

Les astuces pour la recherche avancée sur Google

A moins d'avoir vécu loin, très loin d'un ordinateur durant les 10 dernières années, vous ne pouvez pas ignorer l'existence de Google, le moteur de recherche le plus célèbre de l'Internet. Une bonne partie d'entre vous l'utilise probablement tous les jours.

Il serait par ailleurs très intéressant de comptabiliser le nombre de minutes que passe un internaute à effectuer des recherches via ce moteur en une journée. Nous pourrions ainsi effectuer un petit calcul de l'efficacité, non pas de Google, mais de l'utilisateur. Car tout le problème est là : comment rendre efficace une requête sur un moteur tel que Google ? Comment accélérer nos recherches parmi les dizaines de milliards de pages que recense le géant de Mountain View ? Finalement, vous qui utilisez Google tous les jours, êtes-vous sûrs de le connaître en détail ?

Les réponses se trouvent tout d'abord dans la compréhension de ce qu'est un moteur de recherche, mais aussi dans une utilisation avancée de Google. Pour accéder à celle-ci, vous pourrez recourir à l'interface dédiée que Google met à disposition de tous. Toutefois, nous verrons que cette solution limite quelque peu les possibilités offertes par le moteur de recherche, et que la connaissance de la syntaxe utilisée par Google vous permettra d'effectuer des requêtes beaucoup plus précises. Nous évoquerons également les services spécialisés, tels que Google Music ou Google Translate. Finalement, nous aborderons en quelques lignes les autres moteurs de recherche que sont Yahoo!, Ask et Live Search, afin d'apporter quelques éléments de comparaison avec Google.

Google, les robots, les index, le PageRank

En informatique (et pas seulement d'ailleurs), savoir se servir efficacement de quelque chose nécessite souvent de l'avoir préalablement compris ! Le moteur de recherche illustre bien cette idée, car si l'expérience peut apporter de bons réflexes, elle ne remplace pas la connaissance du fonctionnement de l'outil. Commençons donc par expliquer le principe du moteur de recherche en général, et celui de Google en particulier.

La valeur d'un moteur de recherche se juge sur deux critères principaux :

- sa capacité à rassembler une base de données la plus importante possible, c'est-à-dire à référencer un maximum de pages. Plus ce nombre est important, plus le moteur aura de choix pour vous apporter la réponse la plus adéquate à votre recherche.
- la pertinence des résultats de la recherche, c'est-à-dire l'adéquation entre les pages proposées par le moteur et la recherche formulée. Cette pertinence va dépendre de deux facteurs : la façon dont le moteur gère la syntaxe des requêtes (c'est-à-dire la façon dont on pose la question au moteur : un ou plusieurs mots, une expression...) et l'ordre d'affichage des résultats, puisque les liens les plus pertinents doivent être accessibles rapidement (dans les 20 premiers résultats par exemple).

Trois étapes caractérisent le fonctionnement d'un moteur de recherche : l'exploration, l'indexation et la recherche forment l'arsenal technique créé pour répondre aux problématiques posées juste avant.

- **L'exploration** (ou *crawl*) :



Des robots d'indexation parcourent le Web et suivent les liens trouvés récursivement. Les GoogleBots (les robots de Google), implantés dans les milliers de serveurs de la compagnie, visitent ainsi un maximum de pages Web, en analysent la structure et suivent les liens vers les autres pages Web pour construire une carte la plus complète possible du réseau mondial.

- **L'indexation** :



Au cours de cette exploration, le moteur va extraire les mots considérés comme significatifs (presque tous) correspondant à chaque ressource. Ceux-ci sont alors enregistrés dans une base de données. Les termes non significatifs s'appellent des mots vides. Ce premier critère de jugement distingue déjà les moteurs de recherche entre eux, même si ce n'est pas ici que la différence est la plus marquée.

- **La réponse aux requêtes** :

Ce sont les deux étapes précédentes qui permettent à Google de vous apporter une réponse. Mais celle-ci doit être la plus pertinente possible. C'est ici qu'intervient un algorithme, qui va donner un poids (et même une valeur) aux différentes pages. Le but est de pouvoir présenter les résultats par ordre de pertinence vis-à-vis de votre recherche. C'est cet algorithme qui constitue la plus grande différence entre les divers moteurs de recherche.

