



Aujourd'hui, les outils de recherche foisonnent pour fouiller les entrailles du Web.

N'hésitez pas à les utiliser, l'un puis l'autre, pour effectuer une même recherche, et comparez la pertinence des résultats ! C'est, avec une formulation correcte de la question, encore et toujours la clé du succès.

r.lassaux@iec-iab.be

“ NE COUPEZ PAS... ON CHERCHE ”

THE DEEP BLUE ...

Le Web, c'est bien connu, ce sont des millions et des millions d'informations et de services répartis entre presque autant de sites. Ce que l'on sait moins, c'est l'ampleur du support : rien moins, fin 2000, que deux milliards de pages Web, dont la moitié ne serait d'ailleurs pas accessible. Si l'on ajoute que 80 pages supplémentaires se créent à la seconde (7 millions par jour), vous comprendrez aisément qu'aucun outil de recherche n'est actuellement capable de référencer la totalité des mots, des expressions contenues dans toutes les pages de cette mine exceptionnelle de données en tout genre.

Au risque de paraître manquer d'audace à ceux qui affichent clairement leurs prétentions de couvrir toute la toile dès 2001, on peut toutefois se demander si la question se pose vraiment en ces termes. Car, pour trouver rapidement l'oiseau rare, le site ou la page qui répond parfaitement à vos attentes, le problème, c'est avant tout de savoir où aller le chercher.

C'est précisément la raison d'être des outils de recherche que nous allons examiner dans le cadre de cet article. Ils ont pour nom : annuaires, moteurs de recherche, forums de discussion et sites spécialisés. Tous ont leur points forts et leurs points faibles, et la tâche première de tout internaute débutant consiste évidemment à savoir le(s) quel(s) utiliser en toutes circonstances.

Comment fonctionnent-ils et surtout quelles informations permettent-ils de chercher ? Quels sont les plus

performants ? Voilà donc les balises du choix. Autant de questions auxquelles nous allons à présent répondre pour chaque outil de recherche. Une sorte d'état des lieux de la question que nous compléterons en guise de conclusion par quelques conseils pratiques à garder en mémoire lors de vos pérégrinations sur la toile. Les piments de votre recherche, en quelque sorte ...



PLUSIEURS FACES D'UNE MÊME RÉALITÉ...

Des myriades d'informations à portée d'un seul clic de souris ? Si vous n'en êtes pas convaincu, lancez votre navigateur. Sauf modification de la page de démarrage par défaut, le portail multiservices de Microsoft ou de Netscape ne tarde jamais à envahir votre écran. Guère difficile d'admettre alors qu'en canalisant et regroupant, sous forme de chaînes thématiques, les principaux centres d'intérêt des internautes - actualités, sorties, loisirs, culture... - ces portes d'entrée du net vous mâchent le travail. A l'évidence, une première source d'information. Mais elle laisse rapidement sur sa faim : un classement pas toujours très pertinent, une recherche nécessairement longue et qui vous éloigne toujours un peu de votre point de départ, sont autant d'arguments qui ne plaident pas en leur faveur. Leur grand intérêt réside plutôt dans la concentration des outils de recherche qu'ils présentent : annuaire thématique de sites, moteur de recherche, liens vers les forums et les groupes de discussion qui animent la toile ... s'y côtoient fréquemment. Une rampe de lancement, en quelque sorte, des différents outils de recherche que nous allons examiner dans les lignes qui suivent.

