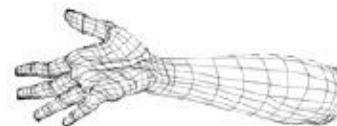


numéro soixante-sept (#67)

le mercredi 12 février 2003



Sommaire



- [Sommaire](#)
- [Editorial](#)
- [Le monde selon Google](#)
Faire des recherches sur Google :
utilisation d'opérateurs
- [La page des moteurs et métamoteurs](#)
Les moteurs – 5%
- [La chronique du Web invisible](#)
Le web invisible : comment rechercher des
bases de données
- [Enfin, les informations](#)
- [Les outils de recherche francophones](#)
- [Entretien avec Sun Wu](#)
de Openfind
- [Conclusion](#)

[Le dernier Moteurzine en
html](#)

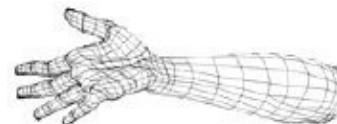
[Le dernier Moteurzine en
PDF](#)

[Les archives de
Moteurzine](#)

[Ecrire à Moteurzine](#)



Éditorial



Une actualité bien chargée...

Suite aux conseils judicieux de certains d'entre vous, Enfin.com est en train de subir un "lifting"

majeur. Non, non, je vous rassure, Enfin va rester le « guide des outils de recherche francophones », mais il n'en reste pas moins qu'il va y avoir un bouleversement important aussi bien dans la forme, que dans le contenu et la manière d'appréhender l'information. Normalement, le nouvel Enfin.com devrait sortir avant le prochain numéro de MoteurZine... mais, attention, c'est juste normalement.

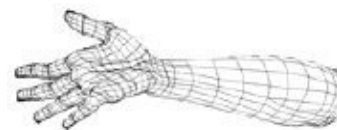
David, le responsable de la chronique sur les "métamoteurs", est l'heureux propriétaire d'un moteur de recherche développé maison, Apocalx Now, avec plus de 1 million de pages dans la base. Comme vous vous en doutez, l'entretien, le développement d'un tel moteur demande du temps, il n'a donc plus la possibilité de s'occuper de cette chronique sur les métamoteurs. Donc, si l'un d'entre vous est intéressé par reprendre cette chronique paraissant un numéro sur deux, qu'il ou elle me contacte de suite... la prochaine parution est dans moins de 10 jours.

Bonne lecture à toutes et tous,

par CHRIS HÉDÉ



Le monde selon Google



Faire des recherches sur Google : recherches spécialisées

Après les bases de la recherche sur Google, les premières recherches avancées et les opérateurs de recherche spécialisée, voici l'utilisation d'opérateurs pour modifier des recherches standard :

Recherche dans un site donné

Pour limiter la recherche à un site donné, tapez vos mots à chercher puis l'opérateur "site:" suivi du nom de domaine du site.

Exemple : *recherche avancee site:www.google.fr*

=> recherche les pages contenant les mots "recherche" et "avancee" situées sur le site www.google.fr

Remarques :

- S'il y a un mot présent sur toutes les pages d'un site (son nom par exemple), ceci permet de connaître le nombre de pages indexées du site.
- S'il n'y a pas de mot présent sur toutes les pages, on peut chercher par exclusion... Exemple :
– *itjgfuhrhfe site:www.google.fr* (cherche toutes les pages du site www.google.fr ne contenant pas le terme "itjgfuhrhfe")

Recherche dans le titre exclusivement

Pour limiter la recherche aux mots du titre des documents, tapez l'opérateur allintitle: suivi (avec espace) de vos mots à chercher. Seul le titre des documents est utilisé pour cette recherche.

Exemple : *allintitle: google recherche*

=> recherche les pages contenant les mots "google" et "recherche" dans le titre.

Recherche dans le titre

Pour effectuer une recherche mixte entre le titre et le reste du document, tapez l'opérateur intitle: devant (sans espace) chaque mot à chercher dans le titre. Les autres mots seront cherchés partout dans le document.

Exemple : *intitle:google recherche*

=> recherche les pages contenant le mot "google" dans le titre et le mot "recherche" n'importe où dans

le document (y compris dans le titre).

Recherche dans l'URL exclusivement

Pour limiter la recherche aux mots de l'URL des documents, tapez l'opérateur `allinurl:` suivi (avec espace) de vos mots à chercher. Seule l'URL des documents est utilisée pour cette recherche. L'ordre des mots ne peut pas être imposé ; on ne peut pas non plus faire une recherche comme "`allinurl: google/recherche`" pour chercher les documents ayant un répertoire "google" contenant un fichier ou un sous-répertoire commençant par "recherche".

Exemple : `allinurl: google recherche`

=> recherche les pages contenant les mots "google" et "recherche" dans l'URL.

Recherche dans l'URL

Pour effectuer une recherche mixte entre l'URL et le reste du document, tapez l'opérateur `inurl:` devant (sans espace) chaque mot à chercher dans l'URL. Les autres mots seront cherchés partout dans le document.

