

Net recherche : Comment choisir ses mots-clés ? Comment effectuer un "sourcing" de qualité ?

[Retour au sommaire de la lettre](#)

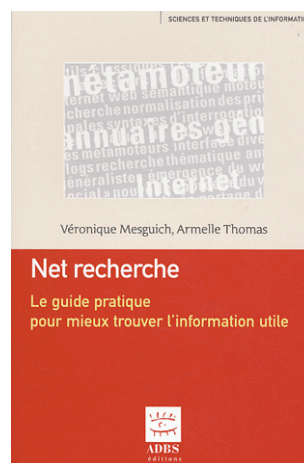
Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Nous vous proposons dans cet article deux "fiches pratiques" extraites du livre "Net recherche" qui vient de sortir aux éditions ADBS. La première nous rappelle les principes de base permettant de bien choisir les mots clés qui vont constituer une requête sur un moteur de recherche, et la seconde nous indique quelques pistes pour identifier des sources "de confiance" sur le Web. Indispensable.....

Le livre "**Net recherche 2009 : le guide pratique pour mieux trouver l'information utile et surveiller le web**" vient de paraître aux éditions ADBS (<http://www.adbs.fr/net-recherche-2009-le-guide-pratique-pour-mieux-trouver-l-information-utile-et-surveiller-le-web-61812.htm>). En voici une courte présentation proposée par l'éditeur :

Sous l'apparente facilité d'utilisation des moteurs de recherche se cache en effet une réalité complexe, et le secret de la réussite d'une recherche ou d'une veille passe autant par la maîtrise des aspects techniques que par la capacité à évaluer et sélectionner les sources pertinentes.

En cette troisième édition profondément renouvelée, développée et mise à jour, Net recherche vise à offrir à toute personne amenée à effectuer des recherches sur Internet un panorama des outils et méthodes existant à ce jour, en intégrant les dispositifs qui permettent de surveiller le web à moindre coût. Cet ouvrage s'adresse notamment aux professionnels de l'information (documentalistes, bibliothécaires, veilleurs), aux enseignants, chercheurs et étudiants, et à tous les autres "travailleurs du savoir" confrontés à la complexité croissante et à l'inflation de l'information en ligne. Ils y découvriront des conseils méthodologiques mais aussi nombre de "trucs et astuces" destinés à optimiser le processus de recherche ou de veille, des informations précises sur les évolutions en cours, ainsi que des présentations d'outils et des listes d'adresses utiles.



Les auteurs ont demandé à Olivier Andrieu, éditeur du site Abondance.com d'en signer la préface. En accord avec les auteurs et l'éditeur, nous vous présenterons, pendant quelques mois, quelques "fiches pratiques" présentes dans ce livre. Nous espérons qu'elles vous permettront de mieux chercher l'information sur le Web...

Fiche 1. Comment choisir ses mots-clés ?

Quand ?

La sélection des mots-clés s'effectue après le choix d'une stratégie de recherche. En effet, le choix sera fondamentalement différent selon que l'on cherche un portail thématique, une source susceptible de fournir l'information, ou bien une information précise immédiatement.

Pour simplifier, disons que, dans le premier cas, les mots-clés seront "le plus large possible", dans le second cas, ils seront "le plus précis possible".

Le choix des mots-clés peut résulter également d'un processus itératif : on peut, au cours d'une recherche, ajouter des mots-clés que l'on aura repérés au sein des réponses trouvées par le moteur, ou au contraire en exclure pour mieux préciser sa requête.

Quel type de mots-clés ?

On a vu que les moteurs de recherche indexent indifféremment la quasi-totalité des mots d'une page web : noms communs, noms propres, verbes, adjectifs, adverbes, etc. Généralement, seuls les articles et pronoms sont ignorés. Certains termes peuvent se révéler utiles pour affiner ou préciser une recherche : par exemple, dans le cas d'une requête sur un marché particulier, ne pas hésiter à employer des termes complémentaires comme "parts de marché", "statistiques", etc.

Pour une recherche précise, on peut utiliser les termes que l'on s'attend à trouver dans le document (voir page 182).

Dans quelles langues ?

Il faudra bien sûr traduire ses mots-clés dans les langues pertinentes eu égard au besoin identifié et donc à la recherche à effectuer. [...]

