

Recherche et Référencement - Numéro 30 -- Septembre 2002

-- Lettre d'actualité mensuelle sur la recherche d'information et le référencement de sites Web --

Au sommaire ce mois-ci :

- > [Offres spéciales pour les abonnés](#)
- > [L'Open Directory en 20 questions](#)
- > [Les ressources en ligne pour tout savoir sur Google](#)
- > [Reportage photos : Espotting Londres](#)
- > [Positionnement publicitaire : un point sur la situation en France et aux USA](#)
- > [Fichier Robots.txt : les syntaxes pour Altavista, Exalead, Fast, Google, Inktomi et Voila](#)
- > [Bruits et chuchotements](#)
- > [En bref](#)
- > [Les nouveaux entrants dans l'annuaire des outils de recherche régionaux](#)
- > [Cherchez. Référenciez-vous](#) (nouveaux outils de recherche)
- > [Contenu](#) : sites proposant du contenu ou des fonctions intéressants
- > [Outils](#) : logiciels et sites Web qui aideront les webmasters et les chercheurs d'information dans leur travail quotidien
- > [Revue d'URL](#) : les meilleurs articles sur les outils de recherche parus sur le Web

Le contenu de cette lettre est accessible sur la zone "Abonnés" du site Abondance, à l'adresse :

<http://abonnes.abondance.com/archives/acturech/0209.html>

La lettre "Recherche & Référencement" paraît aux alentours du 15 de chaque mois (un seul numéro pour les mois de juillet-août)

Pour tout renseignement : © Olivier Andrieu, oa@abondance.com

LE SITE ABONDANCE VOUS FAIT PROFITER DE 3 OFFRES SPÉCIALES CE MOIS-CI :

[Retour au sommaire de la lettre](#)

OFFRE SPECIALE POUR LES ABONNÉS A LA LETTRE R&R

FORUM ET SÉMINAIRE BENCHMARK

Les 5 & 6 novembre prochains, Benchmark organise un **forum** ("Du positionnement publicitaire au référencement payant") et un **séminaire** ("Stratégies de référencement sur mesure") à Paris.
Plus d'infos : <http://www.benchmark.fr/forumreferencement2002/>

Les abonnés à la lettre "R&R" du site Abondance ont droit à **10% de remise** sur la journée du 5 novembre (forum) et sur la journée du 6 (séminaire).

Pour en profiter, envoyez-moi un message (oa@abondance.com), avec votre nom et votre société, en indiquant la journée qui vous intéresse (forum OU séminaire OU les 2), je ferai suivre votre message auprès de Benchmark afin qu'ils vous fassent une proposition adéquate !

OFFRE SPECIALE POUR LES ABONNÉS A LA LETTRE R&R

ABONNEMENT DECOUVERTE A NETSOURCES

"Connaître et bien utiliser l'Internet professionnel"

Tous les deux mois, recevez une lettre (papier) de 16 pages, qui vous aidera à **optimiser vos recherches sur le Web visible et invisible**.

En vous abonnant à **NETSOURCES**, vous apprendrez à **construire des méthodologies de recherche performantes**, vous disposerez de tous les conseils utiles pour choisir le moteur le mieux adapté à votre question, vous découvrirez des sites professionnels qui enrichiront votre bookmark. Plus d'infos sur **NETSOURCES** : <http://www fla-consultants.com>



Profitez de l'offre Découverte réservée aux abonnés de R&R :

45 euros TTC pour un abonnement de six mois (3 numéros)

N'hésitez pas : envoyez un message à l'adresse contact@bases-publications.com avec pour titre : **Offre R&R**. Vous recevrez en retour un message vous expliquant comment profiter de cette offre exceptionnelle.

OFFRE SPECIALE POUR LES ABONNÉS A LA LETTRE R&R

COPERNIC AGENT PROFESSIONAL **Téléchargez le logiciel 3 semaines avant son lancement** **et profitez d'une remise de 40% sur son prix !**

La société Copernic lance le 8 octobre prochain la prochaine version de son logiciel appelée "Copernic Agent professionnel". Du 16 septembre au 7 octobre, les abonnés à la lettre R&R ont la possibilité de télécharger et de tester ce logiciel en se connectant à l'adresse (pour la version française du logiciel) :

[https://secure.element5.com/esales/checkout.html?PRODUCT\[507482\]=1&languageid=6&p;cookies=1&COUPON1=ABONDANCECP1¤cies=EUR,USD,GBP,CAD,CHF,AUD](https://secure.element5.com/esales/checkout.html?PRODUCT[507482]=1&languageid=6&p;cookies=1&COUPON1=ABONDANCECP1¤cies=EUR,USD,GBP,CAD,CHF,AUD)

Autre lien pour la version anglaise du logiciel :

