

Les moteurs de recherche

Sources:

- <https://www.lelivrescolaire.fr/page/30329052>
- https://cache.media.eduscol.education.fr/file/SNT/20/4/RA19_Lycees_G_SNT_2nd_Page_rank_1156204.pdf

Contenus:

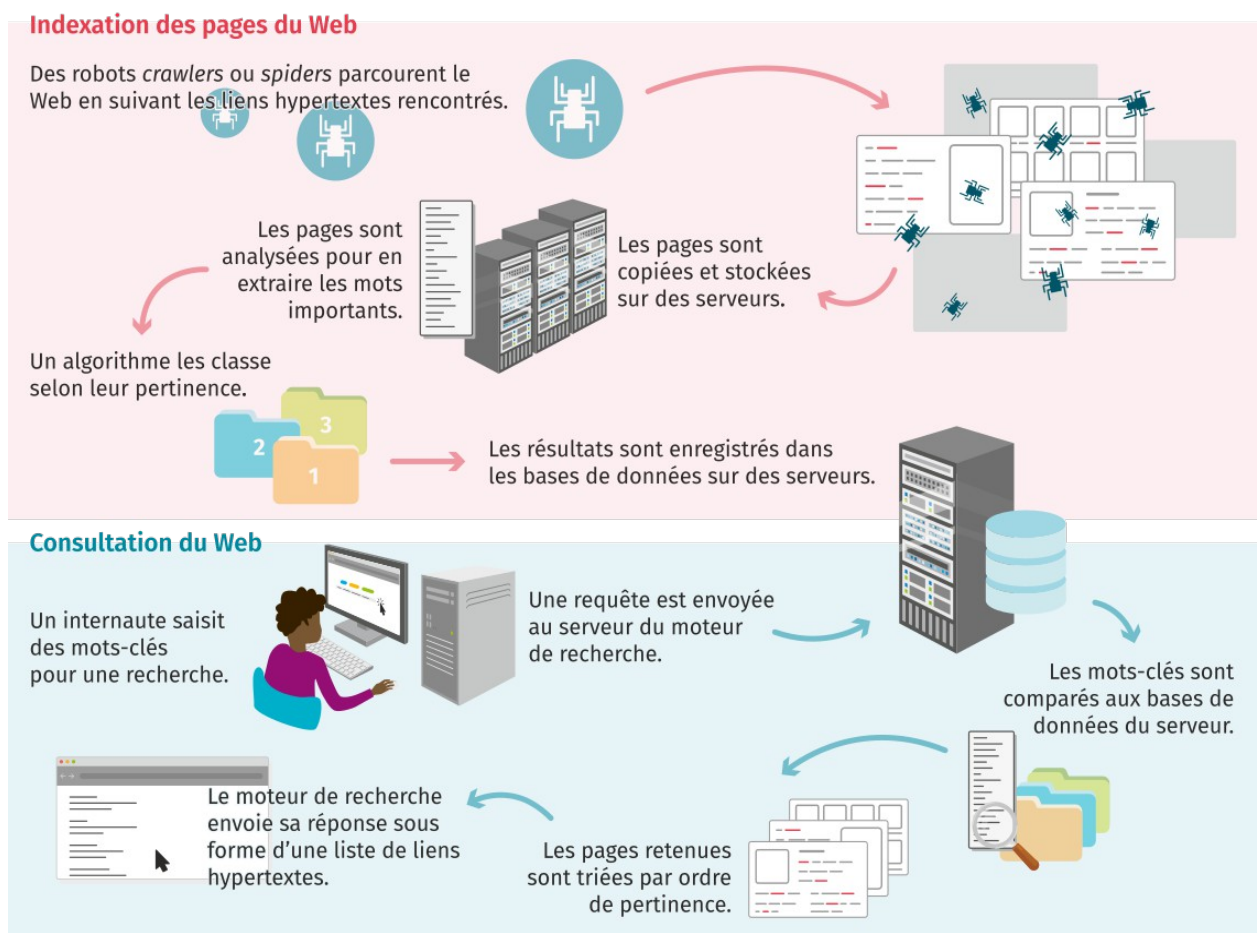
Les moteurs de recherche, principes et usages

Capacités attendues:

Mener une analyse critique des résultats fournis par un moteur de recherche.
Comprendre les enjeux de la publication d'informations.

