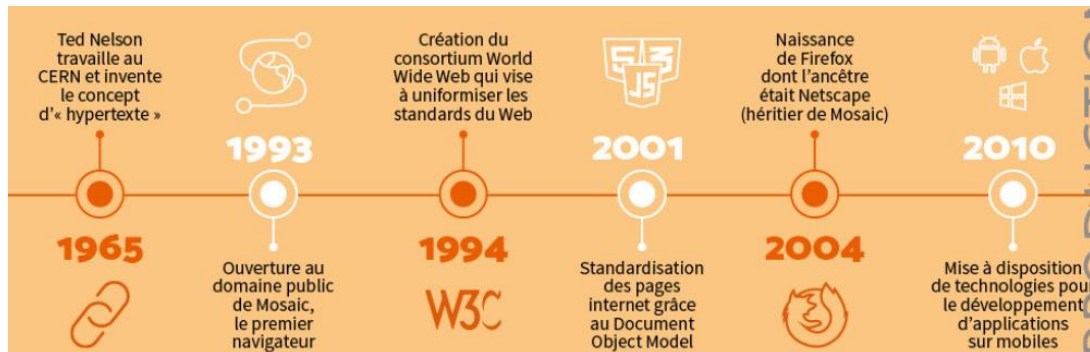
2. Les composants du Web

Le World Wide Web, couramment appelé « Web », est une merveille technologique complexe, façonnée par divers éléments interdépendants. Ces composants, bien que souvent invisibles à l'utilisateur moyen, sont les piliers qui soutiennent l'architecture du Web. Plongeons dans une exploration détaillée de ces constituants essentiels :

- **URL (Uniform Resource Locators) :** Les coordonnées numériques Au cœur de chaque exploration sur le Web se trouve une URL. Comparable à une adresse postale dans le monde réel, une URL guide votre navigateur vers la destination souhaitée sur le Web. Chaque fois que vous tapez quelque chose comme « www.questceque.fr » dans la barre d'adresse de votre navigateur, vous utilisez en réalité une URL pour diriger votre demande vers un serveur spécifique qui héberge ce contenu ;
- **Navigateurs :** Les fenêtres sur le monde numérique Les navigateurs, comme Google Chrome, Mozilla Firefox ou Safari, sont des outils inestimables qui transforment les codes complexes en informations visuellement digestes pour les utilisateurs. Ces logiciels décodent les données des serveurs web et les présentent sous forme de pages web interactives, avec des textes, des images, des vidéos et d'autres médias ;
- **Sites Web :** Les destinations de notre voyage numérique Chaque site web est une collection organisée de pages web qui fonctionnent de manière synchrone pour offrir une expérience utilisateur unique. Ces pages sont stockées sur des serveurs, qui sont des ordinateurs puissants conçus pour traiter et répondre

aux demandes des utilisateurs du monde entier. Ainsi, lorsque vous visitez un site web, votre navigateur demande en fait à un serveur d'afficher le contenu associé à cette URL ;