[https://secure.element5.com/esales/checkout.html?PRODUCT\[507481\]=1&languageid=6&cookies=1&COUPON1=ABONDANCECP1¤cies=EUR,USD,GBP,CAD,CHF,AUD](https://secure.element5.com/esales/checkout.html?PRODUCT[507481]=1&languageid=6&cookies=1&COUPON1=ABONDANCECP1¤cies=EUR,USD,GBP,CAD,CHF,AUD)

Les abonnés à la lettre "R&R" du site Abondance ont droit à **40% de remise** sur l'achat du logiciel jusqu'au 7 octobre (veille de la date de lancement officielle du logiciel). Le prix d'Agent PRO sera de 79,95 Euros et le prix proposé par Copernic **uniquement aux abonnés de la lettre R&R jusqu'au 7 octobre** est de 49,99 Euros !! Pour obtenir cette réduction, utilisez le **Code de coupon (rabais)** suivant : **ABONDANCECP1**

Attention : à partir du 8 octobre, l'offre de remise ne sera plus disponible !!

L'Open Directory en 20 questions

[Retour au sommaire de la lettre](#)

On ne présente plus l'Open Directory (<http://dmoz.org/>), communément appelé ODP (Open Directory Project). Cet annuaire de plus de 3 millions de sites web en toutes langues est devenu un acteur majeur dans le domaine des outils de recherche, et ce pour plusieurs raisons :

- Sa gratuité.
- Son système de maintenance et de suivi, basé sur le bénévolat des éditeurs qui s'occupent des rubriques.
- Son nombre de "clients", de sites web qui l'utilisent.
- Le fait que ce soit l'annuaire utilisé par Google et que, pour voir son site référencé sur le moteur de recherche, une astuce bien connue consiste à le référencer d'abord sur l'Open Directory, Google indexant la plupart des sites de son annuaire associé.

Mais l'Open Directory a aussi ses détracteurs. De nombreuses rumeurs circulent. L'annuaire serait une secte. Certains "méta-éditeurs" (les responsables des rubriques générales) feraient la pluie et le beau temps, acceptant et refusant des éditeurs, voire virant certains sans raisons, etc.

Bref, il fallait faire le point, à l'occasion de cette rentrée. C'est ce que nous faisons ce mois-ci avec Michelle Carrupt, formatrice indépendante à Genève, spécialisée dans les formations informatiques pour les professionnels de la santé et les médecins. Elle a fait ses premiers pas dans l'ODP en octobre 1999. Elle est aujourd'hui responsable de la catégorie World/Français, privilège qu'elle partage avec 9 autres personnes. Elle accorde une attention toute particulière à la catégorie World/Français/Santé, où elle a inscrit la plupart des sites.

Michelle avait déjà accordé une interview pour notre lettre professionnelle (<http://abonnes.abondance.com/archives/acturech/0102.html>). Vous pouvez vous y référer pour mieux connaître l'ODP, son mode de fonctionnement et son historique.



Michelle Carrupt, Open Directory

1. Pouvez-vous nous présenter l'Open Directory en chiffres, au niveau français comme au niveau mondial ? (nombre de sites, nombre de catégories, nombre d'éditeurs actifs, etc.).

L'ODP aujourd'hui c'est :

- Environ 3,8 millions de sites classés ;
- Plus de 10 000 éditeurs dans le monde (51,735 éditeurs depuis 1998) ;
- Plus de 460,000 catégories ;
- Pour la partie francophone <http://dmoz.org/World/Fran%E7ais/> :
- Environ 420 éditeurs actifs ;
- 90 179 sites (une progression de 42% en un an)
- 9 570 catégories.

L'ODP francophone se situe en 3^{ème} position pour les sites en d'autres langues que l'anglais, après l'allemand (215 834) et l'espagnol (94 584).

2. Combien de "clients" compte l'ODP ? En d'autres termes, combien de sites web utilisent-ils l'ODP ? Existe-t-il des statistiques précises à ce sujet ?

263 utilisateurs de données sont répertoriés dans :

http://dmoz.org/Computers/Internet/Searching/Directories/Open_Directory_Project/Sites_Using_ODP_Data/

et 22 pour la partie francophone.

3. Que faut-il faire pour intégrer son site sur l'ODP ? Est-ce payant ?

Non, l'inscription est et demeure "absolument" gratuite pour tous les types de sites. Pour insérer un site dans l'annuaire, il suffit de cliquer sur le lien "proposer un site" que l'on trouve sur chaque catégorie. Par exemple :

<http://dmoz.org/cgi-bin/add.cgi?where=World/Fran%e7ais/Informatique/Internet/Formation>

Si vous trouvez des offres de référencement payant pour insérer un site dans l'annuaire dmoz ou pour accélérer son classement, il s'agit d'un abus et doivent être dénoncés en tant que tels auprès du staff@dmoz.org

4. Le fait d'être inscrit sur l'ODP est-il une garantie de voir son site indexé par Google, comme on le dit souvent ?

La confusion est fréquente ! Le répertoire dmoz n'est pas le moteur Google ! Cependant, comme l'a dit Matt Cutts dans son interview, (<http://chat.abondance.com/google.html>), être référencé dans l'annuaire dmoz peut aider le PageRank de Google, par le nombre de liens qu'un classement dans l'ODP apporte.