En 2020, le site Web le plus visité au monde était celui du moteur de recherche Google. Ce type d'outil permet d'explorer le Web et de trouver des ressources numériques à l'aide de mots-clés.

1 – Fonctionnement d'un moteur de recherche



2 – Référencement naturel SEO (Search Engine Optimization)

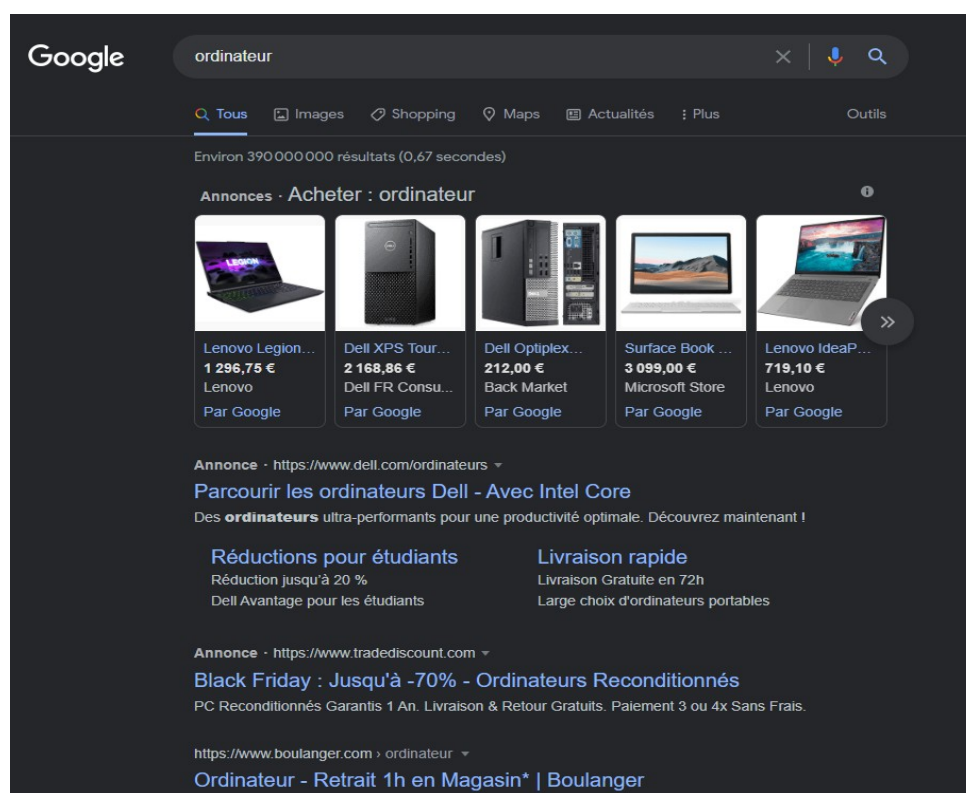
Pour apparaître dans les 3 premières suggestions d'un moteur de recherche, plusieurs conditions doivent être réunies parmi les suivantes:

- Permettre aux robots des moteurs de recherche d'atteindre les ressources sur le site sans le bloquer
- Maximiser les liens entrant sur la page en provenance de sites fréquentés et pertinents
- Optimiser le code de la page Web, dont les métadonnées, pour faciliter le travail d'indexation

Les algorithmes des moteurs de recherche sont complexes et en constante évolution. Certains paramètres tels que la localisation de l'internaute ou son historique de navigation peuvent avoir une influence sur les réponses du moteur de recherche.

→ [Fiche méthode](#) pour comparer les principaux moteurs de recherche.