ANNUAIRES

Un annuaire est un outil de recherche qui recense un certain nombre de sites au travers de fiches descriptives comprenant, en règle générale, le titre, l'adresse (l'URL), et un bref commentaire d'une longueur allant le plus souvent de 15 à 25 mots au maximum. La sélection des sites est le fruit du travail d'une équipe de cyber-documentalistes payés par l'annuaire, entre autres pour examiner si le contenu des sites correspond aux propositions qui leur sont faites par les internautes webmasters (dossiers thématiques, actualités, articles de fond, revues de presse, FAQ, liens intéressants). Chaque site sélectionné est ensuite inscrit dans une ou plusieurs catégorie(s), comme les divertissements, les sports, l'éducation, l'informatique, ... Une fois le mot clé saisi dans le formulaire proposé, l'annuaire va donc rechercher dans sa collection de fiches celles qui contiennent les termes demandés. S'il trouve des résultats, il va proposer, dans un premier temps, les catégories adéquates, puis les sites repris sur les fiches. Un seul clic sur le site désigné vous permet alors d'en lire le contenu des pages et obtenir la réponse à votre question. Pour les internautes anglophones, la référence est [Yahoo! International](#), une célébrité dont le système d'arborescence très simple et efficace a fait de nombreux émules : environ 100.000 sites y sont recensés, divisés en 14 catégories de base, subdivisées elles-mêmes en sous-catégories.... Nous vous conseillons également [Snap / NBCI](#), l'[Open Directory](#), [Looksmart](#) et [w.msm.](#), l'annuaire de Microsoft qu'il est

inutile de présenter. Au niveau francophone, les plus utilisés sont [Yahoo! France](#), [Nomade](#), le [Guide de Voila](#), l'[annuaire de Lycos France](#) et l'[Open Directory francophone](#).

Bilan ? Si leur force réside dans leur secret de fabrication, c'est également leur faiblesse. A défaut d'engager les milliers de travailleurs qui y seraient nécessaires, les annuaires possèdent un champ d'action forcément limité et leur actualisation s'avère parfois fort aléatoire. Mais leurs résultats sont précis, car ils sont hiérarchisés par catégorie... Bref, des outils efficaces pour une première étape dans le processus de recherche ... les pages jaunes du Web.

MOTEURS GÉNÉRALISTES

Tout le monde connaît l'un ou l'autre moteur de recherche, ne fut-ce que ceux directement implantés au cœur du PC, que sont Internet Explorer ou Netscape. Ce qu'on ignore souvent, en revanche, c'est la manière dont ils fonctionnent. C'est pourtant relativement simple. Contrairement aux annuaires qui requièrent l'intervention d'être en chair et en os, le moteur de recherche n'est rien d'autre qu'un logiciel généraliste, autrement dit un robot (crawler ou spider) qui aspire les pages Web



qu'il récupère sur la toile et les recense toutes en suivant les liens qu'il trouve dans chacune de ces pages. Au fur et à mesure de ses pérégrinations, le contenu textuel des

documents rencontrés constitue un "index", c'est-à-dire une collection plus ou moins grande de pages Web. Lorsque vous lancez une requête, le moteur va donc "fouiller" dans sa base les pages Web dont le contenu textuel - la recherche se fait en mode intégral - correspond aux mots clés que vous avez entrés. Une fois qu'il a identifié le "lot" de pages contenant le terme demandé, il classe les pages par ordre de pertinence, selon un ordre et un algorithme (basé sur certains critères de tri) qui lui est spécifique. Plus ou moins fréquemment, le moteur repasse ensuite sur les pages qu'il a indexées au préalable, pour en sauvegarder une version plus récente. C'est la phase de rafraîchissement.

Voilà pour le principe de fonctionnement. Valable pour tous, c'est à peu de choses près leur seul point commun. Pour rester pertinents, les moteurs de recherche - il y en a des centaines - multiplient les astuces pour classer, ordonner et indexer la toile. En voici une petite sélection, arguments à l'appui.

Comme on pouvait s'y attendre, les plus importants sont américains. Star incontestée outre-atlantique, [Google](#)

(560 millions de pages en base de données et un objectif supérieur à un milliard dès 2001) base son succès sur un système de recherche novateur : l'indice de popularité¹ des pages (IPP). En clair, un site est sélectionné, non pas seulement en fonction du contenu de telle page qu'il contient, mais aussi en fonction du nombre de liens qui pointent sur ce site sur Internet. Plus une page est référencée sur les sites Internet, plus cette page est dite "populaire", plus elle est considérée comme intéressante et apparaîtra en tête des résultats de la recherche. Citons également [Altavista](#) (300 millions de pages indexées) dont le nouveau moteur [Raging](#) est particulièrement vélocité, sans oublier évidemment [Hotbot](#). Souvent référencé comme l'un des meilleurs moteurs de recherche généralistes du moment sur Internet, ce moteur permet de n'afficher que les pages dont la date de création est inférieure à la période donnée : une, deux semaines, un mois, ...