Pour établir ce degré de pertinence, on peut se baser sur plusieurs critères différents :

- le pourcentage, dans une page, de mots correspondants à votre recherche,
- le nombre de fois qu'un de vos mots-clés apparaît dans la page,
- la présence du mot demandé dans le titre du document,
- sa présence dans l'adresse de la page,
- sa présence dans le texte de la page, en considérant sa place dans la page, ses attributs (taille des caractères, gras, souligné),
- sa présence dans certains champs ajoutés par le concepteur dans la partie invisible de la page (balises « Meta »). Ce critère, utilisé de façon abusive par certains webmasters, a vu son importance nettement amoindrie.



L'algorithme de Google comporte une particularité supplémentaire, celle de faire appel au *PageRank*, sorte de formule de pertinence qui a fait le succès de ce moteur de recherche. Sans entrer dans les détails que pourrait comporter un article sur le référencement, nous pouvons résumer le principe du *PageRank* ainsi : Google évalue le degré de pertinence d'une page en fonction du nombre d'autres pages qui redirigent vers elle. La qualité d'une page est alors mise en avant, et Google donne une note de 0 à 10 pour la page évaluée. Plus la page aura une note élevée, plus elle aura de chance de figurer en haut de liste lors de l'une de vos recherches. C'est ce principe qui a permis à Larry Page et Sergei Brin (les deux concepteurs de Google) de proposer un moteur qui se démarquait de ceux qui existaient à la fin des années 90.

Quelques conseils et astuces pour vos recherches

Après cette brève introduction sur le fonctionnement d'un moteur de recherche, intéressons-nous plus précisément au cas Google. Quelles sont les spécificités propres à ce moteur de recherche ?

Tout d'abord, Google n'utilise pas la lemmatisation, c'est-à-dire la réduction des mots au masculin et/ou au singulier, à l'infinitif, etc. Par exemple, le terme « innocentes » ne sera pas interprété par Google comme « innocent », et « partîmes » ne deviendra pas « partir ». De même, la requête *cheva** ne vous donnera pas de résultat sur chevalerie ou chevaux, mais uniquement sur cheva suivi de l'astérisque. Sachez aussi que Google ne tient pas compte de la « casse » des mots, c'est-à-dire qu'il ne différencie pas les majuscules des minuscules, mais il tient compte des accents ou autres signes diacritiques (cédille, tilde espagnol, umlaut allemand, etc.). Toutefois, l'utilisation du symbole « + » placé juste devant le mot-clé peut forcer une recherche sur une orthographe précise.



[La recherche cheva* n'amène pas de pages contenant le mot chevalier par exemple](#)

Autre information à prendre en compte : Google ignore les chaînes de caractères dont le poids sémantique est trop faible, comme le, la, les, du, avec, vous, etc., ou des mots spécialisés tels que « http » et « .com », qui ne sont pas pris en compte dans la recherche. Les lettres ou les chiffres d'un seul caractère sont également ignorés. Google part du principe que ces éléments ne jouent que rarement un rôle intéressant dans les recherches, et considère que la prise en compte de ces termes peut ralentir notablement le processus de recherche. Une fois de plus, vous avez la possibilité de passer outre ces automatismes de Google en ajoutant un « + » juste devant vos mots-clés (par exemple, une requête comme « +le comte +de Fourques » prendra en compte tous les termes).