3 – Réponses à une recherche sur google



Découvrir le système de "ranking"

Le système de "ranking" fait la force des moteurs de recherche. Notons que les algorithmes exacts ne sont jamais connus car ils sont considérés comme un secret industriel. Il existe au moins 3 méthodes principales pour classer les pages:

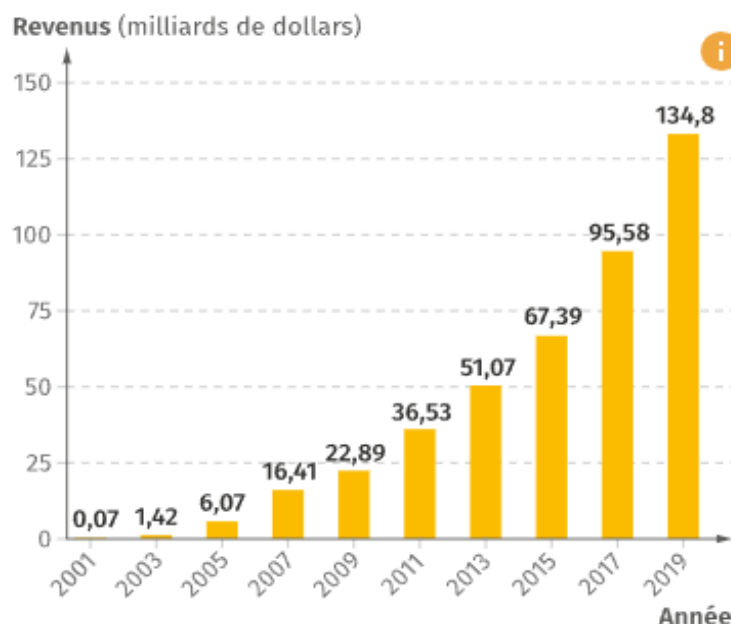
- Le **tri par pertinence**: évalue entre autres le nombre de fois qu'un mot ou un document apparaît dans une page web, ou encore, la correspondance entre l'expression cherchée

et le contenu du document.

- Le **tri par popularité**: comme le *PageRank* de Google, qui note la popularité d'une page web. Plus une page possède de *backlinks* (c'est-à-dire de liens qui pointent vers elle), plus la page est populaire.
- Le **tri par rapport à l'audience**: évalue le temps resté par l'internaute sur une pageweb, synonyme de son intérêt. Des études ont montré que 90% des internautes ne consultent que le premier résultat d'une recherche

4 – Revenus des moteurs de recherche

Les mots-clés peuvent être achetés aux enchères auprès d'un moteur de recherche afin de s'assurer d'apparaître en haut de la liste des propositions faites par le moteur de recherche. Il s'agit du référencement payant ou SEA (*Search Engine Advertising*). Certains bloqueurs de publicités permettent néanmoins de masquer ces annonces.



En 2019, les revenus publicitaires de Google représentaient près de 85 % des revenus de l'entreprise.

