

Moteur de recherche Google

Un moteur de recherche

C'est une application web permettant de retrouver des ressources (pages web, articles de forums Usenet, images, vidéo, fichiers, etc.) associées à des mots quelconques. Certains sites web offrent un moteur de recherche comme principale fonctionnalité ; on appelle alors moteur de recherche le site lui-même (<http://scholar.google.fr/>) .

Les moteurs de recherche ne s'appliquent pas qu'à Internet : certains moteurs sont des logiciels installés sur un ordinateur personnel (PC). Ce sont des moteurs dits desktop qui combinent la recherche parmi les fichiers stockés sur le PC et la recherche parmi les sites Web — on peut citer par exemple Exalead Desktop, Google Desktop et Copernic Desktop Search, etc. etc.

des annuaires .

Ce sont des instruments de recherche élaborés par des personnes qualifiées , qui répertorient et classifient des sites web jugés dignes d'intérêt, et non des robots d'indexation — on peut citer par exemple **Voilà** et **Yahoo!**

Les Métamoteurs

On trouve également des métamoteurs, c'est-à-dire des sites web où une même recherche est lancée simultanément sur plusieurs moteurs de recherche (les résultats étant ensuite fusionnés pour être présentés à l'internaute) — on peut citer:

Ixquick, <https://www.ixquick.com/fra/> **Mamma**, <http://www.mamma.com/>
Kartoo, <http://www.kartoo.com/> **Seek.fr**, <http://www.seek.fr/>
metacrawler . <http://metacrawler.com>

La « toile visible » et la « toile invisible »

On parle souvent de la toile visible et de la toile invisible sur le web, mais qu'est-ce que cela signifie ? En fait une part importante des ressources Internet n'est pas explorée par les robots des moteurs (soit parce les pages sont en accès réservé, soit parce qu'elles appartiennent aux banques de données). Schématiquement:

- La Toile visible peut être assimilée à celle explorée par les moteurs type Google.
- La Toile invisible est donc la partie complémentaire non accessible aux moteurs classiques, donc les pages appartenant aux banques de données (Medline, Inist, Cismef...) et toutes les pages en accès réservé (pages à accès professionnel restreint, aux abonnés d'un journal, aux membres d'une association...).

Le robot – explorateur (« spider »)

- C'est un robot logiciel qui explore de façon autonome le « Web ».
- Son efficacité est capitale pour le moteur.
- Identifie les liens des pages, puis les visite à son tour, parcourant ainsi rapidement la totalité du site, puis ceux qui lui sont rattachés.
- Examine périodiquement des millions de pages et constitue de cette façon une base de données de celles déjà visités.