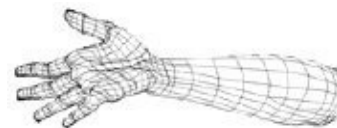
Exemple : `inurl: google recherche`

=> recherche les pages contenant le mot "google" dans l'URL et le mot "recherche" n'importe où dans le document (ou même dans l'URL).

par Olivier DUFFEZ de WebRankInfo



Espace publicitaire



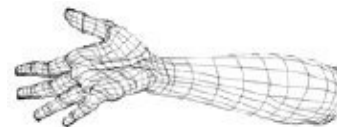
Développez votre indice de popularité...

- pour 0 euros en passant 3 jours à référencer votre site dans les **150** outils généralistes d'Enfin ([ICI](#)).
- pour 55 euros en passant 5 minutes en commandant un référencement dans les **150** outils généralistes d'Enfin ([LA](#)!).

– **vosre publicité lue par 15.000 personnes !** –



La page des moteurs et
moteurs



Les moteurs – 5%

Derrière cette valeur bien ridicule se cache une réalité bien étonnante.

En effet, ce chiffre annoncé sur un forum anglophone concerne le pourcentage réel d'indexation des moteurs de recherche par rapport au nombre de pages disponibles sur le Web. Même Google et ses 3.000.000.000 de pages mondiales ne dispose pas de l'exhaustivité tant prétendue. Il faut maintenant s'interroger sur la manière dont les moteurs de recherche trouvent les nouvelles pages.

Reprenons l'exemple DeepIndex (tant qu'à faire)... ;–)

La plupart des webmasters soumettent leur site sur les moteurs. Pourtant, ce n'est pas en recevant des centaines de soumissions par jour voire des milliers que DeepIndex ou Google peuvent couvrir tous les thèmes et sujets de l'Internet francophone voire européen (pour DeepIndex) ou mondial (pour Google).

En fait une veille est réalisée sur des sites un peu particuliers et souvent "oubliés" par le référencement. Ce sont les "petits" annuaires. Ces "petits" annuaires sont pourtant de vraies mines d'or, s'ils sont correctement tenus et s'ils sont un tant soit peu "professionnels".

Mais que vaut le "petit" annuaire du coin face à un Y..... ou à un V.... ? Les "petits" annuaires, quelque soit leur thématique, sont peut-être moins selectifs et moins importants en terme de trafic, mais ils sont certainement moins chers lorsqu'ils sont payants. Ces désavantages par rapport aux "gros" sont aussi des qualités car ces "petits" annuaires sont accessibles à tous. Les informations et les liens proposés traitent de sujets (ou de mots clés) que les gros ne couvrent pas ou alors d'une façon non exhaustive. Pourquoi ? Tout simplement à cause de critères de sélection trop stricts se basant sur autre chose que la qualité du contenu.

Vous l'aurez certainement compris, je parle ici des "petits" annuaires comme Poossin, Cyber-annuaire et / ou les autres qu'il est possible de trouver dans un annuaire d'annuaires comme Enfin. Dans le milieu on parle aussi, souvent à tort, des liens FFA ("free for all") qu'il ne faut surtout pas confondre avec les "petits" annuaires.

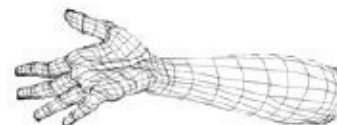
Ces "petits" annuaires ont une importance non négligeable. Il est, aujourd'hui, absolument impensable qu'un webmaster, quelqu'il soit, néglige ce vivier. De même, il est également impensable que ces annuaires disparaissent, car dites le vous bien, l'exhaustivité passe aussi par eux.

Et si certains vous demandent quelques euro pour traiter votre inscription, il ne faut pas réagir en pensant forcément trafic. Dites-vous plutôt que ce lien supplémentaire peut vous apporter une indexation car les moteurs viennent enrichir leur base dans ces "petits" annuaires.

par Gilbert WAYENBORGH de DeepIndex



La chronique du Web invisible



Le web invisible : comment rechercher des bases de données

Après avoir évoqué le concept, intéressons nous maintenant à quelques outils qui permettent de rechercher sur le web invisible.

1. Les bases de données

Certaines bases de données sont partiellement indexées par les moteurs (ainsi Google indexe près d'1/3 des 700 000 pages de l'Internet Movie Database) mais d'autres, protégées, ne le sont pas du tout. Reste à les identifier.

a. Les bases gratuites

Récemment modernisé, le site de l'URFIST (Unité Régionale de Formation et de Promotion pour l'Information Scientifique et Technique) de l'Université Lyon I propose une sélection de plus de 800 bases de données gratuites classées en 10 catégories : Généralités, Philosophie et psychologie, Religion, Sciences sociales, Langage, Sciences de la nature et mathématiques, Techniques (sciences appliquées), Arts, Littérature et rhétorique, Géographie et Histoire...

b. Les banques de données payantes

Parmi les plus importantes :

Thomson–Dialog

L'une des banques les plus volumineuses donne accès à plus de 700 bases de données sur l'intelligence Economique, la chimie, les marchés et produits, le biomédical, la santé, la pharmacie, l'ingénierie et Technologies, l'environnement, les données gouvernementales ainsi que des annuaires d'entreprises et des informations financières...