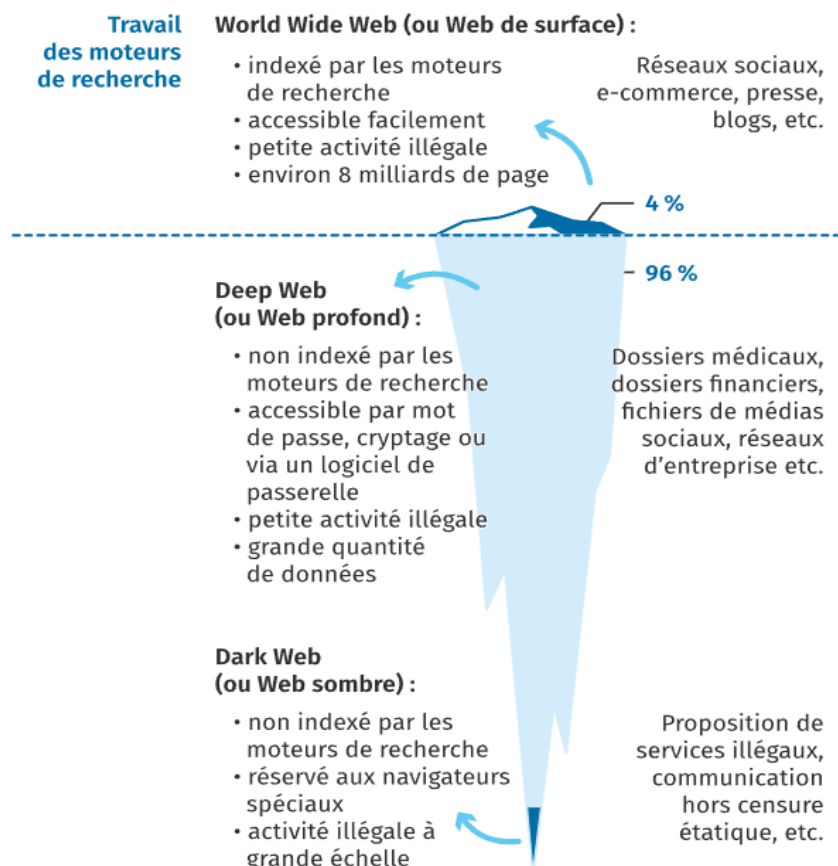
Activité - Différence de réponses

- 1 – Ouvrir un navigateur et se rendre sur le moteur de recherche Google
- 2 – Effectuer une recherche en tapant un mot-clé, par exemple: ordinateur
- 3 – Ouvrir un nouvel onglet et répéter les étapes précédentes avec les moteurs de recherche cité dans "5 - Le web - activité n°1", pour le même mot-clé
- 4 – Comparer les 5 premiers sites renvoyés par les moteurs de recherche utilisés

5 – Les différentes couches du Web

Numérique

Découvrez les différentes couches du Web.



Complément:

Note d'intention:

Le moteur de recherche, véritable révolution, a bouleversé de nombreux métiers. Cette activité propose de comprendre un des algorithmes qui peut se cacher derrière. Nous nous intéresserons ici au moteur de recherche du PageRank de GOOGLE.

Comment un moteur de recherche classe-t-il les sites ? Position du problème:

Pour hiérarchiser la pertinence de site web sur un domaine, plusieurs solutions assez naturelles existent:

- Il est possible de faire appel à une sommité (personne ou groupe de personnes) du domaine pour classer les pages. Le classement serait certes très pertinent pour ce domaine, mais il faudrait solliciter un grand nombre de spécialistes de différents domaines pour obtenir un algorithme pouvant traiter des requêtes dans tous les domaines possibles et imaginables.

Il se pose aussi la difficulté de gérer le très grand nombre de pages évoquant une requête (plusieurs millions au minimum) ainsi que l'actualisation fréquente de ce classement.

- Une deuxième solution serait de demander aux internautes eux-mêmes de voter, pour chaque domaine, et de choisir le classement sur la base de ce vote, considérant que, compte tenu du grand nombre de votants, le classement obtenu serait pertinent.

Ces deux modèles, pourtant intéressants et logiques (ils sont mis en œuvre par Wikipédia), font intervenir l'humain et trouvent leur limite dans le trop grand nombre de pages à gérer qui, de plus, sont évolutives.

Ce qui conduit à privilégier des protocoles entièrement automatisables sous forme d'algorithme.

Dans cette optique, Google a cherché un modèle de hiérarchisation qui soit exploitable dans tous les domaines, utilisable pour tous les mots clés, adaptable à un très grand nombre de données, même évolutives, tout en étant automatisable et suffisamment efficace.

C'est en répondant à ce cahier des charges que ce nouveau venu a réussi l'exploit, en quelques mois et malgré l'émergence de Bing ou encore Qwant, à obtenir le quasi-monopole de la recherche thématique sur le web.

L'idée à la base du modèle de Larry Page et Sergey Brin, fondateurs de Google, revient à attribuer à chaque page un nombre positif entre 0 et 1, appelé score (en anglais PageRank) de la page, qui caractérisera la pertinence de cette page. Ils proposent alors de déterminer ce score à partir des deux règles suivantes:

- R1: Le score attribué à une page doit être d'autant plus élevé que celle-ci est référencée dans une page faisant autorité (dont le score élevé)
- R2: Le score attribué à une page doit être d'autant moins élevé que celle-ci est référencée dans une page contenant un grand nombre de références.

Leur idée: utiliser un surfeur aléatoire.

Principe du surfeur aléatoire:

Après avoir fait la liste (sans classement) de tous les sites traitant la requête, le surfeur aléatoire en choisit un au hasard. Puis il s'intéresse aux liens hypertextes du site sur lequel il se trouve vers les autres sites qu'il a listés. Il en choisit alors un au hasard et répète cette opération sans s'arrêter en comptant pour chacun des sites combien de fois il l'a visité.