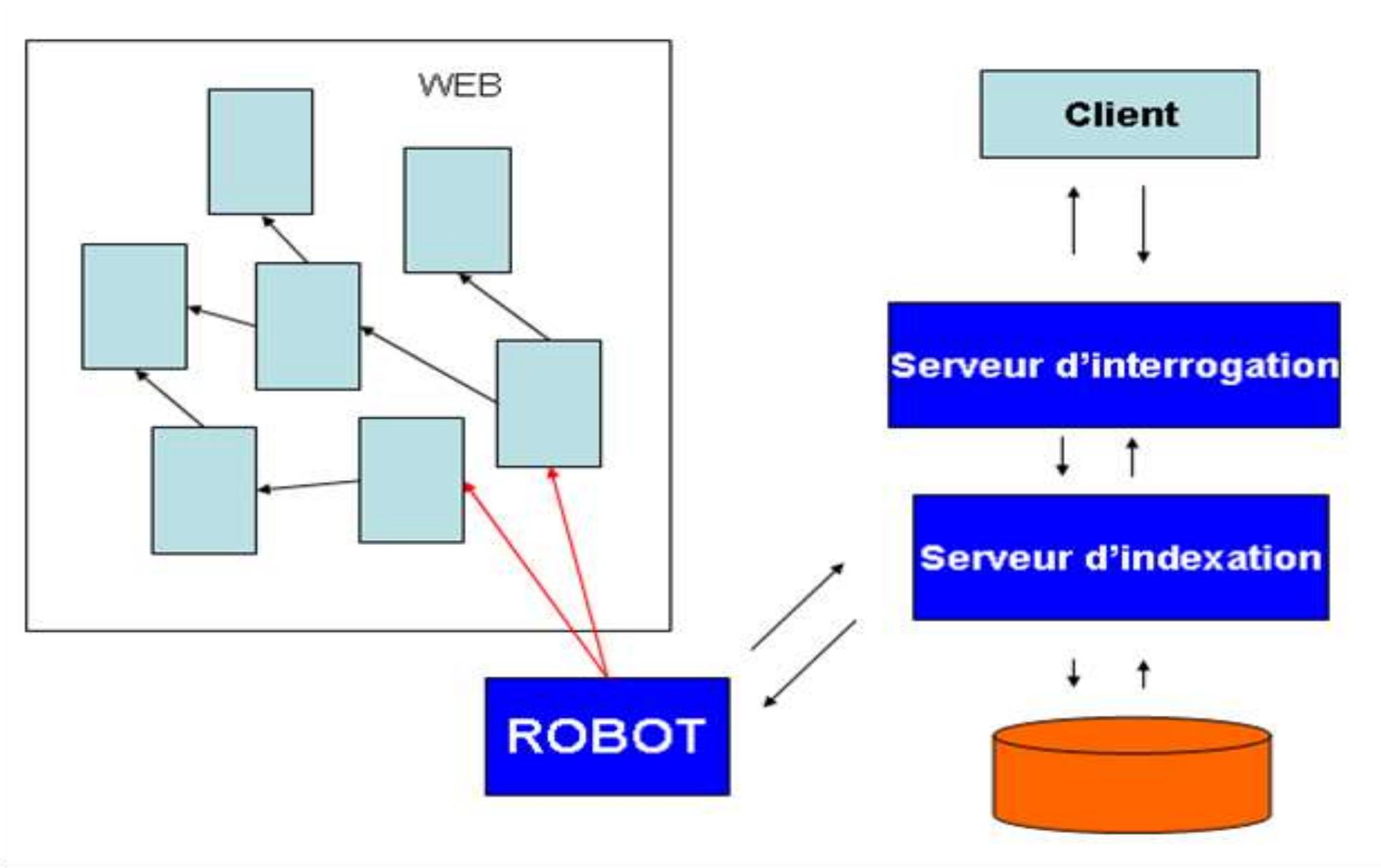
Le système d'indexation

- Analyse les informations collectées, construit un index des mots rencontrés (et des pages correspondantes), puis stocke l'ensemble dans une base de données.
- Convertit certains fichiers ne pouvant pas être indexés à cause de leur format.
- Utilise des outils d'extraction pour ne récupérer que « l'essence » des documents. Ex: Fulcrum, Infoseek, Intelliserv, Livelink.

Module de recherche (« searcher »)

- Le searcher est le frontal de l'utilisateur.
- Grâce à son interface graphique, on peut poser une question, sélectionner les options disponibles, et lancer une recherche.
- Un script faisant alors appel au système d'indexation pour qu'il exécute la requête sur la base de données.
- Les résultats affichés sous forme de page Web intégrant généralement les réponses sous forme de liste.





Google (1998)

- L'Université de Stanford produit les concepteurs d'un moteur de recherche baptisé Google, Sergei Brin (à gauche) et Larry Page (à droite).



Origine du nom

En 1938, le mathématicien américain Edward Kasner, se demandant quelle nom donner au nombre formé du chiffre 1 suivi de 100 zéros, eue la réponse de son neveu (9 ans): mot enfantin “un gogole”, dont l’ortographe est devenue “googol”.

Kesner reporta ce mot dans son traité “Mathematics and the imagination”, qui a en suite inspiré en 1998 les créateurs de “Google” (au lieu de “googol”), plus simple et facile à retenir.

- On peut ainsi noter la similarité avec le mot anglais “goggles” (“lunettes”), dont les deux zéros de la marque rappellent la forme.

Technologie Page Rank

- Google repose principalement sur l'exploitation de la technologie Page Rank.
- Le premier brevet, déposé en janvier 1997 et enregistré le 9 janvier 1998, est la propriété de l'Université de Standford qui à licencié cette licence à Google en 1998, deux mois après sa fondation.
- La licence est exclusive jusqu'en 2011.
- Les recherches qui ont abouties au développement de la technologie du Page Rank ont été financées en partie par la National Science Fundation. Il est précisé dans le brevet que le gouvernement a certains droits sur cette invention.