Le répertoire Google <http://directory.google.com> reprend les données de l'ODP, votre site sera à la fois dans l'annuaire dmoz et dans celui de Google, avec toutefois un délai d'un à deux mois.

5. Certain sites ayant intégré l'ODP comme annuaire ne semblent pas proposer en ligne la dernière version de l'annuaire. Comment cela se fait-il ?

C'est une des questions que l'on nous pose le plus souvent ! L'ODP met sa base à disposition des utilisateurs de données chaque semaine. Libre à ceux-ci de la télécharger et de faire leur propre mise à jour. L'ODP n'en est aucunement responsable. Cette page donne des informations plus complètes :

<http://www.dmozed.org/special/french/www/help/getdata.html>

6. On entend parfois dire que l'ODP est "squatté" par des référenceurs professionnels ou d'autres personnes qui sont devenus méta-éditeurs et qui font aujourd'hui la pluie et le beau temps sur l'acceptation ou le refus de nouveaux éditeurs, de nouveaux sites, voire qui "éjectent" des éditeurs existants sans réelle raison. Que répondez-vous à ce type d'attaque ?

Les éditeurs de l'ODP proviennent de tous les milieux socioprofessionnels, de toutes cultures, de tous âges. Parmi ceux-ci se trouvent des référenceurs professionnels. Ce métier n'est pas incompatible avec le fait d'être éditeur, pour autant que les référenceurs-éditeurs ne se contentent pas de classer les sites de leurs partenaires ou clients. Ils doivent faire preuve de la même attitude pour les sites soumis par d'autres et traiter leurs propres sites avec impartialité.

<http://dmozed.org/special/french/guidelines/conflict.html>

"Des méta-éditeurs qui feraient "la pluie et le beau temps" et qui éjecteraient des éditeurs sans raisons valables ?" Il s'agit d'un mythe ! Certes, il y a eu quelques cas de méta-éditeurs abuseurs, cela reste extrêmement rare. Il est très difficile, voire impossible à un seul éditeur, même méta-éditeur, d'agir seul. La "surveillance" par les pairs est extrêmement efficace, les éditeurs sont vigilants et repèrent aisément les conduites abusives.

Quant aux renvois d'éditeurs, il ne peut se faire qu'après discussion et consensus de plusieurs méta-éditeurs et le contrôle du staff. Les raisons de renvois sont décrites dans les directives générales d'édition qui sont publiques.

<http://dmozed.org/special/french/guidelines/accounts.html#removal>

7. Certains anciens éditeurs semblent quelque peu amers, suite à leur expérience au sein de l'Open Directory. Leurs remarques sont aisément détectables dans les forums et les listes de discussion qui traitent du sujet. Que leur répondez-vous ?

Pour mener à bien notre activité d'édition de sites web, nous formons une communauté virtuelle multiculturelle, avec un forum interne, réservés aux éditeurs. Ce forum donne lieu à toutes sortes d'échanges, sur de nombreux sujets en rapport avec notre activité. C'est la partie "cachée" de la vie des éditeurs, dont le public ne soupçonne pas l'existence. Certains ne supportent pas d'en avoir été exclu. (Ou de s'en être exclu eux-mêmes par un comportement inadéquat) d'où les ressentiments exprimés sur différents forums et listes, ressentiments exprimés par un tout petit nombre de personnes. Ce n'est pas propre à dmoz, ce type de comportement existe dans la plupart des communautés virtuelles. Lire à ce sujet : <http://agora.qc.ca/textes/virtuel.html>

8. Autre critique entendue : l'Open Directory serait une secte. Quel grade portez-vous à l'intérieur de celle-ci ? ; -)

Cette critique m'a toujours amusée. In "la dérive sectaire", Anne Fournier, historienne, et Michel Monroy, psychiatre, in Paris : PUF, 1999. -234 p. disent :

"Le groupe développe une idéologie *alternative radicale*, exclusive et intolérante". Pas le cas de dmoz, nous sommes libres de consacrer au projet 10 minutes, 1 heure, un jour ou plus par mois. Aucune obligation de rendement.

"Sa structure est autoritaire et *autocratique*, sous la forme d'un gourou vivant ou d'une organisation héritière du message."