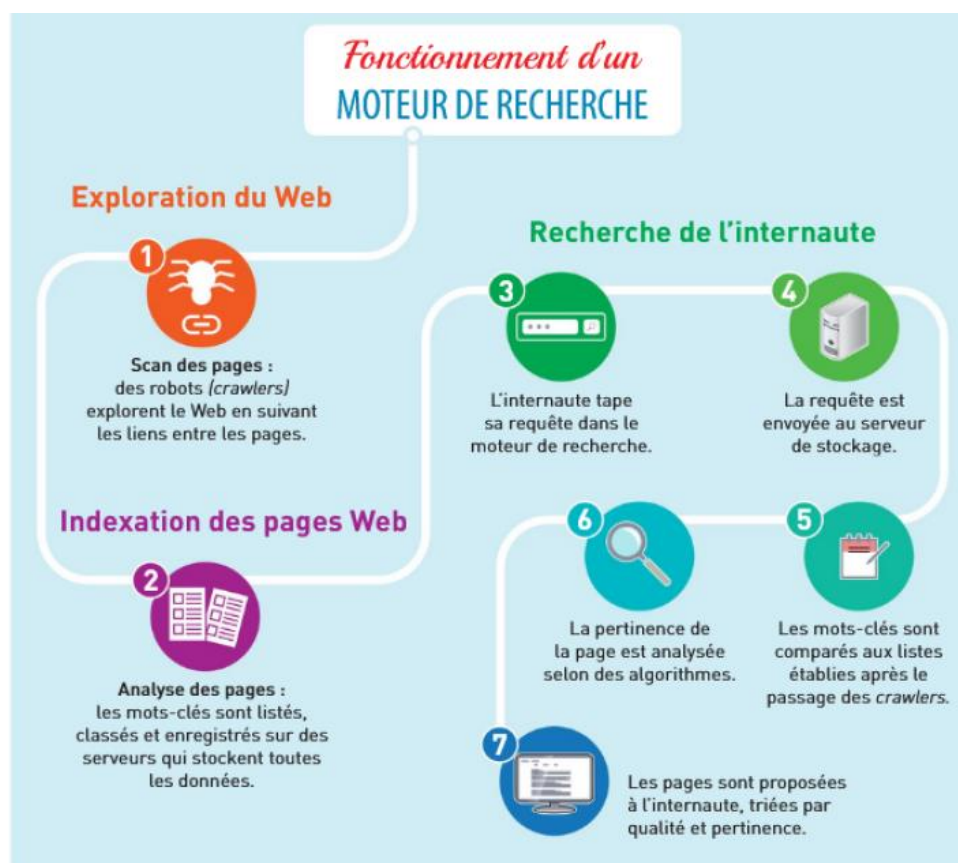
- **Hyperliens** : Les portails vers de nouvelles découvertes Les hyperliens sont peut-être l'un des éléments les plus révolutionnaires du Web. Ces liens, généralement soulignés ou colorés différemment, permettent aux utilisateurs de sauter d'une page ou d'un sujet à un autre avec un simple clic. Ils tissent une immense toile de contenu interconnecté, d'où le terme « toile » pour décrire le Web. Grâce aux hyperliens, le Web est devenu un espace dynamique, interconnecté et en constante évolution.



En combinant tous ces éléments, le Web offre une plateforme d'information et d'interaction sans précédent, redéfinissant sans cesse notre façon de communiquer, d'apprendre et de partager à l'ère numérique.

I. Fonctionnement d'un moteur de recherche :

Les **moteurs de recherche** fonctionnent en trois temps. D'abord, l'exploration consiste à faire parcourir le Web par des robots pour collecter des informations sur les pages Web. Puis, c'est l'indexation : les données sont analysées et classées dans des bases de données afin de permettre leur exploitation. Enfin, un internaute effectue une requête. Un **algorithme** est appliqué pour identifier dans l'index les sites qui correspondent le mieux aux mots-clés de la recherche.



Exercice 1 :

Question 1 : Documents 1

Résumer le fonctionnement d'un moteur de recherche.

Document 2 : Popularité d'une page web :



Calcul de la popularité

La principale difficulté des moteurs de recherche consiste à classer, de la manière la plus pertinente possible l'ensemble des pages contenant les mots

clés demandés, pour choisir quelles pages présenter en premier en réponse à une requête.

PAGE RANK DEFINITION

Le PageRank est un **indice de classement déterminé par un algorithme** de Google, qui permet à ce moteur de recherche de déterminer la popularité d'une page web en vue de la positionner dans ses résultats.

©SEWIP

C'est essentiellement sur ces algorithmes de classement que les moteurs de recherche se font une concurrence acharnée et cachent leurs secrets de fonctionnement. **En 1998**, les informaticiens américains **Larry Page et Sergey Brin** proposent l'algorithme PageRank qui a conduit à la création du moteur de recherche Google et permis de proposer une définition à l'idée de « popularité » d'une page.

L'algorithme PageRank, (de l'anglais Rank : score) repose sur le principe de calculer la popularité d'une page à partir de la popularité des pages qui la citent. Cela évite de chercher à comprendre ce qui est écrit dans une page, ce qui serait difficile pour une machine. L'idée est de faire confiance aux auteurs des pages web : en citant une autre page, ils lui attribuent de la confiance. Cette confiance a de la valeur si elle émane d'un site lui-même très populaire.

Ce calcul est en fait très complexe car il oblige à calculer en même temps la popularité de toutes les pages du web qui dépendent les unes des autres, ce qui aboutit à une gigantesque équation avec un nombre d'inconnues égal au nombre de pages du web !

Il ne faut pas confondre popularité et pertinence d'une page web.

Question 2 : Documents 2

A l'aide des documents et des vidéos sur le site, répondre aux questions suivantes :

1. Qu'est-ce que la pertinence d'une page web ?
2. Qu'est-ce que la popularité d'une page web ?
3. Quel est l'intérêt principal d'indexer les pages web suivant leur popularité ?
4. Quel « outil » les moteurs de recherche utilisent-ils pour calculer la popularité d'une page web ?
5. Citer un algorithme bien connu permettant de calculer la popularité d'une page web. Quels sont ses créateurs ? Quand a-t-il été créé ? Quel moteur de recherche l'utilise ?