Tout en gardant à l'esprit que seuls 10 % des pages du Web sont rédigées en français, c'est avec plaisir que l'on peut constater que les moteurs francophones n'ont actuellement plus grand chose à envier à leurs homologues anglophones. Les vedettes incontestées dans l'hexagone sont [Voilà](#) (150 millions de pages indexées) et [Altavista France](#). Mais d'autres, comme [Ecila](#), [Excite](#), [InfoSeek](#), [Lokace](#), [OpenText](#), [SuperSearch](#), [Caloweb](#) et, bien sûr, [Lycos](#) dégagent également des résultats honorables. Le constat vaut également pour les célèbres moteurs belges, que sont [Advalvas](#) ou [Webbel](#).

Qu'en penser ? Parce qu'ils passent à la loupe l'intérieur même des pages Web, les moteurs de recherche généralistes sont plus précis que les annuaires, mais il est difficile de se fier à un seul guide. De plus, si votre requête n'est pas correctement formulée ou trop vaste, vous serez rapidement submergé par l'abondance des résultats. Moralité : pour trouver sur Internet, tournez-vous plutôt vers les "moteurs" et vers les "moteurs spécialisés", si vous êtes davantage en quête d'informations pointues.

Démonstration.



MÉTAMOTEURS

Sous ce vocable curieux, digne des meilleurs films fantastiques, se cache un logiciel capable d'interroger en une seule fois plusieurs moteurs de recherche d'Internet et de combiner les réponses fournies par ces derniers.

Le plus connu est [Copernic 2000](#), dont la version basique est un gratuiciel, librement téléchargeable (2,3 Mo) sur le site de Copernic Technologies Pour les sites en anglais, [all4one](#), [mamma](#), [infind](#) et [dogpile](#) donneront des résultats à la hauteur de vos espérances. Parce qu'il autorise la consultation des catalogues de dix des sites de téléchargement parmi les plus importants

([Shareware](#), [Download](#), [SoftSeek](#), [FileMine](#), [Zdnet](#), [Hotfiles](#), [Tucows](#), [Alberts](#), [LinuxFiles](#), [32bit](#) et [Lycos](#)), nous ne pouvons, enfin, passer sous silence [SoftCrawler](#), un précieux utilitaire.

A l'heure du bilan, les métamoteurs n'accusent que deux points faibles : leur relative lenteur et, combinaison oblige, leur difficulté à effectuer des recherches pointues en raison de l'incompatibilité de syntaxe entre les différents moteurs.

OUTILS SPÉCIALISÉS

Dans la mesure où ils concentrent l'information sur un domaine spécifique (commerce électronique, médecine, tourisme, législation, etc.), c'est le rendez-vous des amateurs d'informations pointues.

Des exemples ? D'abord, et de toute évidence, les sites spécialisés que nous avons épinglés dans nos précédentes rubriques : euro, fiscalité, droit des sociétés, auxquels il convient d'ajouter le tout récent portail du droit belge (www.droitbelge.net). Appartiennent également à cette catégorie, les sites recensant les adresses e-mail à partir d'un nom (**Four 11** ou **WhoWhere**) ou les numéros de téléphone, comme Infobel. Parmi tant d'autres de toute sorte³, épinglons également www.scour.com, un sacré moteur de recherche spécialisé dans le contenu multimédia, [ftpsearch](http://ftpsearch.com), la référence actuelle pour trouver le pilote de périphérique, le fichier, le programme free-ware ou shareware qui vous manquent, [babylon](http://babylon.com), un excellent dictionnaire français-anglais dont la version 3.0 est en téléchargement libre sur le site de l'éditeur ...

L'avantage de la formule est évident. Outre un accès rapide aux informations, le ciblage précis de la recherche écarte également toute possibilité de confusion homonymique. Mais la médaille a un revers. A trop vouloir restreindre le champ des recherches, le risque de passer à côté de l'information en parfaite adéquation avec l'objet de la recherche est bien réel.

FORUMS, GROUPES DE DISCUSSION

Pourquoi ne pas également poser vos questions aux internautes qui partagent vos centres d'intérêts ?