Lexis Nexis

Lexis Nexis permet d'obtenir des informations personnalisées à travers la définition d'un profil de recherche concernant des acteurs, des marchés, des technologies ou des concurrents à partir de 35.000 sources (journaux, bdd, rapports, brevets etc.) indexées dans tous les domaines et pour de nombreux pays (plus de 90)

Lexis.com : archives juridiques

Nexis.com : archives de journaux

Lexis–Nexis Alert personal : revue de presse personnalisée par mail

Lexis–Nexis Professional : service de recherche multi–critère sur base de donnée.

Factiva

Issu de la fusion de *Reuters Business Briefing* et *Dow Jones Interactive*, Factiva rassemble les dépêches de Dow Jones, de Reuters, le Wall Street Journal et plus de 8 000 autres publications du monde entier. Ces ressources permettent d'accéder à l'actualité, à des articles archivés et à des publications en langue locale (Les Echos, Le Monde, Le Figaro...), à des études de marchés, à des rapports d'analystes ainsi qu'aux cotations de bourse. Accès à 22 langues provenant de plus de 110 pays.

One Source

Une des plus grandes bases de donnée disponibles en ligne sur les entreprises et les secteurs d'activité. One source fournit des informations sur plus d'un million d'entreprises et leurs branches d'activité : Rapport d'analystes, biographie des dirigeants, études de marchés, prévisions, cotations, statistiques sectorielles. Choix possible par nom de société, par secteur, code produit.

Parmi les produits disponibles :

– US Business Browser : 250 000 entreprises aux USA et Canada

– UK BB : 350 000 entreprises au Royaume Uni

– Global BB : 350 000 entreprises US + Monde

– European BB : 300 000 entreprises en Europe et 50 000 au Royaume Uni

Parmi les services :

– Company profiler : rapport sur historique, produit, concurrence, structure

– WatchListUpdate : suivi automatique de news et documents financiers

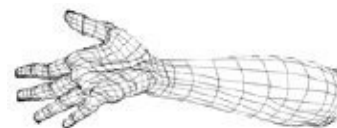
– Industry Profiler : Rapport sur marché, segmentation, tendances et prévision

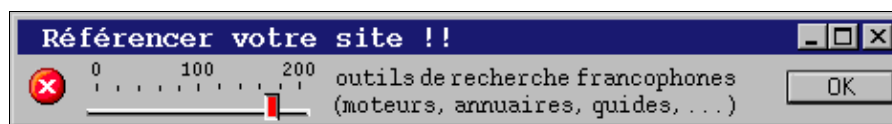
Retrouvez la suite (Les moteurs spécialistes du web invisible) dans le prochain numéro de MoteurZine.

par Christophe ASSELIN de Intelligence–center.com



Espace publicitaire

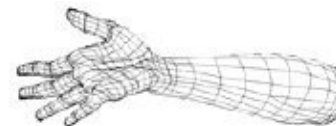




– **vosre publicité lue par 15.000 personnes !** –



Enfin, les informations



Vous êtes le responsable d'un moteur ou d'un annuaire ? Vous travaillez dans le référencement ?
Communiquez et insérez gratuitement votre actualité ici.

07/02/2003 : "Synomia: un moteur de recherche français en ASP"

« Un nouvel acteur français positionné sur le créneau des moteurs de recherche se dévoile. Lancée en juillet 2000, la société en question (Synomia), qui s'adosse à une enveloppe initiale de 2 millions d'euros d'investissement, naît de la rencontre entre un chercheur du CNRS et trois entrepreneurs en quête d'une technologie de recherche Web susceptible d'être implémentée sous forme de produit fini. »

07/02/2003 : "Lycos UK nomme une nouvelle société pour gérer sa publicité"

Lycos Angleterre vient de nommer A Vision pour gérer son budget publicitaire qui s'élève tout de même à 4 millions de livre sterling (3 millions d'euro ?????). Ce budget concerne uniquement le marché anglais et l'objectif fixé à A Vision est d'augmenter le nombre de visiteurs de 18 à 34 ans sur le portail Lycos Angleterre.

07/02/2003 : "Yahoo renforce son équipe commerciale"

Yahoo! vient de recruter deux personnes pour renforcer son équipe commerciale. Beth–Ann Eason, qui a travaillé entre autre chez DoubleClick Media et Ziff Davis Publishing, rejoint Yahoo! comme Vice–Présidente de la gestion des catégories et de la formation des forces de vente. Beth Lawrence, qui a travaillé chez America Online, DoubleClick entre autres, rejoint Yahoo! comme Vice–Présidente des ventes de l'Est.

07/02/2003 : "Qui est placé premier sur le terme "search engine" sur Google ?"

Voici, de nouveau, un résumé des quatre années de la vie de Google en commençant par sa naissance et en finissant avec les différents jeux existant autour de Google.