Vous le savez par expérience ou grâce aux éléments que nous venons de vous apporter, la langue dite « naturelle » (ou encore « parlée ») n'est pas la plus adaptée à la recherche via Google. Entrer toute une phrase dans le moteur de recherche ne vous fournira que rarement le résultat attendu, mis à part peut-être pour les messages d'erreur en informatique. De plus, le temps de traitement de votre requête sera significativement plus important si vous entrez une phrase plutôt que les mots-clés qui la composent. Préférez donc *résultats ligue 1 11 novembre* à la phrase *Quels sont les résultats du championnat de France de Ligue 1 du dimanche 11 novembre ?*



La recherche en langage naturel n'amène pas de réponses satisfaisantes la plupart du temps

Dernier conseil enfin : le nombre de mots que vous entrez a une réelle importance. S'il est certain que plus votre requête comportera de mots, plus son traitement sera long, il est aussi évident que vous la rendrez également plus précise si vous utilisez plusieurs mots. Il y a donc un compromis à trouver, et il vous faut pour cela choisir avec soin les mots que vous utiliserez, ce qui reste la partie la plus importante de votre recherche.

La recherche avancée : explications

Vous connaissez tous la page d'accueil de ce site, qui vous sert probablement tous les jours à effectuer des recherches diverses et variées. Mais connaissez-vous l'interface de la recherche avancée sur Google ? Nous vous la présentons ici, d'une part en évoquant une à une ses possibilités, d'autre part en présentant en parallèle le « langage » utilisé par Google.

- [La page de recherche avancée](#), que vous pouvez trouver à droite de la barre de recherche Google.



Le module de recherche avancée de Google en images

De quoi est composé ce module de recherche avancée ? De conseils de recherche tout d'abord, situés en haut à droite de la page. Vous y trouverez principalement des liens et peu d'explications précises. Passons au premier bloc, dont le fond est bleu. Vous pouvez entrer ici des requêtes plus précises que sur le module de recherche traditionnel :

- « tous les mots suivants » : ET (ou +)

Ce champ sert à placer plusieurs mots que vous tenez absolument à voir figurer dans la réponse. L'ordre des mots importe peu. Cette recherche équivaut à l'utilisation de l'opérateur booléen ET (ou du symbole +), qui va forcer le moteur à ne prendre en compte que les pages où tous les termes précisés apparaissent.

- Exemple : la question *clubic +photo +mosaïque* vous donnera un résultat différent de la requête *clubic photo mosaïque*.
- « au moins un des mots suivants » : OR (ou |)

Vous pouvez ici préciser plusieurs mots synonymes par exemple, puisque le moteur de recherche vous donnera en réponse les pages qui contiennent au moins un des mots spécifiés et pas forcément tous. Cela est utile lors d'une recherche très large sur un thème général.

- Exemple : une recherche sur *crocodile* | *caïman* | *alligator* dans le moteur de recherche classique vous fournira tout ce que vous voulez savoir sur les différentes espèces de l'ordre des crocodyliens.



[Une requête avec l'opérateur OU permet d'agrandir son champ de recherche](#)

- « aucun des mots suivants » : NOT (ou -)

Cette case est très utile, car elle va vous permettre d'exclure un mot-clé. En effet, il existe de nombreux cas où la recherche d'un mot-clé vous amène systématiquement vers une association. Exclure un de mot de l'association permet alors d'effectuer une recherche sans interférence.

- Exemple : si vous voulez des informations sur la ville de Marseille, mais pas sur son équipe de football, tapez *Marseille -olympique* dans le moteur de recherche classique.
- « cette expression exacte » : « »

Les guillemets servent à forcer un terme ou une expression, comme le symbole « + », mais la notion d'ordre est ajoutée ici. Ainsi si le « + » forçait le moteur de recherche à vous fournir des pages contenant les mots-clés, l'utilisation des guillemets forme une requête qui aura pour réponse les pages contenant les termes dans un ordre précis.

- Exemple : « *clubic photo mosaïque* » est différent de *clubic +photo +mosaïque*.



Si vous cherchez les paroles d'une chanson par exemple, la recherche exacte pourra vous être utile

Notez également que Google privilégie les pages dans lesquelles vos termes de recherche apparaissent aussi près que possible les uns des autres. Enfin, prenez garde à bien utiliser les symboles « + », « - », « | » en les collant au mot qui les suit.

- Exemple : *clubic +photo +mosaïque* sera une requête précise et interprétée par Google, alors que *clubic + photo + mosaïque* équivaut ni plus ni moins qu'à « clubic photo mosaïque ».