Disponibles sur la plupart des annuaires généralistes (**Yahoo**, **Lycos**...) ou spécialisés (**forumone**, **forums.multimania**, **oragora**), voire via des moteurs de recherche comme **InfoScan**, et surtout **Déjà-News**, rien n'est, en effet, plus facile que de se connecter aux forums de discussion et d'effectuer, par exemple, des recherches sur des bases de données regroupant des millions d'articles postés.

S'il est a priori séduisant, le système génère en lui-même des contraintes dont il faut être conscient. D'abord, parce

que si l'on y trouve parfois le meilleur - la réponse émanera souvent des spécialistes du secteur -, on est parfois confronté au pire - chacun est libre de publier ce qu'il veut sur la toile. Ensuite, parce que la communication de votre adresse e-mail n'est pas sans risque dès lors qu'elle ouvre la porte à l'envoi de spam (messages publicitaires), de mail bombing (envoi de messages en masse), mais surtout de virus. Pour rester le plus discret possible, nous ne saurions dès lors trop vous recommander d'utiliser une seconde adresse e-mail créée pour l'occasion, voire d'envoyer votre demande via des sites qui vous garantissent, en principe, l'anonymat (**anonymizer**).

A L'HEURE DU BILAN ...

Du fait de sa croissance exponentielle, le Web n'a jamais renfermé autant d'informations. Paradoxalement, alors



que les outils de recherche de plus en plus conviviaux multiplient les astuces pour classer, ordonner et indexer la toile, il n'a jamais paru plus difficile d'y trouver des informations pertinentes. Chercher au hasard sur le Web, c'est vouloir trouver une aiguille dans une botte de foin. Mais souvent, un peu de savoir-faire autorise les miracles.

Voici donc, à la lumière des pages qui précèdent, dix conseils qui devraient vous mener sur la voie d'une recherche pertinente.

❶ Si les outils de recherche dégagent tous des résultats souvent honorables, il faut bien connaître les règles du jeu. En d'autres mots, leurs potentialités et surtout leurs limites. Tous ont leurs points forts et leurs points faibles. N'hésitez pas à les utiliser, l'un puis l'autre, pour effectuer une même recherche, et comparez la pertinence des résultats. Bien souvent, le sésame de la recherche résidera dans leur utilisation combinée...

❷ Parmi ces outils, les moteurs de recherche sont parmi les plus efficaces. Comme les meilleurs vins, ils se bonifient avec les années et les nouveaux venus regorgent d'améliorations⁴ originales pour se tailler une place au soleil.

Du moteur de recherche idéal, on pourrait penser qu'il doit posséder les caractéristiques suivantes : très rapide à la détente, apte à tout trouver, tout en présentant les résultats de manière pertinente, doté d'une interface claire possédant des fonctionnalités originales et pratiques (recherche avancée, analyse sémantique...), capable d'envoyer des rapports de recherche par courrier électronique, gestionnaire d'une fonction d'alerte détectant des changements ou des nouveaux documents sur un thème donné, sécurisé, souple, évolutif et, cerise sur le gâteau, capable de détecter les liens vides, histoire d'éviter le sibyllin, mais irritant message "error 404 - page not found"... Comme on le voit, les qualificatifs ne manquent pas pour le décrire.

Passons du rêve (plus très lointain) à la réalité quotidienne. Si cet outil idéal, celui qui prendra le meilleur de chaque technique pour être le plus efficace possible, n'existe encore que dans l'imagination pourtant débordante de concepteurs, il n'en est pas moins vrai que l'on retrouve déjà l'une ou l'autre de ses caractéristiques dans les moteurs actuels. Dans l'attente, si vous souhaitez d'ores et déjà être certain d'obtenir le meilleur du Web en une seule manipulation, la route du succès passe par les métamoteurs.



❸ En dehors de ces grands moteurs généralistes, les outils spécialisés vous ouvrent les portes des sites de téléchargement, des forums de discussion, des adresses électroniques, voire des annuaires en ligne. Leur visite vaut généralement le détour.

❹ Même si la formule est encore peu usitée dans le cadre d'une recherche d'information, les forums et les groupes de discussion ne sauraient être systématiquement écartés. S'il arrive d'y trouver le pire, le meilleur est parfois au rendez-vous.