07/02/2003 : "Yahoo va mettre encore plus en avant les liens payants"

Yahoo! est en train de tester une nouvelle présentation pour sa page des résultats où les liens payants seraient mis beaucoup plus en avant. Cette nouvelle page permettrait à Yahoo! de générer plus de revenus et surtout va faire plaisir à son partenaire, Overture.

07/02/2003 : "Sharelook, le retour en France"

En fin d'année 2002, Sharelook (annuaire généraliste à la Yahoo! de nationalité allemande) avait mis à jour son interface pour son pays d'origine, l'Allemagne. Aujourd'hui, c'est au tour de Sharelook France de bénéficier de ce rajeunissement. Ce nouveau design laisse toujours en avant l'annuaire, fait disparaître le bandeau publicitaire et surtout nous subjugué avec une couleur prédominante qui est l'orange.

07/02/2003 : "Looksmart change sa charte graphique"

Looksmart devait être fatigué de son design d'un jaune douteux (que vous pouvez encore voir sur la version française du site Looksmart) et propose désormais une mise en page beaucoup plus épurée, beaucoup plus dynamique avec, pour la première fois, la mise en avant de Zeal.

06/02/2003 : "Des liens entre votre hébergement et votre référencement ?"

Le fait de changer de serveur peut–il nuire à votre positionnement auprès des moteurs de recherche ? C'est une question que beaucoup de référenceurs amateurs ou professionnels se posent régulièrement.

06/02/2003 : "Les résultats de Findwhat explose pour le 4ème trimestre de 2002"

Findwhat, le concurrent direct à Overture sur le marché américain, a réalisé un excellent 4ème trimestre 2002. Les résultats sont 15 fois plus importants que prévus.

06/02/2003 : "Terra Lycos se sépare encore d'une partie de son personnel américain"

Terra Lycos (société espagnole) a décidé de licencier une partie de son personnel américain, soit 147 personnes sur les 649 employés (à peu près 20%). Après cette annonce, l'action Terra Lycos a augmenté de 3,4%.

05/02/2003 : "Une option originale proposée par AllTheWeb"

Grâce au "Keyword Search", vous pouvez taper directement votre requête, précédée par atw, dans votre barre de tâche.

05/02/2003 : "Kvasir : moteur de recherche numéro 1 en Norvège"

La Norvège a une population sur Internet de 2,7 millions d'habitants, soit 60% de la population totale. Kvasir a comptabilisé 347 millions de requêtes pour l'année 2002, soit une moyenne de 128 requêtes par an et par habitant.

04/02/2003 : "Ask Jeeves, après tous les excès, les restrictions et la réussite"

A l'époque de la folie de l'Internet, Ask Jeeves y a contribué en faisant mettre, par exemple, des autocollants publicitaires sur des pommes et bananes. Aujourd'hui, après être passé de 850 à 347 employés, Ask Jeeves est une société financièrement stable. Le rachat de Téoma pour 4,5 millions de dollars a permis à Ask Jeeves de se positionner comme un challenger de Google et les 25% de part de marché le confirme.

04/02/2003 : "Hotbot, un outil de recherche qui ne vit que sur sa gloire passée"

En 1997, Hotbot était un élément incontournable du net international. Il a pendant plusieurs années fait parti du top 10 des principaux outils de recherche. Aujourd'hui, Hotbot n'est plus qu'un fantôme de ce qu'il était avant. La chute a débuté en 1999 quand Lycos et son "va chercher" a racheté Hotbot.

04/02/2003 : "Un nouvel outil de recherche anglais au paiement par clic"

Un nouvel outil de recherche anglais vient de voir le jour. Cet outil est produit par Thomson Directories et s'appelle WebFinder.com. Il est destiné aux personnes et sociétés désirant utiliser le PPC (paiement par clic) pour se faire connaître. AltaVista, Freeserve et Streetmap reprendraient déjà le contenu de cet outil.

04/02/2003 : "Sortie de Google search 1.2"

Google search est une extension qui permet de chercher dans la base de données de Google à partir de l'interface de Webmin. Pour ceux qui ne le savent pas, Webmin est l'outil majoritairement utilisé pour la gestion de serveurs Internet sous Linux. Un outil à destination des techniciens de l'Internet.

03/02/2003 : "Deuxième mois de baisse pour Google"

« Avec -0,63% ce mois, Google affiche pour le deuxième mois consécutif une baisse de trafic. Difficile de ne pas tenir pour responsable direct de ce décrochage l'annuaire qui (re)monte, Yahoo!, qui affiche +0,85% ce mois réduisant encore son écart sur Google qui reste toujours largement en tête du peloton.

