

1. DESCRIPTION

Un moteur de recherche est une application web permettant de trouver des ressources à partir d'une requête sous forme de mots. Les ressources peuvent être des pages web, des articles de forums, des images, des vidéos, des fichiers, etc. Certains sites web offrent un moteur de recherche comme principale fonctionnalité; on appelle alors moteur de recherche, le site lui-même (Google, Bing, Yahoo, Qwant..).

Un moteur de recherche est un instrument de recherche sur le web sans intervention humaine. Il est basé sur des robots (ou *bots*, *spiders*, *crawlers*...) qui parcourent les sites à intervalles réguliers et de façon automatique pour découvrir de nouvelles adresses (URL). Ils suivent les liens hypertextes qui relient les pages les unes aux autres. Chaque page identifiée est alors indexée dans une base de données, accessible ensuite à partir de mots-clés.

2. FONCTIONNEMENT D'UN MOTEUR DE RECHERCHE

Il se décompose en trois processus principaux :

- L'exploration ou *crawl* : un robot d'indexation suit tous les hyperliens qu'il trouve et récupère les ressources jugées intéressantes
- L'indexation des ressources récupérées consiste à extraire les mots considérés comme significatifs. Les mots extraits sont enregistrés dans une base de données.
- La recherche correspond à la partie *requêtes* du moteur, qui restitue les résultats (documents correspondant le mieux aux mots contenus dans la requête).

Des modules complémentaires sont souvent utilisés en association avec les trois briques de bases du moteur de recherche. Les plus connus sont les suivants :

- Le correcteur orthographique : permet de corriger les erreurs introduites dans les mots de la requête.
- L'anti-dictionnaire : utilisé pour supprimer à la fois dans l'index et dans les requêtes tous les mots vides (tels que de, le, la) qui sont non discriminants et perturbent le score de recherche.

3. EFFECTUER DES RECHERCHES

L'interface (partie la plus importante, souvent simple et rapide), présente une zone de recherche pour la saisie des mots-clés. Une zone de recherche, située en haut de chaque page de résultat, permet de reformuler la requête sans revenir sur la page d'accueil.

Après appui sur **Entrée** ou clic sur le bouton *Rechercher*, la réponse affiche les liens vers les pages qui correspondent à la requête ainsi qu'un extrait de chacune de ces pages.

Pour accéder directement au premier résultat sur Google (page correspondant au premier lien de la page de résultats), clic sur le bouton *J'ai de la chance*. Au lieu d'afficher la page de résultats, Google affiche la page qui semble la plus pertinente (1^{er} résultat, sans tenir compte des liens commerciaux).

Les moteurs de recherche affichent aussi des propositions de requête en fonction de la saisie au clavier. Il suffit de sélectionner une proposition pour lancer la recherche et afficher les pages de résultat.

4. CHOISIR LES TERMES DE RECHERCHE

Pour des résultats plus pertinents, il est indispensable de bien comprendre le fonctionnement pour lancer des requêtes efficaces.

- Utiliser les mots se trouvant sur les pages souhaitées. Par exemple pour trouver la date de décès de Napoléon 1^{er}, la recherche "napoleon 1 mort le" sera plus pertinente que "date décès napoleon 1".
- Ne pas commencer les requêtes par les termes : discussion sur..., article sur..., documentation sur..., page sur... Se limiter aux termes de recherches qui sont censés apparaître dans les pages.
- Eviter les termes trop généralistes : une recherche sur "développement fœtus" ou " étapes fœtus" sera plus précise que "fœtus".
- Etre bref : par exemple, Google limite la requête à 10 mots mais moins permet de meilleurs résultats. Au lieu de "programme pour supprimer la dépendance à la cigarette", choisir "programme arrêter fumer".
- Termes ignorés : les mots courants sont ignorés (le, la, un, du, avec, vous, je...) ainsi que la plupart des termes d'un seul caractère (à, y...). Si un seul mot dans la requête (mot ignoré), la recherche sera lancée.
- Majuscules, minuscules et accents : aucune différence entre majuscule et minuscule (la casse est ignorée). Tous les signes diacritiques sont aussi ignorés (accents, cédille, tréma, tilde...).
- Exclusion de la ponctuation : la majorité des signes de ponctuations est ignorée (pique-nique et pique nique, Dr. Freud et dr freud...).
- Ordre et proximité des mots-clés : l'ordre des termes de recherche a une importance. Les pages contenant les mots dans le même ordre que la requête sont placées en premier. Les pages contenant les termes de recherche proches l'un de l'autre seront favorisées.
- Opérateur d'exclusion : permet d'exclure les résultats de recherche incluant un mot ou site précis (précéder ces derniers d'un tiret). Ce caractère est particulièrement utile pour des homonymes.