Un ou plusieurs mots-clés ?

On procédera par étape pour affiner éventuellement sa recherche à l'aide de plusieurs mots-clés. Si, avec un seul mot-clé, le nombre de résultats est faible (une centaine ou moins) mais les résultats pertinents, inutile de préciser davantage.

En règle générale, afin de mieux cibler les résultats, il ne faut pas hésiter à employer plusieurs mots-clés (au moins deux ou trois), les plus précis possible (de préférence les mots que l'on s'attend à trouver dans les résultats). Cependant, si la terminologie ou l'association terminologique est très spécifique, on peut utiliser un seul mot-clé ou expression.

Selon une étude réalisée en 2005 par la société IDfr, les internautes utilisent en moyenne des requêtes composées de "2,44 mots" sur les moteurs de recherche. On note environ 30 % de requêtes comportant un mot, et autant de requêtes à deux mots. Les requêtes à trois mots représentent 20 % et celles utilisant quatre termes 10 %.

Pour ou contre le SAUF ?

On peut isoler les mots-clés à exclure absolument car générateurs de bruit (opérateur SAUF ou signe -). Attention toutefois à ne pas aller trop vite, de peur de passer à côté de documents pertinents. Ainsi, si je cherche des informations sur les énergies alternatives autres que solaires, je peux être tenté de soumettre au moteur une équation du type énergies alternatives - solaires. Mais je n'obtiendrai pas les ressources qui abordent successivement toutes les énergies alternatives.

Majuscules, minuscules, accents ?

De façon générale, les moteurs sont insensibles à la casse des caractères et retourneront le même nombre de résultats pour python, PYTHON, ou Python. Il en va de même pour les accents : la majorité des moteurs traitent de manière identique les mots accentués ou non. Google, quoiqu'en dise l'aide en ligne, ne gère pas de manière claire cette question des accents et autres signes diacritiques : théoriquement insensible aux accents, le moteur donne toutefois des résultats différents avec ou sans, surtout s'il s'agit de mots peu fréquents, et lorsqu'ils figurent dans des expressions.

Troncatures ?

La troncature sur le web permet de remplacer plusieurs caractères sur la fin des mots, généralement avec le caractère *. La troncature à la fin d'un mot peut être limitée à un caractère ou illimitée. L'option de troncature, qui était disponible auprès de plusieurs moteurs comme AltaVista, a tendance à devenir rare. Les moteurs les plus utilisés par les internautes ne proposent pas cette option en tant que telle.

En revanche, Google (suivi par Yahoo! depuis février 2009) utilise – de façon assez peu explicite et sans intervention de l'internaute – une forme de troncature, qui consiste à élargir la recherche sur un terme à des termes considérés comme proches (par exemple ce terme au pluriel).

Sur un certain nombre de moteurs, on peut utiliser l'opérateur OR pour rechercher un terme à la fois au singulier ou au pluriel : l'équation de recherche *avion OR avions* permet de prendre en compte les deux orthographes.

Exalead est à ce jour l'un des rares moteurs à proposer l'option troncature, à droite.

Et les synonymes ?

Il est important d'explorer la terminologie du domaine de recherche pour repérer les synonymes et les mots sémantiquement proches. De façon générale, n'oublions pas que les premiers documents récupérés – les plus intéressants – permettent de valider, compléter ou revoir ses mots-clés.

Astuces pour identifier des synonymes ou mots associés

- Explorer les balises méta (keywords) de quelques documents pertinents.
- Utiliser un dictionnaire de synonymes, tel celui du laboratoire de linguistique du CNRS pour les termes en français (<http://www.crisco.unicaen.fr/Présentation-du-dictionnaire.htm>). Pour l'anglais, Kwmap est un outil visuel fort intéressant. À partir d'un mot-clé saisi dans le formulaire ad hoc, il génère deux axes : l'un pour les mots-clés "en relation" (par exemple "Formula 1" pour "car") et l'autre pour les "variations" possibles d'un mot (par exemple "car insurance" pour "car"). Les différents nœuds sont cliquables.
- Utiliser un moteur de recherche travaillant à partir de dictionnaires, encyclopédies, thésaurus tels que, pour les termes en anglais, www.thesaurus.com ou le métamoteur Surfswax en utilisant l'option Focus.