Si le réveil de Yahoo! ne fait pas de doute, on peut se demander si Google aura les ressources nécessaires pour repartir à la hausse. Facteur déterminant, l'épée de damoclès représenté par Inktomi si Yahoo! le choisit en lieu et place de Google. Cela représentant encore une perte en notoriété pour le moteur leader. Rendez-vous est pris pour une réponse possible aux alentours du mois de mars. »

03/02/2003 : "Antidot distinguée par deux fois"

« L'innovation technologique et l'originalité de la solution AFS (Antidot Finder Suite) ont permis à Antidot d'être nominée pour l'European IST Prize 2003 et de recevoir la qualification "Entreprise Innovante" (Label FCPI) décernée par l'Agence Nationale de Valorisation de la Recherche (ANVAR). »

03/02/2003 : "Apocalx Now : 1 012 802 pages dans la base"

Apocalx Now, le moteur de recherche du site spécialisé dans la recherche et plus particulièrement célèbre par son métamoteur, vient de dépasser la barre du million de pages dans la base. Souhaitons donc la bienvenue au club des millions à Apocalx et rendez-vous au club des milliards (beaucoup plus petit, je vous rassure).

02/02/2003 : "Les mots-clés sponsorisés, champions de la mémorisation des publicités"

« Chaque année, DoubleClick publie un rapport tiré de ses activités d'Ad Serving avec son outil DART et fournit ainsi au marché de nombreux indicateurs concernant l'évolution de la publicité en ligne. Les données récoltées sont basées sur plus de 630 millions de publicités délivrées par DoubleClick pour le compte de plusieurs milliers de clients. Sur l'année 2002, on observe une forte hausse du taux de mémorisation des campagnes, qui passe de 0,36 % au premier trimestre à 0,53 % au quatrième trimestre. Cela représente une hausse de 47 %. Dans le même temps, DoubleClick constate l'emprise croissante des mots-clés et du contenu dans le système de ciblage des campagnes publicitaires. Dans plus de 80 % des cas, le ciblage est fait par mots-clés, loin devant un ciblage géographique (12 %). »

01/02/2003 : "Yahoo et Inktomi vont-ils freiner l'ascension de Google ?"

« NetBooster et Weborama tablent sur une éventuelle bataille entre Google et son partenaire cette année, suite au rachat d'Inktomi par Yahoo le mois dernier. »

30/01/2003 : "Yahoo! allume sa webradio musicale payante"

« Après avoir réglé mardi dernier le contentieux entre sa filiale LaunchCast (webradio gratuite) et Sony Music pour violation de copyright, Yahoo a dévoilé hier son tout nouveau service de webradio payante : LaunchCast Plus. Pour 3,99 dollars par mois ou 35,99 dollars par an, les abonnés auront accès à des douzaines de radios musicales non commerciales, dont une cinquantaine de stations thématiques (classique, country, jazz, etc.). »

30/01/2003 : "Arisem se renforce dans le KM avec MS et Business Interactif"

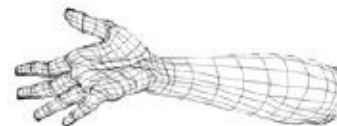
« Arisem, éditeur français spécialisé dans la gestion des connaissances, propose une solution globale de KM avec le soutien de Business Interactif et Microsoft. »

30/01/2003 : "Dell et FAST search &transfer, une longue histoire d'amour"

Dell communique autour de son partenariat avec FAST search &transfer et le fait que ses serveurs sont utilisés depuis le début par FAST. Alltheweb est passé de 80 millions de pages en 1999 pour atteindre 2,1 milliards, aujourd'hui. Tous les serveurs de FAST, plus de 1000, sont des serveurs PowerEdge et l'espace disque, plus de 200 TB, est fourni par des PowerVault.



Les outils de recherche
francophones



08 Fevrier 2003 :

TOP | sexe | adulte

E-sexe.ch : Annuaire adulte avec de nombreuses catégories, une topliste des meilleurs sites de qualité.

07 Fevrier 2003 :

TOP | internet | guide généraliste | f

Fouille : Annuaire francophone.

TOP | art et culture | encyclopédie

Wanadoo Encyclopédie : Encyclopédie Hachette avec Wanadoo.

TOP | commerce et économie | les entreprises

Space business : Le portail des entreprises.

06 Fevrier 2003 :

TOP | régional | europe | france | rhône alpes | grenoble

City38 : Enfin un portail pour les grenoblois, annuaire, cinéma, logement, bons plans...

05 Fevrier 2003 :

TOP | méta–moteurs

Cyberneto : Meta–Moteur : Cyberneto

04 Fevrier 2003 :

TOP | technologie | referencement | prest. referencement

Synodiance : Synodiance est une société de conseil spécialisée dans la conception, le déploiement et l'optimisation des stratégies de communication on–line.

TOP | commerce et économie | marketing

01expo.net : Annuaire pour particuliers et professionnels sur les foires, expositions, salons, congrès...

01 Fevrier 2003 :

TOP | technologie | referencement | prest. referencement

Brioude Internet Référencement : Prestataire spécialisé en référencement suivi de positionnement et trafic généré par les outils de recherche.

TOP | sexe | rencontre | rencontrez–vous

Africa vip : Rencontre réservée aux échanges entre l'Europe, l'Afrique, l'Ile de la Reunion, la Polynesie et les Antilles .

