

**Informatique et création numérique (ICN) & Education aux médias et à l'information (EMI)
COMPRENDRE UN MOTEUR DE RECHERCHE**

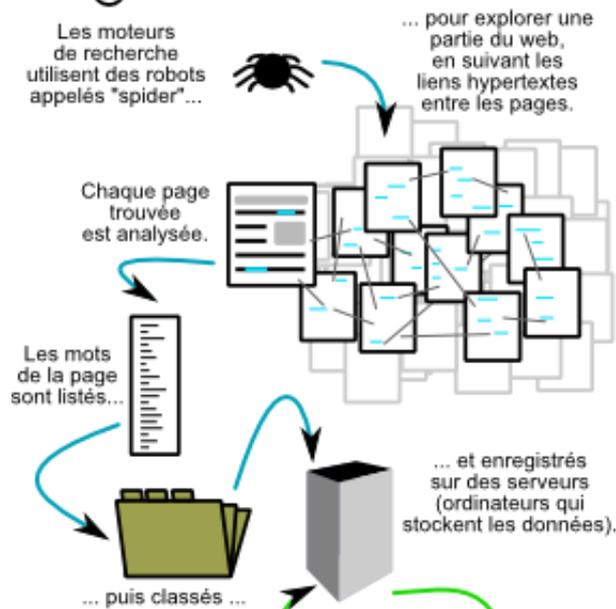
SYNTHESE de CONNAISSANCES

Activité 1 : Fonctionnement d'un moteur de recherche

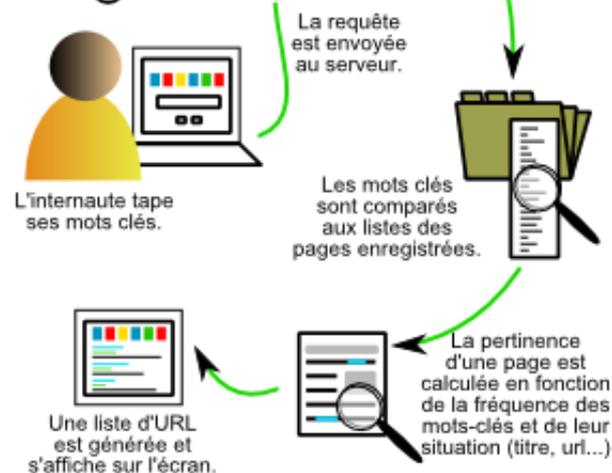
COMMENT FONCTIONNE UN MOTEUR DE RECHERCHE ?

Un moteur de recherche est un type de site Web. Il permet à l'internaute de rechercher des informations et des adresses de sites sur le Web.

PHASE ① : INDEXATION DU WEB



PHASE ② : RECHERCHE DES DONNEES



Par : fenetresur.wordpress.com - CC-BY-NC-ND

Mots-clés : Crawler ou spider / indexation / liens hypertextes / requête / serveur / mots clés / base de données / interrogateur / algorithme / pertinence, classement, positionnement des pages web / page rank /

<https://www.youtube.com/watch?v=iKMm6SXO0wA>

Activité 2 : Explorer différents moteurs de recherche

Quel autre moteur de recherche connaissez-vous ?

Yahoo – Bing – Qwant – Duckduckgo

Les autres : Yippy, Xaphir, Quora, Lilo, Startpage, Ecosia

Baidu (le Google chinois) / Yandex (le Google russe)

https://fr.wikipedia.org/wiki/Liste_de_moteurs_de_recherche



Ne pas confondre navigateur et moteur de recherche

En bref...