Les sites sont alors affichés dans l'ordre décroissant de leur nombre de visites.

Ainsi pour un certain mot clé rentré, il s'intéresse aux sites qui évoquent ce mot-clé, mais également aux liens hypertextes qui permettent de passer d'un site à l'autre.

Activité – Comment classer les pages web issues d'une recherche ?

Les moteurs de recherche permettent de trouver des informations dans des pages dont on ne connaît pas l'adresse, voire même, dont on ignore l'existence. On souhaite faire une recherche avec les mots-clés "Apollo 11", sur 3 moteurs de recherche différents (Google, Qwant, Yahoo).

1 – Effectuer la recherche "Apollo 11" sur les 3 moteurs proposés et compléter le tableau suivant avec vos résultats de recherche:

Moteur de recherche	Présence d'une annonce publicitaire en tête des résultats	Position du premier liens vers Wikipédia	Position du premier lien vers un média culturel	Position du premier liens vers le film Apollo 11
Google				
Qwant				
Yahoo				

2 – Expliquer pourquoi le classement des résultats est différent en fonction du moteurs de recherche:

3 – Quel moteur de recherche conseilleriez-vous ? (justifier votre réponse):

4 – Pourquoi est-il important pour une page web d'être en tête d'un classement de recherche ?:

5 – La méthode du "*spamming*" consiste à écrire du texte blanc sur fond blanc dans une page web. Expliquez pourquoi le **tri par pertinence** est considéré comme peu fiable par les moteurs de recherche actuels:

Testez-vous ! Sécuriser sa navigation sur le web

Depuis les [révélations d'Edward Snowden \(2013\)](#) ou encore le [scandale Cambridge Analytica \(2016\)](#), les internautes sont désormais plus sensibles aux questions de sécurité et d'anonymat lors d'une navigation sur le web. Menons l'enquête pour savoir si votre navigateur web est bien paramétré !

Afin d'évaluer votre niveau de sécurité sur le Web, faites le test suivant en surlignant la réponse:

1 – Mettez-vous à jour votre navigateur web ?

- ① Régulièrement
- ② De temps en temps
- ③ Jamais

2 – Utilisez-vous un moteur de recherche sans cookies, comme Qwant ou DuckDuckGo ?

- ① Oui

- ② Parfois
- ③ Non

3 – Supprimez-vous régulièrement les cookies et l'historique de vos visites sur le Web ?

- ① Oui, chaque jour
- ② Parfois
- ③ Non, jamais

4 – La connexion Wi-Fi à votre box est-elle sécurisée par un mot de passe *fort* (c'est-à-dire composé de majuscules, de minuscules, de chiffres, de caractères spéciaux, en nombre supérieur à 15 caractères) ?

- ① Oui
- ② Non
- ③ L'accès à ma box n'a pas de mot de passe

5 – Les mots de passe que vous utilisez sont-ils:

- ① Fort ?
- ② Un mot du dictionnaire ?
- ③ Ma date de naissance ?

6 – Vos mots de passe sont-ils différents pour chaque applications ?

- ① Oui, bien sûr !
- ② Oui, je les enregistres tous dans un fichier nommé "motdepasse" sur mon ordinateur !
- ③ Non, surtout pas !

7 – Limitez-vous la diffusion de vos informations personnelles sur le Web (réseaux sociaux, etc) ?

- ① Systématiquement
- ② Cela dépend du contenu à diffuser
- ③ Pas du tout

8 – Utilisez-vous systématiquement le bouton "déconnexion" en quittant une application Web ?

- ① A chaque fois
- ② Parfois
- ③ Jamais

9 – Vous vous connectez à un réseau Wi-Fi public en HTTP avec un smartphone (aéroport, lycée, etc):

- ① Vous ne composez aucun de vos mots de passe
- ② Vous ne changez rien à vos habitudes
- ③ Vous en profitez pour vous connecter à votre messagerie ou à votre banque

Résultats: Comptabilisez le nombre de numéros de chaque type