30 Janvier 2003 :

TOP | régional | europe | france | pays de loire

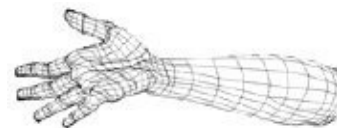
Info La Baule .com : Le portail de la Baule, Guérande et leur région.

TOP | art et culture | audiovisuel | télévision

Annuaire de Séries Max : Séries Max, un magazine et un annuaire de sites liés aux séries.



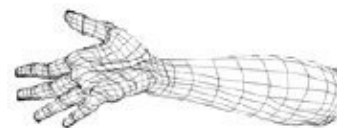
Espace publicitaire



– votre publicité lue par 15.000 personnes ! –



Entretien avec...
<http://www.openfind.com/>



Sun Wu, fondateur de Openfind

ATTENTION, cet interview est une traduction de l'américain sachant que ce n'est pas notre langue d'origine à lui comme à moi. Il est donc possible que je ne traduise pas littéralement les paroles de Sun Wu mais j'essaye de m'adapter au mieux à l'esprit de ses réponses !

«**Moteurzine**». — Bonjour. Tout d'abord, est-ce que vous pouvez vous présenter à nos lecteurs ?

Sun Wu. — Je m'appelle Sun Wu. Je suis Professeur Associé en Informatique à l'Université Nationale Chung-Cheng à Taiwan.

J'ai créé Openfind en 1998 et depuis, je travaille comme consultant pour Openfind.

«**Moteurzine**». — Vous avez travaillé sur Glimpse, la technologie de moteur de recherche. Vous pouvez nous en dire plus sur votre contribution ?

Sun Wu. — Ce que j'ai fait sur Glimpse, c'est une extension d'un précédent travail sur un outil d'approximation de champs ("an approximate pattern matching tool") s'appelant "agrep". C'était une partie de mon travail de ma thèse. J'ai implanté cet élément dans Glimpse en 1993.

«**Moteurzine**». — Vous pouvez nous raconter comment tout a commencé avec Openfind ?

Sun Wu. — Depuis 1994, je travaille sur un projet s'appelant "GAIS" (qui signifie Global Area Information Search|Service) qui est devenu en 1996 le premier moteur de recherche de langue chinoise (le moteur a été lancé fin 1995).

En 1997, des investisseurs m'ont encouragé à créer une société. Ayant pour objectif de mettre en place un moteur qui ait un bon accueil et un certain succès, je me suis lancé dans cette aventure (Il est important de noter que le projet GAIS a débuté sans aucun fond, le laboratoire n'avait même pas ses propres serveurs. :(A cause de cette absence de ressources en serveur et bande passante, je n'ai pas été capable de construire, alors, un important moteur de recherche.).

Openfind a été créé officiellement en janvier 1998 avec la technologie développée au sein du laboratoire GAIS et avec un investissement d'environ 0,6 millions de dollars américains.

«**Moteurzine**». — J'ai lu dans un communiqué de presse de Yahoo! qu'en 1998, Yahoo! Chine a intégré Openfind comme moteur de recherche. Aujourd'hui, c'est le rêve de beaucoup de moteurs de recherche de pouvoir être intégré dans Yahoo! et vous, vous l'avez réalisé en 1998. Comment ça s'est passé ?

Sun Wu. — GAIS était déjà reconnu dans le milieu chinois à cette époque. Yahoo! a ouvert sa version chinoise en avril 1998. En mai 1998, Openfind était lancé officiellement. Le jour suivant le lancement du moteur de recherche Openfind, j'ai reçu un coup de téléphone de l'équipe de Yahoo!, qui connaissait bien GAIS, pour une prise de contact. Après plusieurs mois de négociation, nous avons commencé à coopérer.

Openfind était, en 1998, une petite structure sans aucune expérience. Il y avait 20 employés, la moitié

était des ingénieurs et nous n'avions même pas une personne en charge du commercial et du marketing au commencement.

«**Moteurzine**». — Aujourd'hui, Openfind est toujours en "béta test". Qu'est-ce qui manque pour qu'Openfind passe vers une version définitive ?

Sun Wu. — Pour parler honnêtement, le projet est en quelque sorte en hibernation.

Nous sommes principalement 4 personnes à avoir travaillé sur ce projet depuis juillet 2001.

Après août 2003, nous allons être occupé avec nos contrats avec Yahoo! Taiwan et HongKong et un contrat avec YAM (le troisième plus grand portail à Taiwan). Ce sont deux gros marchés que nous avons gagné face à Google. Nous risquons donc de mettre de côté notre travail sur le moteur de recherche.

Le problème que nous avons, c'est le manque de personnel sur les projets liés au moteur de recherche, alors que l'activité principale d'Openfind est de proposer une solution professionnelle globale autour de la recherche et d'un système webmail.

Nous sommes en train de mettre en place une nouvelle société dédiée uniquement à la technologie des moteurs de recherche et sa commercialisation.

La version finale d'Openfind sera très différente de la version béta avec une nouvelle interface pour les utilisateurs et de nouveaux services comme la recherche d'images, de sons, d'actualités, de logiciels, etc.