Il faut utiliser l'un pour accéder à l'autre

- **Navigateur** = fenêtre permettant de regarder les pages internet
- **Moteur de recherche** = application permettant de trouver le contenu recherché

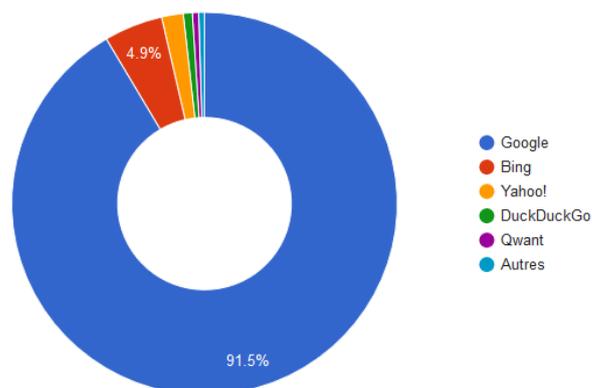
Exemples de navigateurs:

- Internet Explorer 
- Google Chrome 
- Mozilla Firefox 
- Safari 

Tester et comparer les moteurs de recherche

Voici les parts de marché des moteurs de recherche en FRANCE :

- Google : 91,49%
- Bing : 4,93%
- Yahoo! : 1,81%
- DuckDuckGo : 0,76%
- Qwant : 0,52%



Mai 2018

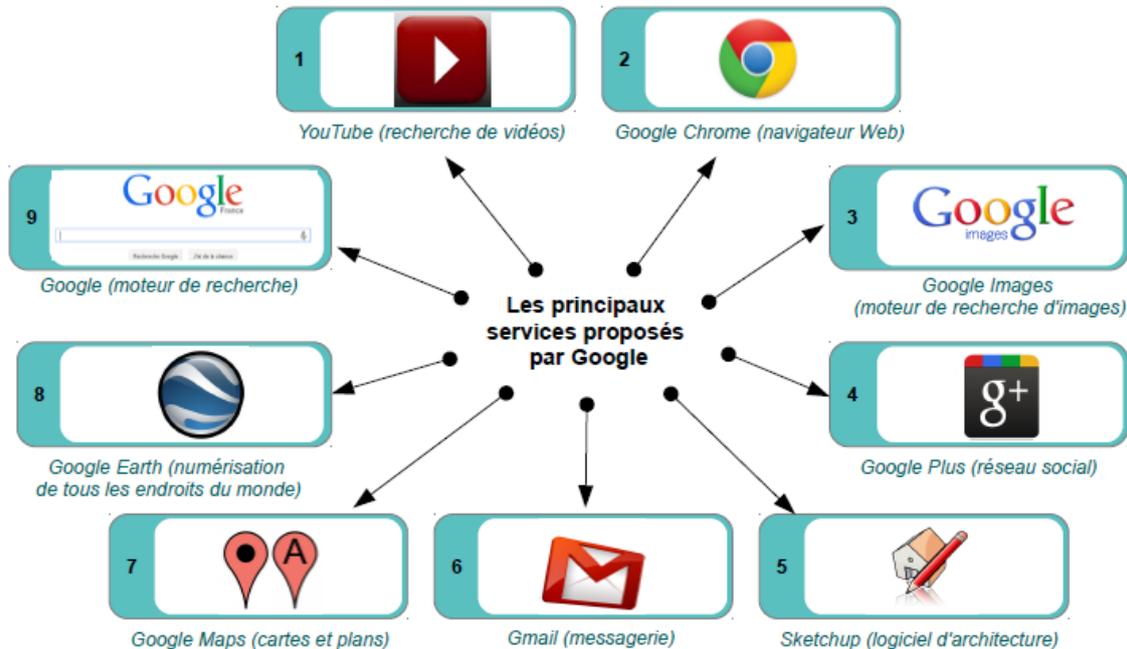
	Date de création et nationalité	Popularité	Protection de la vie privée	Fonctionnalités et outils connectés
	1998 Américain	91,49%	<p>La multiplication des services proposés par Google engrange une demande accrue de renseignements sur les utilisateurs : suivi de la navigation et stockage des mots-clés, scan des mails dans Gmail, des informations livrées dans les formulaires, entre autres.</p> <p>Google croise ces données pour affiner le profil des utilisateurs, et améliorer le ciblage des publicités sur internet.</p> <p>Une telle concentration d'informations sur les individus et leur conservation inquiètent les organisations de défense de la vie privée sur internet.</p>	<p>Capacité à trouver la réponse rapidement et faculté qu'il a à comprendre nos besoins et nos attentes. Le plus efficace, le plus fonctionnel et le plus complet.</p> <p>220 produits, services, brevets ...</p>
	2013 Français	0,52%	<p>Qwant garantit que les informations personnelles des utilisateurs ne sont pas stockées et utilisées à des fins commerciales. Aucun cookie ne trace l'internaute. Les résultats sont très orientés par l'actualité et les tendances.</p>	<p>Page de résultats très riche.</p> <p>Qwant junior.</p> <p>La SERP (page de résultat) affiche un tableau de bord.</p>