5. INTERPRÉTER LES RÉSULTATS D'UNE RECHERCHE

Après avoir lancé la recherche, la page de résultats s'affiche. Elle contient les informations et les liens vers d'autres pages.

- La partie supérieure sous le bandeau web peut contenir des liens commerciaux (publicité achetée par des sociétés), idem colonne de droite ou bas de page.
- Les parties supérieure et inférieure de la page affichent des informations sur la recherche en cours et offrent des outils pour affiner celle-ci :
 - Liens vers d'autres outils (web, images, actualités, maps, vidéos...).
 - Outils de recherche (web, page francophone, page France...).
 - Barres de statistiques (nombre de résultats et temps).
 - Zone de recherche semblable à la page d'accueil avec les termes de la recherche.
 - Boutons pour passer d'une page à l'autre.
 - Recherche associées aux termes de la recherche (les plus pertinentes pour le moteur de recherche).

Résultats de la recherche : les pages les plus pertinentes sont affichées en premier.

- Liens vers les résultats : les titres des pages web s'affichent sous forme de lien. Si la page n'a pas de titre, affichage de l'adresse URL. Titre = meilleur moyen de trouver une page.

- Extrait de la page : extrait du texte, dans lequel les termes de recherche apparaissent en gras, mais pas toujours si ils sont dans l'adresse web ou le texte d'un lien. L'extrait permet de connaître l'information recherchée sans avoir à ouvrir la page, un aperçu du contenu de la page liée, les termes à utiliser pour une recherche plus approfondie.
- Informations techniques : l'adresse du résultat (URL) s'affichant sous le titre permet de connaître le site web.
- Liens Traduire cette page : une traduction d'une page écrite dans une langue étrangère, peut être réalisée par ordinateur. Cette traduction est loin d'être parfaite, mais permet d'en obtenir le sens général.

Conversion de fichiers : un lien peut renvoyer vers des fichiers de divers formats. Certains formats sont indexés (convertis au format HTML ou texte). Utile si l'on ne possède pas le logiciel ou si le fichier ne s'affiche pas (pdf, xls, ppt, doc...).

Plusieurs résultats pour un même site : lorsque plusieurs résultats se trouvent sur un même site, le premier s'affiche et les autres en retrait.

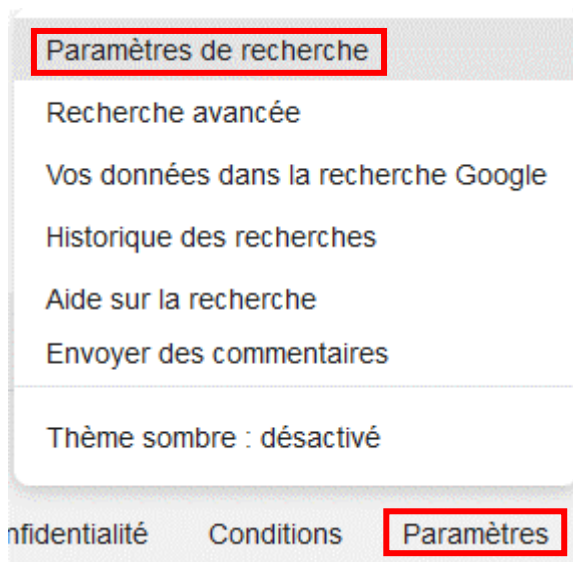
Afficher les différentes pages : pour passer d'un page de résultat à une autre, clic sur un numéro qui apparaît en bas de la page de résultats ou suivant.

Effectuer des recherches avancées. Pour affiner une recherche, formulaire de recherche.

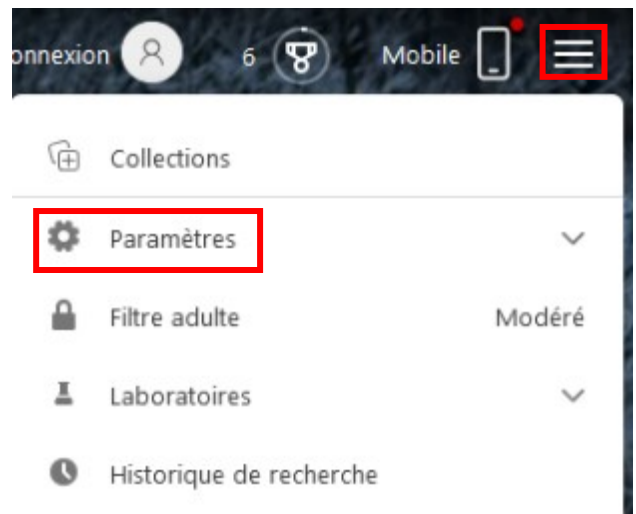
Historique de recherche. Il permet d'appuyer les nouvelles recherches sur les recherches passées.

6. PARAMÉTRER LE MOTEUR DE RECHERCHE

Il est possible de paramétrer le moteur de recherche :



Google



Bing