

SNT – MOTEURS DE RECHERCHE



Quel moteur de recherche utilisez-vous ?

– sur votre smartphone :

– sur votre PC :

COMMENT ÇA FONCTIONNE, UN MOTEUR DE RECHERCHE ?

Le fonctionnement d'un moteur de recherche se décompose en trois processus principaux :

1) L'**exploration** : le web est systématiquement exploré par un robot (le *crawler* ou *spider*) suivant récursivement tous les hyperliens qu'il trouve et récupérant les ressources jugées intéressantes.

Il analyse des éléments très importants tels que le code source, l'URL, les images et le nom qui leur est donné, les liens internes et externes, les pages sur lesquelles pointent ces liens, etc.

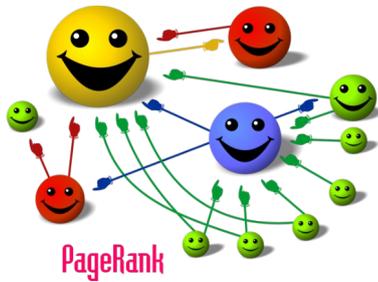
2) L'**indexation** : ce robot indexe alors les ressources récupérées, en extrayant les mots considérés comme significatifs du corpus à explorer. Les mots extraits sont enregistrés dans une base de données organisée comme un gigantesque dictionnaire inverse ou, plus exactement, comme l'index terminologique d'un ouvrage, qui permet de retrouver rapidement dans quel chapitre de l'ouvrage se situe un terme significatif donné.

Remarque : ces termes significatifs sont associés à un poids. Celui-ci augmente (entre autres) proportionnellement au nombre d'occurrences du mot dans le document.

L'indexation s'effectue après un ensemble de traitements sémantiques (suppressions des mots tels que " de ", " le ", " la ", etc.) et algorithmiques selon différentes techniques basées aujourd'hui sur l'intelligence artificielle.

3) La **recherche** dans le moteur : un algorithme est appliqué pour identifier dans le corpus documentaire (en utilisant l'index), les documents qui correspondent le mieux aux mots contenus dans la requête, afin de présenter les résultats des recherches par ordre de pertinence supposée.

Les algorithmes de recherche font l'objet de très nombreuses investigations scientifiques et utilisent massivement les mathématiques.



PageRank

Le PageRank est un des algorithmes d'analyse des liens utilisés par le moteur de recherche Google : il mesure quantitativement la popularité d'une page web.

Ce système a été inventé par Larry Page (de son vrai prénom Lawrence), cofondateur de Google.

D'autres critères que les liens entrants sont pris en compte dans le calcul du PageRank, la recette exacte étant gardée secrète par Google : les liens sortants, les ancres, le trafic de la page, le comportement des visiteurs, le nom de domaine ou encore l'hébergement influencent le score.

Depuis le 15 avril 2016, Google a officiellement arrêté d'afficher publiquement le PageRank de chaque page/site. Mais si Google met fin à la communication publique de ce score, elle utilise toujours ce système en interne pour classer les pages du Web.

Si vous souhaitez comprendre quelle est l'idée principale du PageRank, voir [cette vidéo](#) d'env. 9 min.

UN MÉTAMOTEUR EST-IL UN MOTEUR DE RECHERCHE ?

Google, Baidu, Bing, DuckDuckGo, Ecosia, Qwant... ne sont pas tous vraiment des moteurs de recherche !

Un **métamoteur** est un outil qui envoie ses requêtes à plusieurs moteurs de recherche et retourne les résultats de chacun d'eux. Le métamoteur permet aux utilisateurs de n'entrer le sujet de leur recherche qu'une seule fois tout en accédant aux réponses de plusieurs moteurs de recherche différents.

Un métamoteur élimine les résultats similaires (par exemple, si *Google* et *Yahoo* ! renvoient sur les deux mêmes liens, le métamoteur ne va l'afficher qu'une seule fois dans la liste des résultats).

Un métamoteur trie les résultats comme il le souhaite : certains métamoteurs permettent en outre de mélanger une fonction annuaire (les résultats sont classés par thèmes) et une fonction " moteur ".

Quelques métamoteurs :

- *DuckDuckGo*
- *Ecosia* (à vocation écologique : reverse 80 % de ses bénéfices dans le but de replanter des arbres et ainsi contrer son empreinte carbone)
- *Lilo* (français, permet de contribuer à des projets sociaux et environnementaux de son choix)
- *Searx* (libre, sous licence AGPLv3)
- *Mamma.com* (anglophone, spécialisé pour les recherches à caractère commercial : coupons, bons de réduction, promotions, etc.)
- *YouCare* (français, basé sur les résultats d'un service Microsoft, permet de financer des repas pour les animaux en refuges).