❺ A moins de passer vos journées à surveiller la toile, actualiser ses connaissances n'est pas chose aisée. Ce n'est désormais plus un problème, dans la mesure où plusieurs sites se sont fait une spécialité de suivre l'actualité et l'évolution des moteurs de recherche. Le nirvana en la matière, c'est le site français Abondance (www.abondance.fr). Très régulièrement mis à jour, il propose, outre des informations détaillées sur les moteurs et leur fonctionnement, de fréquents comparatifs entre les différents outils. Des tonnes d'informations détaillées sur les performances et de précieux conseils sur les meilleures stratégies à adopter, vous en trouverez également en anglais par un simple clic sur searchenginesshowdown.com.

❻ Pour gagner du temps, n'hésitez pas à configurer votre navigateur. Rien de plus simple, en effet, que de choisir l'outil de recherche qui vous convient le mieux comme page d'ouverture au lancement de votre navigateur. Avec Microsoft Explorer (version 5), cliquez sur "Outil/options Internet" puis sélectionnez l'onglet "Général". Tapez ensuite l'adresse du portail dans la fenêtre "Page de démarrage". Avec Netscape Navigator (4.7), cliquez sur "Edition/préférences", puis sélectionnez "Navigateur". Tapez l'adresse de votre portail dans la fenêtre "Page d'accueil". Le tour est joué.

❼ Si vous consultez souvent le même site, pensez aux aspirateurs de sites, comme Memo web 3 flash ou web-Whacker... Grâce à ces logiciels spécialisés, rapatrier totalement ou partiellement un site sur votre disque dur s'avère un jeu d'enfant. Si elle évite des coûts de communication, cette méthode de travail ne saurait toutefois s'appliquer aux sites régulièrement mis à jour.

❽ Si une lassitude, au demeurant bien compréhensible, vous envahit dès les premières minutes de recherche sur la toile, ne perdez pas courage pour autant. Adressez votre requête à un expert. Pour obtenir une réponse précise à une question précise et, qui plus est, dans un délai déterminé, la consultation online d'un site expert compte, en effet, au rang des méthodes de recherche en vogue. Moyennant rétribution, des sites comme www.question.fr, www.paypersearch.com, www.woonoz.com, ou, sur un plan international, www.exp.com, www.askme.com et www.veille.com apporteront des réponses pertinentes à vos questions dans des domaines aussi variés que la santé, l'informatique ou les finances. Sans effectuer vous-même les recherches, il est également loisible d'être automatiquement informé de l'évolution d'une information d'une double façon. La première consiste à exploi-

ter Spyonit, un moteur qui scrute le Web et vous avertit (par e-mail, mobile, messagerie instantanée...) des nouveautés concernant l'objet de votre recherche. Il vous est également loisible de vous abonner à une liste de diffusion. Il en existe de nombreuses catégories en français ([francophones.listes](#) ou en anglais ([liszt](#)). Vous recevrez alors toutes les informations par le biais de votre messagerie électronique.

⑨ Gardez toujours à l'esprit que ce que les outils de recherche ne font pas, c'est vous garantir que l'information

trouvée est exacte. Mieux vaudra donc toujours vérifier plutôt deux fois qu'une.

⑩ Tout autant, sinon davantage, la clé du succès passe enfin par une formulation optimale de la requête. Pour cela, il n'y a pas de miracle, mais bien de nombreux trucs et astuces très pratiques pour restreindre le champ des recherches, comme l'utilisation des fonctions de recherche avancée que présentent tous les grands moteurs ou le recours aux opérateurs booléens. Nous y consacrerons le prochain article.