Enfin, en haut à droite de cette zone, vous pouvez régler le nombre de réponses par page (de 10 à 100 résultats). Sous ce rectangle bleu, il est possible de préciser davantage sa recherche en choisissant :

- la langue dans laquelle est écrite la page,
- une région dans laquelle la page est disponible,
- un format de fichier, si vous souhaitez ne rechercher par exemple que des documents .PDF ou .DOC, vous pouvez le préciser dans la case prévue à cet effet,
- la date à laquelle la page Web a été découverte par le moteur, ce qui ne correspond pas tout à fait à l'heure d'actualisation de la page, mais s'en approche si le site est question possède un fort taux d'actualisation,
- l'emplacement des mots-clé précisés : ils peuvent être contenus dans le titre, le corps, l'adresse ou les liens de la page,
- d'inclure ou d'exclure certains domaines comme .org, .com ou .fr par exemple,
- les droits d'utilisation des documents affichés : vous pouvez ainsi diriger votre recherche vers des documents qui sont « libres de droits et d'utilisation ou de distribution » ou « libres de droits et d'utilisation ou de distribution, y compris à des fins commerciales » par exemple.

De l'utilisation avancée de la syntaxe

Nous l'avons vu avec les fonctionnalités présentes dans le rectangle bleu du module de recherche avancée de Google, chaque commande possède un équivalent dans le moteur de recherche classique. Ces fonctions sont activées par une syntaxe précise, comme l'utilisation des symboles « + », « - » ou « | ». Les autres possibilités offertes par la recherche avancée possèdent-elles aussi leurs pendants en commande en ligne, qui peuvent se révéler plus puissants encore :

- la requête « **site:** » :

Placée devant le nom d'un site (clubic.com à tout hasard), cette commande permet d'afficher des résultats contenus uniquement par le site de votre choix.

- Exemple: la recherche *crysis site:jeuxvideo.fr* vous donnera toutes les pages de Jeux Vidéo.fr contenant la chaîne de caractère *crysis*. Vous pouvez également placer derrière la commande site: un domaine précis, pour limiter vos recherches aux pages françaises (.fr, par exemple).



[La recherche « site » permet de cantonner les réponses aux pages d'un seul site](#)

- la requête « **filetype:** » :

C'est le pendant à la fonctionnalité du moteur de recherche avancée. L'intérêt de l'utilisation de la commande en ligne est la possibilité de choisir n'importe quel type de fichier, et non pas seulement ceux listés par Google dans leur interface dédiée.

- Exemple : lancez une recherche *compte rendu chimie filetype:doc* pour voir si le Net ne vous propose pas des pistes pour vos travaux pratiques de chimie.

Lorsque nous évoquons les possibilités offertes par le moteur de recherche avancée, nous vous avons présenté la fonctionnalité de recherche par emplacement dans la page. La syntaxe du moteur de recherche de Google comporte également cette possibilité :

- La requête « **intext:** » permet de restreindre la recherche du mot-clé au corps du texte (entre les balises de la page).

- Vous pouvez également demander à Google une recherche « **allintext:** », dont les réponses devront contenir les mots proposés dans le corps de la page.
- Exemple : si vous tapez *allintext:geek vacances mer*, les trois mots choisis seront inclus dans le texte des pages que Google vous indiquera.



Trouver rapidement un document dont vous connaissez le contenu grâce à la commande allintext:

- De même, « **intitle:** » suivie immédiatement du mot-clé permet de spécifier à Google que le terme recherché doit se trouver dans le titre de la page, et plus précisément dans la balise <TITLE> de la page HTML.
- « **allintitle:** » permet d'effectuer cette même requête, mais avec plusieurs mots,
- Exemple : *allintitle:test 8800 nvidia* affiche toutes les pages dont les titres contiennent ces trois mots.



[La preuve en image...](#)

- Il y a aussi les commandes « **inurl:** » et « **allinurl:** », qui permettent de préciser que le (les) mot(s)-clé doivent se trouver dans l'adresse Internet de la page (son URL).
- Enfin, Google propose également d'effectuer une recherche spécifique sur les liens que contient une page ; ce sont les commandes « **inanchor:** » et « **allinanchor:** » qui se chargent de cette fonctionnalité, pour un ou plusieurs mots respectivement.