CHERCHER SUR GOOGLE

Si je vous demande " Quels sont les impacts du réchauffement climatique sur notre environnement et notre santé ? ", que tapez-vous dans Google pour avoir des éléments de réponse ?

Si vous tapez la phrase entière, Google renvoie près de 14 600 000 résultats !

Si vous tapez *réchauffement climatique*, vous aurez 8 790 000 résultats : Google, dans ce cas, cherche toutes les pages, images, vidéos, etc. contenant ces mots, individuellement, " réchauffement ", " climatique ", mais aussi ensemble " réchauffement climatique ".

Si vous tapez " *réchauffement climatique* ", vous aurez 6 690 000 résultats

À savoir :

- Google ne tient pas compte des accents (été = ete) ni de la casse (majuscule = minuscule).
- La recherche Google ignore généralement les signes de ponctuation qui ne font pas partie d'un opérateur de recherche.
- L'ordre des mots a un impact sur les résultats ainsi que la répétition des termes.
- Par défaut, les résultats de recherche sont liés à votre localisation.

Il existe des techniques qui permettent d'affiner nos recherches.

Voici les principaux opérateurs à connaître :

Exemple de recherche	Explication
"intelligence économique"	Les guillemets permettent de rechercher une expression exacte.
intelligence économique	L'espace est l'opérateur par défaut. Ici Google va rechercher les liens contenant du texte avec les termes <i>intelligence</i> et <i>économique</i> .
amour OR gloire	L'opérateur OR permet de rechercher un terme ou un autre (ou les deux).
communication -veille	Le signe - permet d'exclure un terme. Ici la requête permet de connaître les pages contenant le terme <i>communication</i> , mais exclue celles qui contiennent <i>veille</i> . Le moins doit être collé au terme à exclure.
"barack * obama"	L'astérisque est souvent utilisé pour remplacer un ou des mots (mais pas une suite de lettres dans un mot). Ici la requête permet de trouver rapidement le deuxième prénom de Barack Obama.
(ipad OR iphone) Apple	Les parenthèses regroupent des termes ou des opérateurs de recherche pour aider à structurer une recherche avancée.

Quelques opérateurs avancés, bien utiles :

Exemple de recherche	Explication
<code>intitle:maths</code>	Permet de rechercher dans le titre d'une page un ou plusieurs termes. S'il y a plusieurs termes après <i>intitle</i> l'opérateur recherchera nécessairement le premier terme dans le titre, les suivants seront recherchés soit dans le titre, soit dans le texte, soit dans l'URL.
<code>allintitle:maths seconde</code>	Permet de rechercher dans le titre d'une page tous les termes : ne peut pas se combiner avec d'autres opérateurs.
<code>allintext:maths seconde cours</code>	Permet de rechercher des pages dont le texte contient tous les termes indiqués. Ne peut pas se combiner avec d'autres opérateurs.
<code>inurl:amour</code>	Permet de rechercher dans l'URL d'une page un ou plusieurs termes. S'il y a plusieurs termes après <i>inurl</i> l'opérateur recherchera nécessairement le premier terme dans l'adresse URL, les suivants seront recherchés dans l'URL ou dans le texte.
<code>allinurl:amour gloire</code>	Permet de rechercher dans l'URL d'une page tous les termes : ne peut pas se combiner avec d'autres opérateurs.
<code>filetype:pdf</code>	Limite la recherche au type de fichier spécifié. Ici on recherche uniquement des documents PDF. Attention, ne pas mettre de point avant le type de fichier (ne pas écrire <i>filetype:.pdf</i>).
<code>site:mathemathieu.fr</code>	Permet de rechercher sur un site spécifique.
<code>"lit de dimensions 90..400"</code>	Recherche ici l'expression "lit de dimensions" suivie d'un entier compris entre 90 et 400.
<code>"le nombre 0.." site:mathemathieu.fr</code>	Recherche ici l'expression "le nombre" suivie d'un entier naturel.
<code>veille AROUND(6) "intelligence economique"</code>	Permet de rechercher les pages web intégrant le terme <i>veille</i> et l'expression exacte "intelligence économique", séparés par au maximum 6 mots.

→ Bien sûr, vous pouvez combiner plusieurs opérateurs. Par exemple :

`"nikola tesla" intitle:"top 5..10 faits" -site:youtube.com inurl:2015`

permet de rechercher les pages qui contiennent l'expression *nikola tesla*, ont la phrase *top X faits* dans le titre où X varie de 5 à 10, ne sont pas sur YouTube et dont l'URL contient le terme *2015*.