- Le Page Rank c'est la note de 0 à 10 attribuée par Google aux sites Web.
- Le moteur de recherche base tout son fond de commerce dessus, car la course au PageRank est devenu le sport international des Webmasters.
- Google analyse de façon très précise les pages Web et fichiers disponibles Online:
 - 1) il classe toutes ces pages dans son index de plusieurs milliards de pages.
 - 2) le moteur de recherche étudie la popularité de la page, pour lui donner une note, le Page Rank, qui est le résultat d'un savant algorithme.
- Cette note est un paramètre pour évaluer la pertinence d'une page par rapport aux critères de recherche.
- Google est donc performant car ce sont les résultats les plus pertinents par rapport à la recherche qui vont sortir en premier.

Seul point négatif: les Webmasters les plus malins arrivent souvent à se positionner sur les premières pages de résultat (par exemple utilisant des mots-clé très recherchés sur Internet).

L'algorithme de Google progresse et rattrape les tricheurs, mais ceux-ci trouvent de nouvelles techniques pour positionner de nouveau leurs sites sur le moteur.

- L'index de Google est encore majoritairement propre.
- Le souci est qu'il n'y a que 10 places sur la première page de résultats et trop de sites veulent cette visibilité maximum.

Page Rank

- Page Rank : calcule le nombre de fois qu'un site X est cité par d'autres sites A, B, C, D...
- Mais il prend aussi en compte la popularité des sites A, B, C, D: être cité par des sites qui sont eux mêmes populaires augmente la popularité de X
- Inconvénient : des **sites récents n'apparaîtront pas tout de suite** (pas le temps d'être cités)

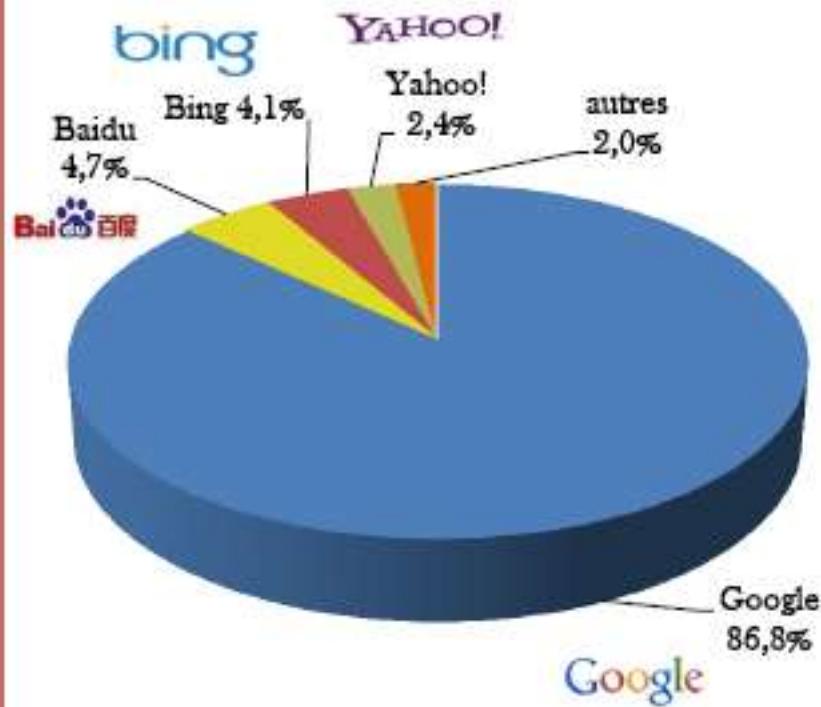
Mission et stratégie

Google est l'outil de recherche le plus utilisé au monde, avec plus de 70% de parts de marché.

- Sa mission est d'organiser l'information du monde entier et de la rendre accessible à tous.
- Le succès de ce moteur de recherche se résume dans le principe du “simple et efficace”:
 - l'interface dépouillée,
 - la rapidité de réponse,
 - la pertinence des résultats,
 - le grand nombre de pages indexées,
 - la variété de types de fichiers indexés.
- Utilisant une stratégie de marketing viral très efficace Google domine radicalement le marché de la recherche sur le Web, malgré les tentatives de ses concurrents pour le contrer.