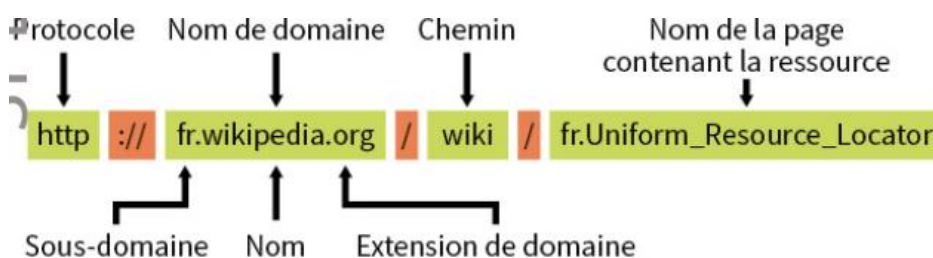
II. Comment accéder et naviguer sur le Web

Document 1 : Le protocole de communication HTTP

Pour s'afficher dans le navigateur, une page Web doit avoir fait l'objet d'une recherche de la part de l'utilisateur. Cette recherche devient une **requête** : le navigateur demande au serveur, qui héberge ce document, de le lui envoyer. C'est cette relation client-serveur que désigne le **protocole HTTP** (*HyperText Transfer Protocol*, « protocole de transfert hypertexte »).

Pour que cela soit possible, il faut que le document en question puisse être localisé sur le World Wide Web par une adresse unique. Celle-ci indique son emplacement sur un ordinateur serveur, adresse à partir de laquelle le navigateur va pouvoir envoyer sa requête. On appelle cette adresse une **URL** ou « *Uniform Resource Locator* », expression anglaise pouvant se traduire par « localisateur universel de ressource ».

Exemple d'URL



Document 2 : Les Extensions de domaines

Les **extensions et noms de domaine** de premier niveau se divisent en deux catégories : celles associées à un pays ou un territoire (« .fr » ou « .uk », par exemple) et les celles associées à des activités (« .edu » pour l'éducation, « .com » pour le commerce, par exemple).

	Extension de noms de domaine	Catégorie	Nombre de noms de domaine en janv. 2019	Évolution depuis 01/18 en %
1 ^{er}	.com	Activité – Sites « commerciaux »	137 781 544	5,34
2 ^e	.cn	Territoire - Chine	22 741 258	6,27
3 ^e	.tk	Territoire - Tokelau	21 500 000	12,57
4 ^e	.de	Territoire - Allemagne	16 204 745	- 0,67
5 ^e	.net	Activité - sites « network »	13 679 758	- 8,80
6 ^e	.co.uk et .uk	Territoire - Royaume-Uni	11 999 151	- 0,37
7 ^e	.org	Activité - sites « organisation »	10 172 458	- 1,24
8 ^e	.nl	Territoire - Pays-Bas	5 843 964	0,79
9 ^e	.ru	Territoire - Russie	5 788 165	8,10
10 ^e	.info	Activité - sites d'« information »	4 724 406	- 21,26

Le top 10 des extensions de domaine, classement 2019

Source : Solidnames.

Document 3 : Les noms de domaines

Sur le Web, un utilisateur ou une entreprise est libre de choisir le nom de domaine associé aux pages Web qu'il crée, sous réserve qu'il soit disponible. Ce choix est stratégique sur le plan de la communication car il est l'une des portes d'entrée de l'identité numérique.

Les noms de domaines ne peuvent pas être achetés à vie, mais loués sur une durée déterminée. Pour en louer un, les utilisateurs doivent se rapprocher de distributeurs agréés par l'ICANN (*Internet Corporation for Assigned Names and Numbers*), l'autorité de régulation d'internet.

Pour qu'une adresse soit accessible *via* son nom de domaine, ce dernier doit être enregistré dans deux DNS (Domain Name System) au minimum. Le DNS fait le lien entre le Web et internet.

Exercice 2 :

A l'aide des documents et des vidéos sur le site, répondre aux questions suivantes :

1. Que veut dire URL et quels sont les quatre principaux éléments ?
2. A quoi sert l'extension de domaine, il en existe deux types lesquels, qui les administre et quelle est l'extension de domaine la plus utilisée ?
3. Qui choisit les noms de domaines ? auprès de quelle société ?
4. **Conclusion :** Pour quelles raisons un concepteur de site Web doit-il tenir compte des éléments décrits dans tous les documents ?