→ Autre exemple : `"devenir * en 2..10 etapes"`.

→ À votre avis, à quoi sert la recherche suivante ?

`site:mathemathieu.fr -inurl:https -filetype:pdf`

→ Vous cherchez à savoir si quelqu'un plagie votre site ? Utilisez une phrase unique de votre texte, mettez-la entre guillemets après un opérateur *intext:*, et excluez votre propre site avec *-site:*. Exemple :

`intext:"ils s'amusaient dans nos entrailles" -site:moz.com`

Enfin, Google permet de filtrer notre recherche à travers quelques outils :

The screenshot shows a Google search bar with the text "citation maths". Below the search bar, there are navigation tabs: "Tous", "Images", "Actualités", "Vidéos", "Shopping", "Plus", "Paramètres", and "Outils". Below the tabs, it says "Environ 9 290 000 résultats (0,47 secondes)".

Les onglets *Images*, *Actualités*, *Vidéos*, etc. Permettent de préciser notre recherche.

Dans *Paramètres*, on peut accéder à une recherche avancée et préciser de nombreux critères pour affiner notre recherche :

Recherche avancée

The screenshot shows the Google Advanced Search page. It is divided into two main sections: "Trouvez des pages avec..." and "Affinez ensuite la recherche par...".

Trouvez des pages avec...

- tous les mots suivants : citation maths
- ce mot ou cette expression exact(e) :
- l'un des mots suivants :
- aucun des mots suivants :
- nombres compris entre :

Pour effectuer cette opération dans le champ de recherche

- Saisissez les mots importants : terrier tricolore
- Ajoutez des guillemets autour des mots exacts : "terrier"
- Saisissez OR entre tous les mots à inclure : miniature OR standard
- Placez un signe - (moins) devant les mots à exclure : -rongeur, -"Jack Russell"
- Placez deux points entre les nombres, et ajoutez une unité de mesure : 10..35 kilos, 300..500 USD, 2010..2011

Affinez ensuite la recherche par...

- langue : toutes les langues
- région : tous les pays/territoires
- dernière mise à jour : à une date indifférente
- site ou domaine :
- termes apparaissant : n'importe où dans la page
- SafeSearch : Afficher les résultats explicites
- type de fichier : tous les formats
- droits d'usage : non filtré par licence

Recherche avancée

Enfin, si je clique sur *Outils*, je peux préciser rapidement la langue de recherche ainsi qu'une date précise ou une plage de dates :

The screenshot shows a Google search bar with the text "Citation de mathématicien". Below the search bar, there are navigation tabs: "Tous", "Images", "Actualités", "Vidéos", "Shopping", "Plus", "Paramètres", and "Outils". Below the tabs, there are filters: "Toutes les langues", "Date indifférente", and "Tous les résultats".

Below this, there is another search bar with the text "maths citation". Below this search bar, there are navigation tabs: "Tous", "Images", "Actualités", "Vidéos", "Shopping", "Plus", "Paramètres", and "Outils". Below the tabs, there are filters: "Taille", "Couleur", "Type", "Période", and "Droits d'usage".

TRAVAIL À FAIRE SUR GOOGLE OU UN AUTRE MOTEUR

1. Tester et analyser les recherches suivantes :

mercure planete "caractéristiques physiques"

film critique "le seigneur des anneaux"

"recettes de cuisine" poisson -saumon

sortie film (italien OR espagnol)

universite montreal site:.ca

cours maths ("terminale OR premiere") -partie filetype:pdf

allinurl:pdf docs maths

allintext:pdf docs maths

"histoire * cinema" site:.fr inurl:https

histoire cinema site:.fr inurl:https

2. **a.** Saisissez "escargot de bourgogne" : constatez que l'on trouve à la fois des sites sur les escargots de bourgogne mais aussi des recettes de cuisine.

b. Modifier la recherche en excluant toute recette de cuisine.

c. Modifier la recherche en limitant les résultats à ceux datant de moins d'un mois.

3. Faire les recherches sur les sujets de votre choix et manipuler les outils découverts aujourd'hui.

4. Voir cette courte vidéo « Comment les algorithmes nous enferment ? » et vérifier avec un ami, si possible sur votre smartphone, que la même recherche ne renvoie pas les mêmes résultats (personnalisation du résultat en fonction de votre historique) :



<https://www.lumni.fr/video/comment-les-algorithmes-nous-enferment>