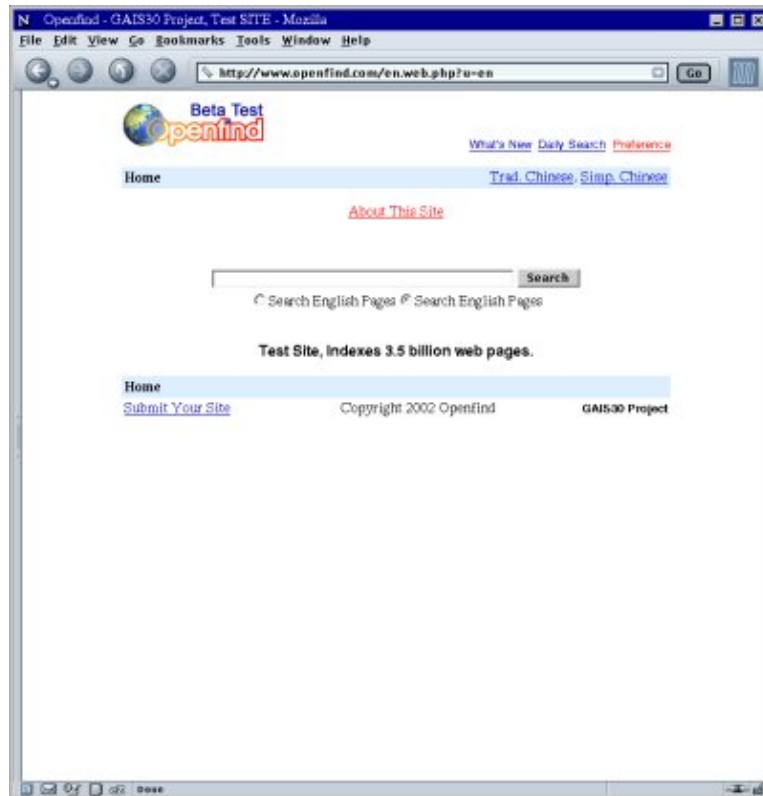
J'espère que la version définitive sera en ligne pour juin 2003.

«**Moteurzine**». — Quelle est l'architecture matériel d'Openfind ?

Sun Wu. — Nous utilisons des PCs sous FreeBSD.

La plupart des machines que nous utilisons ont ces caractéristiques :

- CPU : PentiumIII 1G – 1,26G
- Memoire : 1G ou 2G
- Disque dur : WesternDigital 120G * 4



Copie écran de la page de garde du site Openfind

«Moteurzine». — Et au niveau logiciel ?

Sun Wu. — Tous les logiciels et toute la technologie que nous utilisons sont développés en interne sauf pour l'OS (FreeBSD).

«Moteurzine». — Quels sont les moteurs de recherche que vous estimez être vos principaux concurrents ?

Sun Wu. — Je dirais cela d'une autre manière :-)

Nous voulons être les concurrents de Google et FAST.

«Moteurzine». — Pour tout le monde ou presque, le moteur de recherche numéro 1 est Google. Comment l'expliquez-vous ?

Sun Wu. — Personnellement, je respecte beaucoup Google.

Leur interface utilisateur est clair et leur service de recherche est fabuleux.

«Moteurzine». — Fast et Google font une sorte de course autour du nombre de pages qu'ils ont dans leur base. Openfind est, depuis pas mal de temps, en tête avec 3,5 milliards de pages. Mais, ce nombre n'ayant pas évolué depuis plusieurs mois, Google n'est plus très loin avec ses 3 milliards de pages. Combien en avez-vous réellement aujourd'hui ? Pensez-vous être capable d'atteindre les 5 milliards de pages et quand ?

Sun Wu. — Nous n'allons pas mettre à jour notre base de données avant le prochain lancement d'Openfind. Il y a donc beaucoup de chance que notre base contienne des pages vides.

C'est possible d'atteindre les 5 milliards de pages si nous possédons les ressources nécessaires (bande passante et serveurs). Pour maintenir un moteur de recherche avec une base très importante cela coûte énormément cher au niveau de la bande passante car il faut "crawler" (visiter) tout le web.

Nous sommes en train de créer une nouvelle société pour être capable de gérer ce genre de ressources nécessaire pour ce type de moteur.

«**Moteurzine**». — Connaissez-vous plus ou moins le pourcentage de pages pour chaque langue ?

Sun Wu. — Je peux me rappeler uniquement le pourcentage approximatif pour les principales langues :

anglais : 55%
russe : 7%
allemand : 6%
chinois : 5%
japonais : 5%
français : 3%
espagnol : 3%

«**Moteurzine**». — Pensez-vous proposer Openfind dans différentes langues avec une interface locale pour chaque pays ?

Sun Wu. — Oui, nous comptons, graduellement, construire une interface locale pour chaque pays.

«**Moteurzine**». — Quels sont les éléments pris en compte pour le positionnement dans Openfind ? Lisez-vous les méta tags ?