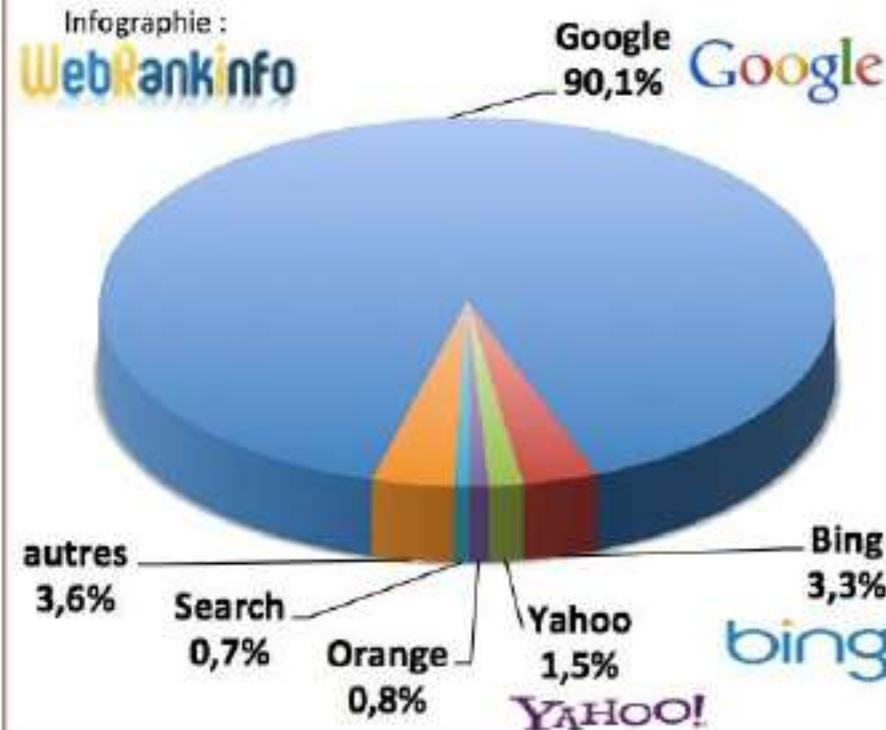
Monde

*Parts de marché des moteurs de recherche dans le monde, mars 2012
(Karmasnack)*



France

Parts de marché des moteurs de recherche en France, décembre 2011 (AT Internet)



Sites les plus visités



Google, les chiffres :

- ✓ Un index de plus de 40 milliards de pages (Google 2013, Bing 3.8 milliards)
- ✓ 44 000 employés dans le monde (2013)
- ✓ Plus de 110 sites régionaux, de 130 interfaces linguistiques
- ✓ La marque la plus populaire au monde(ou 2ème)
- ✓ Plus de 4 milliards de requête/jour(2012)
- ✓ Slogan « Don'tbeevil»

Points faibles

- ✓ la couverture est limitée – Google indexe au plus 1 à 5 % de l’information en ligne. Cela signifie que Google ne tient pas compte de 95 à 99 % du Web?
- ✓ Google présente les résultats les plus populaires, et non les plus crédibles ou les plus importants?
- ✓ les résultats obtenus sur Google ne sont pas toujours exacts?
- ✓ les spécialistes de l’optimisation des moteurs de recherche «appâtent» Google?
- ✓ toutes vos recherches sont enregistrées et pourraient être retenues contre vous par les autorités?

La personnalisation de la recherche

« L'algorithme [de Google] utilise plus de 200 paramètres pour affiner une requête. Cela inclut le PageRank du site internet, la localisation géographique de l'internaute, les liens sur lesquels il clique généralement, la façon dont il modifie ses requêtes quand il n'est pas satisfait, ainsi que son historique de recherche »



comment afficher le moteur de recherche Google?

Cliquez dans la barre d'adresse : toute l'adresse est sélectionnée. Appuyez sur la touche « suppr » du clavier pour faire disparaître le texte.

Tapez **www.google.fr** dans la barre d'adresse et validez (en appuyant sur la touche « Entrée » de votre clavier).

description du moteur de recherche Google

Google représente **un ensemble de services gratuits**. En effet le moteur de recherche Google propose des recherches thématiques: vous pouvez ne rechercher que des images, des vidéos, des cartes, etc. Par défaut, vous êtes dans le service web de Google.