Pour la langue française, l'outil Exploredge (lancé en 2006, avec une version 2 en 2008) apporte une nouvelle voie d'exploitation des univers sémantiques de nos recherches mots-clés, grâce à des partenariats avec des professionnels d'un domaine.

Ce métamoteur assistant de recherche est disponible gratuitement sous la forme d'un *plugin* pour le navigateur Firefox : au moment du téléchargement, l'utilisateur choisit les domaines "dans lesquels il veut être assisté dans ses recherches" parmi l'environnement et l'énergie, les loisirs, la culture, un secteur d'activité particulier, l'entreprise, etc.

Une fois la requête effectuée (un système d'onglets permet de passer des résultats d'un moteur à un autre), une fenêtre Exploredge structure l'univers sémantique de la recherche qu'il est ensuite possible de relancer avec un ou plusieurs mots-clés.

Les internautes peuvent contribuer à enrichir les thésaurus en proposant de nouveaux termes et bénéficient alors d'une fiche de présentation.

Outils de suggestions de mots-clés

Les suggestions de mots-clés sont désormais monnaie courante. Elles peuvent intervenir en amont, les moteurs proposant des termes au fur et à mesure de la saisie par l'internaute [...]. Elles sont aussi générées de façon automatique par certains outils comme Exalead, Ask France ou Kartoo en fonction de leurs occurrences et co-occurrences dans la liste des résultats.

Tous ces outils se sont beaucoup développés au cours des derniers mois. Ils permettent, à partir d'un mot, d'obtenir soit des synonymes soit des expressions faisant partie de l'environnement sémantique du mot-clé de départ. Par exemple, le site Wikiwax, basé sur le texte intégral de Wikipédia, affiche au fur et à mesure de la frappe les mots-clés commençant par les lettres saisies.

L'un des générateurs de mots-clés le plus utilisé par les référenceurs, celui d'Overture, a fermé courant 2008, laissant la voie libre à celui de Google (<https://adwords.google.fr/select/KeywordToolExternal>), également très fréquenté, et désormais "le meilleur s'il est bien utilisé" d'après Olivier Andrieu (<http://docs.abondance.com/question120.html>).

L'américain WordTracker propose des expressions du mot-clé demandé, et pour chacune, le nombre de demandes quotidiennes du mot-clé.

Quant à l'outil Google Trends, apparu courant 2005, il permet d'avoir une idée du volume de requêtes pour un mot-clé sur une période ou une zone géographique donnée. On peut également comparer plusieurs termes.

Mots-clés ou tags ?

On l'a vu au chapitre 4, les tags ont depuis quelques années envahi le web. Ces mots-clés, avec lesquels les internautes étiquettent leurs productions ou favoris ainsi que les documents ou sites qu'ils souhaitent conserver, voire partager, peuvent constituer une alternative aux recherches via les traditionnels mots-clés. C'est une nouvelle manière de penser, d'indexer, totalement libre et ouverte, pour laquelle il faut abandonner au préalable tous ses réflexes de professionnel, si toutefois on en a !

Le secret réside alors dans sa capacité à se laisser guider sans a priori par les autres internautes et par les liens successifs proposés au net-chercheur. Notons bien sûr que la navigation par tags s'effectue uniquement sur les sites qui les utilisent (type Technorati, IceRocket, Delicious ou Yoono), mais ils deviennent nombreux dans le paysage web 2.0.

Fiche 2. Comment effectuer un "sourcing" de qualité ?

L'identification des sources d'information, ou "sourcing", reste une stratégie incontournable sur la Toile pour des besoins professionnels, et des techniques spécifiques peuvent faire gagner beaucoup de temps.

- Dans le cadre d'une recherche : on veut soit obtenir directement une information précise (stratégie privilégiant davantage les moteurs) soit identifier la source susceptible de contenir l'information souhaitée ("stratégie sources"). Sauf dans le cas d'une question-réponse rapide, les deux stratégies s'utilisent souvent conjointement.
- Dans le cadre de la constitution d'un bookmark : pour faciliter ultérieurement les recherches ponctuelles sur des thèmes récurrents, on veut également repérer non pas des informations précises en tant que telles, mais des sources où l'on pourra par la suite obtenir les réponses à des questions.
- Dans le cadre de la mise en œuvre d'une veille : on se situe dans un processus de surveillance continue de sources préalablement identifiées, puis validées comme pertinentes par rapport aux thèmes précis engagés par la veille.