Si nous évoquions plus haut le signe « * » en précisant que Google ne comprenait pas les requêtes du type *cheva** comme étant une recherche sur tous les mots commençants par *cheva*, l'astérisque conserve toutefois un intérêt dans la syntaxe du moteur de recherche. En effet, ce symbole peut être utilisé pour remplacer n'importe quel mot :

- Exemple : effectuer une recherche *clubic * article* vous mènera vers les pages qui contiennent les termes *clubic* et *articles* et un autre mot (magazine, informatique, etc).

Enfin, notez que le symbole « ~ » vous permettra, s'il est placé devant un mot, d'obtenir les pages contenant ce mot, mais également ses synonymes :

- Exemple : la requête *~crocodile* vous donnera des réponses dont le mot-clé est *alligator*.

Bien évidemment, ces opérateurs peuvent être combinés entre eux (et avec des mots) dans une même requête, ce qui peut permettre des recherches très précises et puissantes.

Quelques commandes spécifiques

Google possède également quelques commandes plus spécifiques encore. Parmi elles, on peut par exemple citer la commande « **stocks:** » qui, suivie des trois lettres définissant un titre en bourse, vous donne en réponse un graphique du cours de cette action ainsi que la valeur du titre en temps réel. Cette fonction ne fonctionne que sous Google.com malheureusement, tout comme la commande « **info:** » qui, si elle est suivie d'un site (*clubic.com* au hasard), permet de visualiser la page qui se trouve dans le cache de Google et de trouver les indications concernant les pages similaires à cette page (la fonction « **related:** » le permet aussi), celles pointant vers *clubic.com*, celles du site en lui-même et celles contenant les termes *clubic.com*. Cette dernière fonction est également accessible via les commandes « **inlink:** » ou « **allinlink:** ».





La commande info: apporte des renseignements intéressants sur un site particulier ; dommage qu'elle ne soit disponible qu'en anglais

Notez également l'existence de la commande « **define:** », qui permet de vérifier la définition d'un mot.

- Exemple : tapez *define:HTML* et vous obtiendrez une page de différentes définitions du terme HTML.

Enfin dans les Google News, la commande « **source:** » fonctionne de la même façon que la commande « **site:** », mais se limite à ce type de pages. Les commandes « **author:** », « **insubject:** » et « **location:** » sont également valides dans ce contexte et servent à restreindre les recherches par auteur ou par lieu de publication, la commande « **insubject:** » permettant d'effectuer une recherche de mots-clé contenus dans le sujet de la news. Cette fois, la fonction est disponible pour news.google.fr.

De l'utilisation malveillante de cette syntaxe

Toutes les commandes que nous vous avons présentées jusque-là peuvent évidemment être associées pour former des requêtes particulièrement puissantes, au point de mettre en péril la confidentialité de certaines données ayant trait à la vie privée de nombre d'internautes. En effet, il devient simple, avec les modestes connaissances apportées par cet article, de se lancer à la recherche de liste d'adresses mails, de fichiers de contact MSN ou même de fichiers de configuration de serveur FTP qui peuvent contenir des mots de passe. La recherche de fichiers musicaux est également assez simple, de même que de fichiers vidéo ou de photos. Une commande telle que *-inurl:(html|html|php) intitle:"index of" + "last modified" + "parent directory" + description + size +(jpeg|bmp|png) "vacances"* peut se révéler très indiscreète par exemple...

Lors de nos recherches, les résultats obtenus avec de telles requêtes nous ont quelque peu surpris, car nous ne nous attendions pas à trouver autant de serveurs non protégés. Prenez donc garde à sécuriser les accès qui peuvent mener à vos données, car vous savez maintenant qu'elles sont accessibles via un simple moteur de recherche. Si vous souhaitez en apprendre plus sur ce type de recherche (afin de vous en protéger, évidemment), vous pouvez visiter [ce site](#).