Tout savoir sur Qwant
<http://emi.re/qwant.html>

Tout savoir sur Google
<http://emi.re/google.html>

Activité 4 : Nos vies en Google

➤ Nomme les principaux services de Google



Les résultats de requêtes formulées dans Google

➤ Que constates-tu au sujet des résultats de chaque ligne du tableau ?

- ❶ Par défaut, Google affiche 10 résultats par page sauf si des annonces publicitaires (sites commerciaux) ou des sites appartenant à Google y figurent.
- ❷ Wikipédia est le site le plus visité, c'est pourquoi il se retrouve très souvent en haut de la première page de résultats (notion de PageRank).
- ❸ et ❹ Selon l'actualité, on peut trouver les sites de Google (G.Actualités, G.Images, G.Vidéos, YouTube...) en première page de résultats.

➤ Que peux-tu en conclure sur le fonctionnement de Google ?

Google se comporte comme toute entreprise en développant ses services, son influence, de manière optimale. Pour cela, elle cherche à se rendre incontournable lors de chacune de nos actions sur le Web.

PAGERANK : principal critère de classement de Google

Source : Wikipédia

« Le principe de base est d'attribuer à chaque page une valeur (ou score) proportionnelle au nombre de fois que passerait par cette page un utilisateur parcourant le graphe du Web en cliquant aléatoirement, sur un des liens apparaissant sur chaque page. Ainsi, une page a un PageRank d'autant plus important qu'est grande la somme des PageRanks des pages qui pointent vers elle (elle comprise, s'il y a des liens internes). »

Que pouvez-vous en conclure sur le fonctionnement de Google ?

Comment Google choisit-il l'ordre de présentation des pages ?

On sait que les gens ne consultent généralement que la première page de résultats, et même souvent seulement les deux ou trois premiers sites de cette page.

L'ordre dans lequel les sites apparaissent est donc très important. Google décide de l'ordre en faisant des calculs mathématiques selon des formules que l'on appelle **algorithme**.

On sait que l'algorithme prend en compte la place du mot dans les sites (ils sont plus importants si ils sont dans le titre du site ou de la page par exemple) ou le nombre de fois qu'on retrouve le mot dans la page. Mais d'autres critères compliqués et secret sont pris en compte.

A partir de nos expériences, on peut déduire quelques faits :

- Nous n'obtenons pas tous les mêmes résultats. **C'est parce que Google personnalise ses résultats** en fonction notamment des recherches précédentes. Par exemple si on fait une recherche météo, il y a de fortes chances qu'il nous propose directement la réponse pour notre ville.
- Wikipedia arrive tout le temps en tête. **C'est parce que Google privilégie les sites les plus populaires (les plus consultés, ceux qui sont souvent cités, etc.)** : ce sont ceux qui auront le plus de chance de vous intéresser.
- On trouve très souvent Google images et Youtube en première page. **C'est parce que Google privilégie ses propres services**, c'est de la publicité personnelle en quelque sorte.
- On trouve beaucoup de sites d'entreprises ; on ne sait pas trop comment Google fait pour privilégier certaines entreprises par rapport à d'autres.