Sun Wu. — Oui, nous lisons les meta-tags.

Les éléments pris en compte pour le positionnement sont :

1. les éléments externes basés sur les liens entre les pages web, sur le texte de l'URL, la taille, le temps, etc.

2. les éléments internes basés sur le degré de concordance entre une requête et la page web. L'élément de concordance prendra en compte l'occurrence de la forme de la requête, sa position, etc.

«**Moteurzine**». — Vous pouvez nous dire ce qu'un webmaster doit faire ou ne pas faire pour être bien positionné sur Openfind ?

Sun Wu. — 1. Un site avec une bonne syntaxe et un contenu d'une bonne longueur sera mieux positionné.

2. Openfind va vérifier les spams et les sites qui utilisent des artifices comme beaucoup de liens artificiels seront probablement filtrés (certains sites utilisent beaucoup de domaines différents pour le même site et mettent des liens entre ces pseudo sites pour augmenter leur nombre de liens et donc augmenter leur positionnement).

«**Moteurzine**». — Vous avez développé une technologie spécifique pour Openfind qui s'appelle PolyRank. Vous pouvez nous en dire plus dessus ?

Sun Wu. — PolyRank a deux sens :

1. Cela désigne que le positionnement est basé sur des critères de position,
2. PolyRank, quand ce sera complètement achevé, permettra aux utilisateurs d'ajuster leurs critères de positionnement par eux même. C'est à dire que les utilisateurs pourront contrôler et maîtriser les effets du positionnement d'une façon plus personnelle.

«**Moteurzine**». — Comptez-vous garder la soumission à votre moteur de recherche gratuite, ou faire comme plusieurs outils et devenir payant ?

Sun Wu. — Oui, nous comptons garder une soumission gratuite pour Openfind.

«**Moteurzine**». — Plusieurs portails et moteurs de recherche proposent des résultats avec des "liens sponsorisés". Qu'en pensez-vous ? Est-ce que vous pensez intégrer ce genre de résultats sur Openfind ?

Sun Wu. — Nous n'avons pas de projets à ce niveau là, actuellement.

«**Moteurzine**». — Bien que vous soyez toujours en bêta version, connaissez-vous votre trafic quotidien ?

Sun Wu. — Ca fait très longtemps que je n'ai pas regardé les chiffres. Je ne connais donc pas le trafic que nous pouvons avoir maintenant.

«**Moteurzine**». — Comment voyez-vous l'évolution des outils de recherche dans les mois à venir ?

Sun Wu. — Beaucoup plus de services (plus de formats) seront couverts et offerts par les moteurs de recherche. Ils vont devenir de plus en plus intelligents puisque la technologie évolue. (telle qu'une classification automatique, la clusterisation des résultats, des techniques de réponses des requêtes, etc).

«**Moteurzine**». — Comment voyez-vous le futur pour Openfind ? Allez-vous être capable de faire concurrence aux moteurs de recherche majeurs ?

Sun Wu. — Ce n'est pas sûr.

Si j'ai la possibilité d'avoir assez de ressources à investir, il y a de bonnes chances.

«**Moteurzine**». — Quel moteur de recherche utilisez-vous ? Si vous pouvez construire le moteur de recherche parfait, comment serait-il ?

Sun Wu. — J'utilise Openfind et Google.

Le moteur parfait serait :

- plus intelligent et proposerait un mode plus interactif,
- plus organisé,
- plus facile à contrôler et personnaliser.

«**Moteurzine**». — Pour finir, une question plus personnelle. Quel est votre site préféré et pourquoi ?

Sun Wu. — Honnêtement, je n'ai pas de site préféré. La plupart du temps, sur Internet, je le passe à lire et gérer mes e-mails.

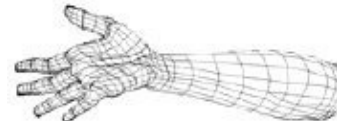
Le système de mail que j'utilise est mail2000.com.tw qui est un autre produit majeur d'Openfind.

«**Moteurzine**». — Sun Wu, merci beaucoup pour avoir trouvé le temps de répondre à toutes ces questions.

Sun Wu. — Merci.



Conclusion



ATTENTION

La liste des abonnés n'est pas disponible. Nous ne la donnons ni ne la revendons à personne. En revanche, vous pouvez sponsoriser notre lettre d'information : contactez-nous pour discuter des modalités.

[Abonnement gratuit](#)
à « Moteurzine »

[Désinscription de](#)
« Moteurzine »

les crédits

Chronique sur « Google »

Chronique sur les moteurs

Chronique sur le Web
invisible

Infodesign

Le reste (mise en page,
édito, actualités, les
nouveauautés et l'interview)

[Olivier DUFFEZ](#) de
[WebRankInfo](#)

[Gilbert WAYENBORGH](#)
de [DeepIndex](#)

[Christophe ASSELIN](#) de
[Intelligence-center.com](#)

[Max](#), le talentueux !

[CHRIS HÉDÉ](#) de
[MoteurZine](#) / [Enfin](#)