Les services spécifiques de Google

Il existe encore d'autres moyens d'effectuer des recherches avec Google, ce dernier proposant toute une gamme de services spécifiques. En voici une liste non exhaustive, car le géant de Mountain View n'a de cesse de la faire grandir à un rythme élevé.

Adresse :	Fonction de la page :
http://blogsearch.google.fr/	Vous pouvez ici effectuer une recherche spécifiques aux nombreux blogs qui existent de par le monde.
http://books.google.fr/	Cette page sert à trouver un livre plus rapidement qu'en passant par une recherche classique, puisque toutes les réponses proposées ici seront des livres.
http://www.google.fr/codesearch	Une des dernières fonctionnalités disponibles sur Google. Elle permet aux développeurs d'avoir accès aux codes libres publiés sur le Net.
http://directory.google.fr/	Google rangé par catégories. Vous pouvez ici restreindre vos recherches à des thèmes comme la littérature, le tourisme, l'emploi ou l'astronomie. Des thèmes assez larges, certes, mais qui constituent déjà un premier filtre.
http://images.google.fr/	Le fameux moteur de recherche d'images de Google, que l'on ne présente plus.
http://local.google.fr/	Le service local.google (plus communément appelé Google Maps) n'a pas la prétention de remplacer les pages jaunes, mais propose tout de même une interface conviviale, agrémentée d'une carte précise, actualisée au fur et à mesure de votre recherche.
http://www.google.fr/movies?q	Un module de recherche intéressant sur les films, qui fournit les critiques de différents critiques ainsi que les cinémas qui passent ce film près de chez vous. Seul hic, et de taille : cette fonction n'est pas disponible en France...
http://news.google.fr/	Voilà un service qui fonctionne correctement dans l'hexagone, et qui permet d'afficher les dernières brèves sur une actualité recoupant la politique, l'informatique ou le sport.
http://scholar.google.fr	Cette page propose des résultats qui sont orientés vers les recherches scolaires. Ici, pas de risque de voir apparaître une page au contenu trop explicite...
http://www.google.fr/translate_t	L'une des possibilités les plus utilisées de Google : le module de traduction en ligne. Entrez un mot, choisissez sa langue, et enfin la langue dans laquelle vous souhaitez voir traduire ce mot, et lancez le processus !

Notez de plus que le moteur de recherche Google peut également servir de calculatrice.

- Exemple : essayez la ligne suivante sous Google et admirez le résultat : $(2*i-5*789+2+2*(9/4-8/6))/7...$ Google comprend même les nombres complexes ! Il connaît également pi, et le symbole « ^ » est utilisé pour les puissances, comme sur une calculatrice.

Une fonction de convertisseur est également disponible :

- Exemple : la requête *3 euros en dollars* ou la question *5 dollars en monnaie japonaise* fonctionnent parfaitement.

Enfin, Google est capable d'afficher des informations sur la météo si vous lui précisez une requête adéquate :

- Exemple : tapez *meteo lyon* (ou toute autre ville bien sûr !) pour avoir immédiatement accès à des données telles que les prévisions sur 4 jours, la température, la vitesse et la direction du vent et le taux d'humidité.

Et les autres moteurs de recherche ?

Google est évidemment le moteur le plus utilisé dans le monde. Mais il en existe de nombreux autres qui, s'ils ne sont peut-être pas aussi connus que Google (les parts de marché en France de Yahoo! - 3,18% en octobre 2007, source Xiti - ou Live Search - 2,63% - le prouvent), ils n'en restent pas moins intéressants dans leurs spécificités.

Il existe principalement 6 technologies différentes sur lesquelles sont basés quasiment tous les moteurs de recherche francophones : Google, Yahoo! Search Technology, Live Search, Exalead, Ask et Voila. Sans évoquer Voila (dont l'influence se limite au portail d'Orange) ni Exalead (dont nous vous parlerons en conclusion), nous pouvons toutefois porter notre attention sur les Yahoo!, Live Search et Ask, qui sont parmi les plus utilisés en France après le géant Google.

Un fonctionnement comparable à celui de Google ?