Web [Images](#) [Vidéos](#) [Maps](#) [Actualités](#) [Livres](#) [Gmail](#) [plus ▾](#)

description d'une page de réponse Google

Google est un service gratuit. Les propriétaires de Google gagnent de l'argent en vendant de l'espace publicitaire aux entreprises pour apparaître sur les pages de réponse de Google lorsque certains mots-clés sont choisis. Vous verrez donc parfois apparaître des **liens commerciaux** avant les réponses de Google et/ou sur la partie droite de la page.

Les vrais réponses du moteur de recherche Google apparaissent à la suite. Elles sont structurées en trois parties.



voyage

Rechercher

[Recherche avancée](#)Rechercher dans : Web Pages francophones Pages : FranceWeb [Afficher les options...](#)

Résultats 1 à 10 sur un total d'environ 70 800 000 pour voyage avec SafeSearch (0,16 secondes)

Voyage Discountwww.promovacances.com/Voyage + de 3000 offres en promotion ! dans 70 Pays. c'est Promovacances[Liens commerciaux](#)**Voyages - Promotions**Nouvelles-Frontieres.fr/voyages + 500 promos Voyages à saisir Cap sur les destinations Soleil ![Liens commerciaux](#)**Voyage : Prix Cassés**www.voyages-sncf.com Offre Spéciale : Voyage Bas Prix ! Vol, Hotel, Séjour jusqu'à -65%**Voyage**Jusqu'à -70% sur vos vacances !
Vente privée voyage, inscrivez-vous
www.voyage-prive.com**Voyage Pas Cher**Comparez les prix de votre voyage
des promotions et dernière minute.
www.VoyagerMoinsCher.com**Voyages Aquatour**Partez avec Aquatour : le N°1
des vacances au départ de Lille
www.aquatour.com**Séjour Disneyland Paris**Séjour tout compris à partir de
132€ et Gratuit pour les - de 7 ans
sejournereve.fr/Sejour-Disneyland
Île-de-France**Voyage**Partez en voyage
à petits prix avec Jet tours**Agence de voyage Thomas Cook en ligne pour réserver vos vacances ...**20 avr. 2010 ... Agence de voyage Thomas Cook Réservez vos vacances, voyages, séjours et week end en ligne, par téléphone, ou sur notre réseau d'Agences de ...
www.thomascook.fr/ - Il y a 11 heures - [En cache](#) - [Pages similaires](#)**Agence de voyages, pas cher: Nouvelles Frontières. Voyage à prix ...**Agence de voyage, NOUVELLES FRONTIERES: voyages dernière minute, voyage pas cher , séjours, hôtel, vol ? Avec notre agence de voyage, partez pour un voyage ...
[Séjours](#) - [Vols](#) - [Enchères](#) - [Promos](#)
www.nouvelles-frontieres.fr/ - [En cache](#) - [Pages similaires](#)**Agence de Voyage FRAM : séjour, week end, billet d'avion dégriffé ...**Agence de voyage Fram. Découvrez nos produits (séjour, week end, billet d'avion, hotel, hotel club, croisière..) et nos promotions voyage (séjour pas cher, ...
www.fram.fr/ - [En cache](#) - [Pages similaires](#)

Goooooooooooooogle ►

1 2 3 4 5 6 7 8 9 10

[Suivant](#)

Description d'une réponse Google

la structure en 3 parties:

1-la phrase en caractère bleu et souligné est le lien internet (ou hypertexte) vers la page internet qui contient la réponse. Cliquer sur ce lien pour visualiser la page-réponse. Notez que cette phrase apparaît en violet lorsque le lien a déjà été visité. **Pour lire une réponse Google, il faut donc cliquer sur la première ligne de la réponse.**

2-la partie en caractère noir est une courte description de la page Internet qui correspond au lien au-dessus.

3-En caractère vert c'est l'adresse Internet de la page.

Principes de base

- MAJUSCULE = minuscule
- AND par défaut
- Exclusion des mots vides (le, la, les, du, avec, de, lettres et chiffres uniques, http, .com, etc)

Principes de base



ordre des mots

A screenshot of a Google search bar. The text "blanche alger" is typed into the search field, and the search button is visible on the right.