LES ENTRAILLES DU WEB ⁵		
ANNUAIRES	Généralistes Yahoo! International www.yahoo.com Nomade www.nomade.fr Lycos www.lycos.fr Msn www.msn.fr Wanadoo www.wanadoo.fr Voila www.voila.fr	Thématiques 7Alpha7 www.7alpha.com/ Enfin www.enfin.com/ Search Engine Guide www.searchengineguide.com/ ...
MOTEURS	Généralistes Altavista www.altavista.fr ou www.altavista.com Lycos www.lycos.fr ou www.lycos.com Caloweb www.caloweb.com Webcheck www.webcheck.com Lokace www.lokace.fr Hotbot www.hotbot.com Google www.google.com Go www.go.com Webcrawler www.webcrawler.com Infoseek www.infoseek.com Alltheweb, all the time http://www.alltheweb.com	Méta-moteurs Copernic 2000 www.copernic.com.fr SoftCrawler www.data-wizard.de/SoftCrawler ALL4one www.all4one.com Mamma www.mamma.com Infind www.infind.com Dogpile www.dogpile.com
OUTILS SPÉCIALISÉS	Question www.question.fr , Paypersearch www.paypersearch.com Expert www.exp.com , Veille www.veille.com	Droit belge www.droitbelge.net Scour www.scour.com Ftpsearch www.ftpsearch.com , Searchability www.searchability.com Babylon www.babylon.com

1. Cette indexation du Web par l'analyse de la fréquence et l'importance des liens mentionnés sur les sites nuit cependant à la prise en compte des pages récentes et/ou traitant de domaines très pointus, ce en raison du faible nombre de sites qui pointent vers elles.
2. Sa nouvelle interface intègre 55 moteurs de recherche (contre 40 auparavant) répartis en six domaines : tout le Web, forums de discussion, annuaires d'adresses électroniques, achat de livres, achat de logiciels, achat de matériel. Une fois la recherche effectuée, il trie les documents, élimine les doublons, vérifie l'existence des pages, et surligne les mots clés. Il conserve un historique des recherches et peut les affiner ou les remettre à jour quelque temps après. Dernier raffinement : il rajoute une commande dans le menu contextuel du bouton droit de la souris, pour lancer une recherche par un simple clic de souris sur un mot. Avant chaque recherche, le logiciel se met à jour automatiquement et télécharge lui-même les derniers modules d'interrogation de sites.
3. De nombreux moteurs de recherche répertorient les sites spécialisés tout en les présentant parfois en quelques lignes : *Beaucoup*, *Internet Sleuth*, *7 Alpha IDF*, *Search Power*, *The Ultimate Search Page*, *Special search engines*, *FinderSeeker*, *Internet Search Engines*, *BIG Search Engine Index*.
4. Certains moteurs pionniers, comme *Excalibur*, tolèrent désormais les requêtes imprécises, en recherchant, outre la réplique exacte de la chaîne de caractères saisie au clavier, les chaînes approchantes. Pour trouver les

pages qui répondent le mieux aux requêtes, *Caloweb* utilise également des techniques sophistiquées de correspondance textuelle. Sans déroger à la règle, le moteur *Webcheck* met à portée de tous une idée intéressante : la recherche par concept. Plutôt que de se fier au nombre de répétitions d'un ou plusieurs mots clés, *WebCheck* fouille parmi 500 millions de pages indexées pour extraire celles qui font référence au thème directement lié à la requête. Citons également *Cobrain*, mais surtout *Oingo*, qui, au lieu de rechercher les occurrences d'un mot dans une page, travaille par associations d'idées, grâce à un dictionnaire auquel il se réfère. A la demande d'un terme, il répond ainsi par une liste de sens, dans laquelle on peut choisir celui qui convient le mieux. Au rayon des fonctions nouvelles, pointons enfin *Copernic Summarizer*, un logiciel conçu pour résumer les pages Web qui s'affranchit de tous les textes. Son utilisation est simple : il suffit d'indiquer au logiciel l'adresse d'une page Web et il la résume en ne retenant que les concepts. Le logiciel s'incorpore dans Internet Explorer et Word. Si l'on ajoute que de nombreux laboratoires travaillent sur des moteurs de recherche capables de retrouver des photos, sans l'aide d'aucun mot, à partir d'un croquis ou d'indications concernant la couleur ou les textures, l'avenir s'annonce incontestablement riche en changements d'importance.

5. Cet article est consultable en ligne sur notre site Internet (<http://www.accountancy.be>). Les liens dynamiques vers tous les sites répertoriés sont activés de sorte qu'il vous suffit d'un clic pour découvrir ce que cache l'adresse URL.