Le but n'est pas ici de dire si Google est meilleur ou pire que ces autres moteurs, mais de montrer les quelques différences qui existent dans leurs façons de fonctionner. Dire si Google est plus pertinent que Yahoo! demande en effet des études poussées, que de nombreux groupes ont tenté de réaliser, et dont les résultats ne concordent jamais puisque les protocoles sont à chaque fois différents.

Quelles sont donc les similitudes (et les différences) entre Google, Yahoo!, Live Search, et Ask ? Au niveau du principe de base - à savoir la récupération de données par des robots -, tous possèdent un fonctionnement similaire. Les différences principales vont se situer au niveau du classement des pages et des fonctions de recherche.

Yahoo! Search Technology



Jusqu'en février 2004, Yahoo! utilisait la technologie de Google. Mais suite au rachat d'Inktomi et d'Altavista, les équipes de recherche de Yahoo! ont développé leur propre moteur, qui leur offre l'indépendance. Le Yahoo! Slurp (le robot de Yahoo!, qui reprend le nom de celui d'Inktomi) indexe les pages en suivant les liens HREF (et non les liens RC, afin de ne pas indexer des frames seules ou sans contenu). Notez que si Google indexe les 101 premiers kilobits d'une page, Yahoo! en recense 500.

En ce qui concerne l'algorithme de classement, Yahoo! reste très mystérieux à ce propos et il est très difficile de trouver une quelconque information à ce sujet. Google l'est tout autant, mais la publication des travaux de recherches de Larry Page et Sergey Brin a permis d'en apprendre un peu sur le fonctionnement de Google, qui a bien évolué depuis.

Yahoo! possède un moteur de recherche incluant les pages Web, les images, les vidéos, l'actualité et le shopping. Les « questions-réponses » et le Guide Web complètent cette liste. La recherche s'effectue avec l'opérateur « ET » par défaut, comme sur Google. Elle inclue les pages en cache, mais ne comprend pas quelques fonctions avancées comme la troncation et nécessite la présence du « http:// » pour valider un lien. La recherche avancée permet d'utiliser les opérateurs « tous ces mots », « un de ces mots » ou « aucun de ces mots ». La requête « phrase exacte » (en utilisant les guillemets) est également possible, toutes ces fonctionnalités pouvant bien sûr être combinées. L'astérisque a la même utilité que sur Google, et tout comme ce dernier, Yahoo! n'est pas sensible à la casse. Enfin, une recherche par date (« il y a plus de 3 mois », « plus de 6 mois », « plus d'un an ») est prévue, ainsi que la recherche dans la page ou celle d'un type de fichier (via la commande *originurlextension*., un peu plus longue que la commande *filetype*:).

Live Search



Successeur de MSN Search et intégré au portail Live.com, ce moteur de recherche est né en septembre 2006. Il permet une recherche de pages Web, d'actualités, d'images, de vidéos. Tout comme Yahoo!, il possède un module de questions-réponses. En revanche, ses fonctions au niveau de la recherche sont limitées : pas de troncation, requête limitée à 10 mots... même le module de recherche avancée, présent partout ailleurs sur la page principale, n'est ici disponible qu'après une première recherche.

Le « ET » est, ici aussi, inclu par défaut entre vos mots-clé, mais vous pouvez tout à fait utiliser les autres opérateurs : « OR », « NOT » ou « AND NOT ». Les guillemets possèdent la même fonction que sur Google ou Yahoo!, et la recherche ne prend pas en compte les majuscules. Vous pouvez, grâce à la recherche avancée, limiter vos recherches à un site, un domaine, un type de fichier. Les mots comme « le » ou « de » ne sont pas pris en compte, mais peuvent être reconsidérés par le moteur s'ils sont précédés d'un « + », tout comme sur Yahoo! ou Google.