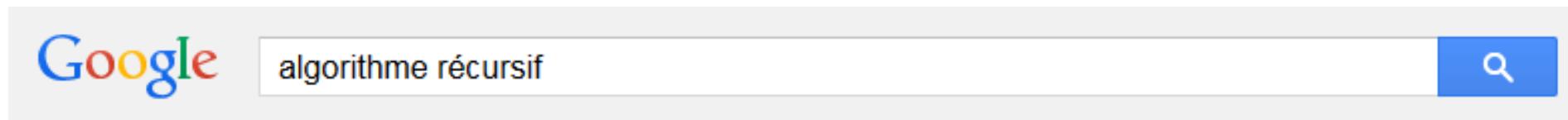
donne 2 fois plus de résultats que

A screenshot of a Google search bar. The text "alger blanche" is typed into the search field, and the search button is visible on the right.

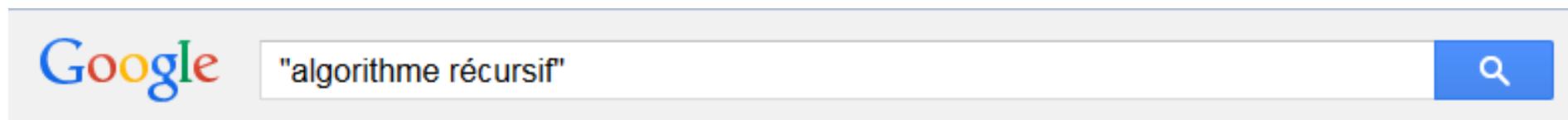
Les opérateurs de recherche

Les opérateurs

"expression exacte "

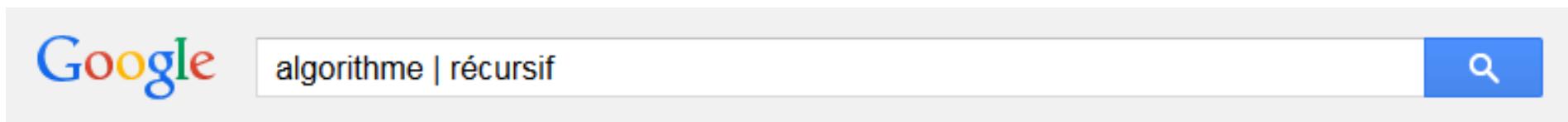


Donne plus de résultats par rapport à

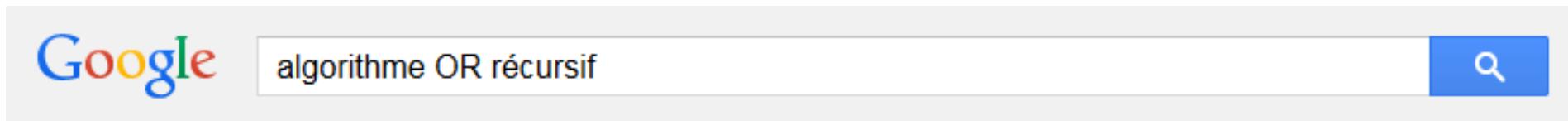


Les opérateurs

OR = |



donne le même nombre de résultats



Les opérateurs

- Exclure des significations :

santé -mentale

Rechercher

recherche sur la santé SAUF la santé mentale

Les opérateurs

+ Tenir compte d'un mot vide

Tome +II

Rechercher

Recherche sur « Tome II » ou « tome 2 »

Les opérateurs

Site: dans un site web déterminé

Recherche de « prêt » sur le site du RBDSanté

Les opérateurs

Define:

define:promotion de la santé

Rechercher

Recherche des définitions de « promotion de la santé »

Les opérateurs

link:

Recherche des liens vers le site RBDSanté

Les opérateurs

Filetype: Limite à un type de fichier

santé publique filetype:pdf

Rechercher

Recherche de pdf sur la santé publique

Les opérateurs

~ élargir aux synonymes

~toxicomanie

Rechercher

Recherche sur « toxicomanie », « drogue », « toxicomane », etc.

Les opérateurs

* Troncature

*Wallonne pour *Handicap*

Rechercher

Recherche d'une expression dont on ne se souvient pas

hospi*

Rechercher

Recherche de tous les mots commençant par « hospi... »

Les opérateurs

« XX..XX »

colloque ehealth « 2000..2009 »

Rechercher

Recherche de « colloque ehealth » dans l'intervalle 2000-2009

Les opérateurs

Allintitle:

allintitle: bruxelles promotion santé

Rechercher

Recherche dans le titre de la page

Allinanchor:

allinanchor: santé enfant

Rechercher

Recherche de « santé enfant » dans les hyperliens