Trouver des sites portails verticaux ou des sites fédérateurs

Les portails spécialisés, on l'a vu, constituent un bon point de départ pour une recherche thématique approfondie, dans la mesure où ils disposent de listes de liens, voire de répertoires complets, qu'il suffit alors d'exploiter. Encore faut-il pouvoir les identifier, ce qui n'est pas facile dans la mesure où il n'existe pas de véritable répertoire de portails. Plusieurs voies d'approche sont néanmoins possibles.

- Utiliser les répertoires généralistes comme l'Open Directory. Pour certaines thématiques, une sous-rubrique "annuaires" ou "directory" est disponible ; pour d'autres, une exploration sera nécessaire à partir des résultats d'une requête la plus large possible. Les liens proposés à la fin de nombre d'articles de Wikipédia peuvent parfois conduire à des sites portails, ou du moins des sites de référence en relation avec le sujet de l'article.
- Exploiter les répertoires d'outils de recherche et de portails verticaux ainsi que les répertoires professionnels et académiques vus auparavant (voir chapitre 3). Attention ! Ils ne sont jamais exhaustifs, et peu critiques pour la plupart.

- S'appuyer sur les sites des associations professionnelles, donnant le plus souvent les liens clés du secteur (sur les répertoires généralistes ou sur des répertoires spécialisés entreprises comme Indexa en France).

- Effectuer une recherche sur Google, qui privilégie la popularité dans son tri de pertinence, avec comme mot-clé sa thématique : il y a de fortes chances que, si un portail reconnu existe sur le sujet, il figure dans les deux premières pages de résultats.

- Lancer une recherche sur un moteur classique en utilisant des mots-clés comme *portail*, *liens*, etc., associés à la thématique souhaitée. Exemple : pour trouver un portail spécialisé en géographie, taper la requête suivante : *portail géographie*. S'il existe un portail dans ce domaine, il apparaîtra sans doute dans les premières réponses du moteur. Pour trouver des portails anglophones, on utilisera la même astuce en employant le terme *portal*.

- Pratiquer le "bouche à oreille" (y compris sous forme électronique avec les listes de discussion et forums) : l'information sur les bons sites circule...

- Compléter également ces portails spécialisés par des "univers Netvibes9" thématiques, lorsqu'ils existent bien entendu, et qui recensent eux-mêmes des flux provenant de sites *a priori* pertinents et de qualité.

Trouver des listes de liens (bookmarks, favoris) déjà constitués

Si les portails ou les répertoires issus d'associations et organismes reconnus sont repérables par les méthodes indiquées ci-dessus, la recherche est plus hasardeuse pour les pages de liens d'experts, de passionnés ou d'organismes plus confidentiels. Elles sont en effet plus complexes à localiser *via* les outils de recherche classiques...

- Recherche directe : il s'agit d'utiliser le mot-clé le plus significatif pour sa recherche (voir fiche "*Comment choisir ses mots-clés ?*") et d'y adjoindre un terme comme "*ressources*" ou "*liens*" ou "*bookmark*", ce qui donne, par exemple, sur Google "*santé animale*" *liens*. On peut aussi restreindre sur le titre et/ou jouer sur l'url avec des équations de recherche telles que "*santé animale*" *intitle:liens*, ou *intitle:liens OR inurl:liens "santé animale"*. Bien sûr, toutes les pages identifiées ne seront pas pertinentes, et beaucoup de pages utiles ne "sortiront" pas de cette manière.

- Utiliser des sites de "bookmarking social" ou partage de favoris, comme la plate-forme Delicious (voir chapitre 3). La recherche s'effectue *via* des tags correspondant à des mots-clés, ou encore à des personnes morales ou organisations. On pourra ainsi repérer des sites *a priori* de qualité parmi les favoris mis en ligne par une communauté d'internautes. Attention ! Les sites répertoriés dans ces plates-formes ne couvrent pas tous les secteurs, notamment les secteurs professionnels.

Pour effectuer un "sourcing" sur des sujets plutôt grand public, on peut utiliser également un service comme Alltop qui répertorie des sites "populaires" classés par ordre alphabétique ou par grands thèmes. Un renouveau des annuaires classiques ?