Ask

Ask (anciennement connu comme Ask Jeeves) utilise une base de données développée à l'origine par Teoma, qui reste bien moins importante que celles de Google ou Yahoo!. Ce n'est pas là sa seule faiblesse, puisqu'Ask ne permet pas la recherche des pages en cache et ne propose qu'une seule réponse par site. Les opérateurs booléens (comme le « OR » ou le « AND », considéré par défaut) fonctionnent, mais leur association n'est pas aussi simple que sur les autres moteurs. La recherche d'une phrase exacte (via les guillemets) fonctionne, elle, parfaitement, et tout comme pour Google et consorts, Ask est insensible à la casse.

Ask est l'un des rares moteurs de recherche à communiquer un peu sur son fonctionnement, et nous pouvons apprendre sur leur site que l'algorithme « ExpertRank » demande de nombreuses analyses supplémentaires, soi-disant non pratiquées par les autres moteurs de recherche. Ask analyserait le Web dans sa configuration réelle – en fonction des communautés consacrées à des sujets spécifiques -, n'incluant pas nécessairement dans son index tous les sites visités.



[La page d'accueil de Ask, très épurée...](#)

Enfin, nous n'avons pas évoqué le cas des moteurs de recherche spécifiques comme Youtube ou DailyMotion, tant la tendance actuelle semble être au rachat et à l'introduction de ces moteurs dans ceux déjà existants.

Conclusion

En septembre dernier, Google franchissait pour la première fois la barre des 90% de parts de visites sur les sites Web audités par XiTi, soit près de 87 00 sites francophones. Ce qui laisse peu de place aux concurrents que sont Yahoo!, Live Search ou Ask, pour ne citer qu'eux. Et même si la firme de Mountain View développe probablement encore son moteur actuellement, ce n'est jamais très bon pour une entreprise de ne pas avoir de concurrence, particulièrement lorsqu'il s'agit d'innover. Et c'est bien là que se trouve le problème actuel des moteurs de recherche, l'innovation.

Car si des services apparaissent aujourd'hui avec une fréquence élevée, le principe de base des moteurs de recherches est le même depuis des années. En effet, les principaux moteurs de recherche ne comprennent pour l'instant pas ce qu'ils « lisent » sur les pages qu'ils indexent ou ce qu'ils affichent dans les résultats de recherche. C'est une linguistique assez basique qui est appliquée dans les algorithmes et le contexte de la page n'est pas compris d'un point de vue sémantique. Ces algorithmes sont pourtant évolués, puisqu'ils permettent de classer les pages avec une certaine réussite, mais le fonctionnement actuel possède des limites que chacun peut apprécier chaque jour. C'est donc vers la sémantique appliquée qu'il faut probablement se tourner pour capter les enjeux futurs de la recherche d'informations.



En effet, un moteur qui analyserait un contenu sémantiquement pourrait aller beaucoup plus loin. Il s'agirait d'associer des mots d'un même champ lexical, et une recherche sur les fleurs par exemple devrait vous conduire vers les abeilles ou les arbres, plus généralement vers la campagne ou la nature. Nous pourrions imaginer pouvoir doser le degré d'ouverture de la recherche, pourquoi pas. Plusieurs moteurs de recherche de ce type, tourné vers la sémantique appliquée, existent à l'heure actuelle : [Exalead](#) et plus encore [Kartoo](#) en sont de bons exemples et semblent être des moteurs très prometteurs.

Exalead, moteur de recherche français, propose en effet un fonctionnement très différent de celui de Google. Bien que basé sur le même annuaire que Google, à savoir l'openDirectory, Exalead en a une tout autre utilisation : politique pour limiter l'usage de la bande passante du site scanné, capture d'écran des sites visités, prise en compte des meta-tags... Mais les différences principales se situent à l'indexation, car si Exalead utilise la catégorisation (comme Ask par exemple), c'est bien l'utilisation de la linguistique, en général, et de la lemmatisation en particulier, qui donne à ce moteur sa particularité. Pour chaque mot ou groupe de mots ayant le même lemme, l'indexeur essaie d'y associer un ou plusieurs documents, ce qui permet d'agrandir la recherche d'autant de termes qui sont associés à votre mot-clé. S'il existe un réel avantage en termes d'efficacité de recherche, ce type de moteur n'est pas encore complètement prêt pour une utilisation grand public, mais cette alternative pourrait sonner la fin du référencement.