Trouver des sites "pointant" vers une source déjà connue

Cette stratégie s'appuie sur des sites intéressants déjà identifiés : on cherche alors les pages liées à une page, un site ou un domaine connu, c'est-à-dire qui ont un lien vers cette page ou ce site (backlinks en anglais), en partant du principe que si un site pointe sur le site spécialisé que nous connaissons, soit ce site traite du même sujet (la thématique peut être connexe ou plus large), soit il propose une liste de liens nous intéressant. Il faut alors utiliser la fonction "link:" sur les moteurs qui la supportent (comme Google, Yahoo! et Live Search).

On écrit, par exemple dans le formulaire de recherche de Google, *link:www.adbs.fr* (la mention *http* n'est pas nécessaire). Mais il n'est pas possible d'adjoindre un mot-clé à cette fonction pour trouver les pages liées à l'ADBS et contenant tel mot-clé. Il n'est pas non plus possible d'utiliser d'autres opérateurs dans la même équation de recherche. Attention ! Google restreint les liens obtenus aux sites les plus populaires.

On peut aussi utiliser le métamoteur Linkpopularity.com pour les trois moteurs majeurs Google, Yahoo! et Live Search : on compare ainsi le nombre de liens trouvés par ces trois moteurs, à partir d'un url.

Sur Yahoo!, on peut chercher les liens pointant sur la page d'accueil de l'ADBS : *link:http://www.adbs.fr* (la mention http est alors indispensable). Mais on peut aussi chercher les liens qui pointent vers un nom de domaine, et donc les pages d'un site ou d'un ensemble de sites du même domaine dans leur intégralité. On utilise alors la fonction "linkdomain:". Par exemple, la commande *linkdomain:agriculture.gouv.fr* cherche les pages ayant un lien vers une des pages du ou des sites du ministère de l'Agriculture. Mais on peut aussi combiner cette commande avec des mots-clés, ou avec d'autres opérateurs. Par exemple *linkdomain:agriculture.gouv.fr "santé animale"* cherche les pages contenant l'expression "santé animale" et pointant sur le domaine agriculture.gouv.fr.

À noter que la fonction "linkdomain:" est disponible également sur Live Search.

Yahoo! propose également une fonction sans équivalent sur les autres moteurs : *Yahoo! Site Explorer*, qui permet aux webmestres d'avoir des informations sur la façon dont un site web est indexé par le moteur.

Trouver des sites "similaires" à une source déjà connue

Cette stratégie est souvent payante et permet de compléter une première liste de ressources. Elle s'appuie également sur les sites intéressants déjà identifiés.

- Utilisation des répertoires comme l'Open Directory : le nom du site devient le mot-clé à utiliser. Il suffit alors de cliquer sur la (ou les) rubrique(s) intégrant ce site.
 - Utilisation des moteurs classiques : on choisit, à partir de la page de résultats, la fonction appelée "Related pages" ou "Sites similaires". Cette option est disponible dans plusieurs moteurs, notamment Google.
- Le logiciel de visualisation graphique Touchgraph pourra être ici d'une aide précieuse. Ce logiciel permet de visualiser sous forme de cartographie les sites "similaires" à ceux trouvés par le moteur Google, suite à une requête correspondant à un mot-clé ou à une adresse url. On bénéficie ainsi d'une vue d'ensemble, non seulement des réponses, mais aussi des sites considérés comme similaires.
- On peut utiliser le moteur Alexa, qui ajoute aux résultats de Google des informations sur la page visitée et indique parfois des sites "similaires" à la page (c'est-à-dire consultés conjointement).
 - Une technique fort utile consiste à choisir comme mots-clés les noms des deux ou trois sources déjà connues : il y a fort à parier que la page résultats citant déjà ces sites ou pages en fournira d'autres similaires.
 - On peut enfin choisir de surfer sur la vague des nouveaux outils de "navigation sociale" (voir chapitre 4).

Véronique Mesguich et Armelle Thomas

Auteurs du livre "Net recherche 2009 : le guide pratique pour mieux trouver l'information utile et surveiller le web". Extraits du chapitre 7.

Réagissez à cet article sur le blog des abonnés d'Abondance :
<http://abonnes.abondance.com/blogpro/2009/05/net-recherche-comment-choisir-ses-mots.html>