Le gourou ? Quel gourou ? Nous avons un éditeur en chef, un Directeur Général en quelque sorte. Les méta-éditeurs sont les leaders, les managers de la communauté. Ils agissent en tant que tels. Nous sommes une centaine aujourd'hui et aucun d'entre nous n'est un gourou, au sens sectaire du terme !

"Il revendique une *référence exclusive* à sa propre interprétation du monde, qu'elle s'applique aux croyances, aux données scientifiques, à l'éthique, aux comportements quotidiens, aux rapports interpersonnels, aux moyens de faire triompher la cause du groupe."

Pas applicable pour dmoz, le seul intérêt commun étant la création du plus grand annuaire de sites web fait par des humains.

"Il préconise des *ruptures de tous ordres* : références antérieures, orientations personnelles, relations, convictions, libre critique, choix affectifs, les relations au monde extérieur devenant marquées par le rejet, la suspicion, voire la diabolisation."

Pas applicable non plus, aucune pression de ce genre n'est exercée.

"Il multiplie *promesses et assurances* de tout genre: développement personnel, salut élitiste, toute puissance sur soi-même, santé, pouvoir collectif, promotion interne."

Aucune promesse n'est émise.

La seule analogie possible serait peut-être dans la relative dépendance que certains éditeurs développent par rapport à la communauté, jusqu'à ne pas supporter de s'en voir exclu pour comportement inadéquat ou abusif.

Je comparerais plutôt la communauté dmoz à une association, un club de sport ou un parti politique!

Les accusations de secte viennent aussi du fait que les discussions internes des éditeurs ne sont pas publiques, de même que les délibérations des méta-éditeurs sur différents sujets concernant la gestion de la communauté. Qui s'inquiète de ne pas avoir accès aux procès verbaux des séances des sociétés concurrentes? Nous ne lisons jamais ce genre d'allégations au sujet des annuaires payants ! Le fait que nous soyons bénévoles semblent incompréhensible à beaucoup d'acteurs de l'Internet marchand !

9. Il arrive souvent qu'une catégorie ne propose pas d'éditeur. Comment faire, en pratique, lors d'une soumission, pour avoir une chance de voir son site évalué par un documentaliste sur une telle rubrique ? Quel est le processus mis en place, au sein de l'ODP, lorsqu'un site est soumis sur une catégorie ne disposant pas d'éditeur ?

Si aucun nom ne figure sur une catégorie, cela ne veut pas dire que personne ne s'en occupe. Tous les éditeurs des catégories supérieures peuvent éditer dans les sous-branches, de même que les éditeurs dits "Editall" qui peuvent éditer dans tout le répertoire. Vous pouvez consulter la date de mise à jour d'une catégorie (date au bas de la page) Si aucune édition n'a été faite durant les 2 mois précédents, vous pouvez re-soumettre votre site une deuxième fois.

10. L'importance de l'ODP a grandi au fil du temps, notamment depuis que Google, qui détient 50% des parts de marché du monde de la recherche d'infos sur le Web, l'utilise

comme annuaire. Google est une entreprise cherchant clairement à gagner de l'argent, et on ne peut pas l'en blâmer ;-) Mais ces objectifs financiers sont-ils "compatibles" avec l'utilisation d'un annuaire "Open" et gratuit comme l'ODP, basé sur le bénévolat de ses éditeurs ? En d'autres termes, est-il logique qu'une entreprise comme Google gagne de l'argent et soit profitable, comme cela est annoncé, en partie grâce au bénévolat de milliers d'internautes qui passent des heures de leur temps à maintenir l'outil ? Ne serait-il pas logique que l'ODP reste gratuit pour les sites à but non lucratif mais soit payant pour les sociétés commerciales ? Ceci permettrait peut-être de rémunérer les éditeurs et ainsi d'obtenir un annuaire au contenu encore meilleur, non ?

Je ne crois pas que Google gagne de l'argent et soit profitable grâce à notre travail, mais plutôt grâce à la commercialisation de leur technologie. L'ODP est et restera gratuit, Netscape Communication s'y est engagé l'an dernier dans son contrat social :

<http://dmoz.org/special/french/www/help/socialcontract.html>

Le fait que notre activité ne soit pas rémunérée est justement ce qui motive bon nombre d'entre nous. C'est d'abord et avant tout un loisir. La plupart des éditeurs quitteraient vraisemblablement dmoz si nous devions abandonner ce mode de fonctionnement. Cela nous donne une très grande liberté, que ce soit dans le temps consacré au projet, que dans nos rapports avec les webmasters et autres professionnels du domaine. Nous pouvons quitter le projet à n'importe quel moment sans fournir la moindre raison, nous pouvons éditer au rythme qui nous convient. Nous ne sommes pas prêts à abandonner cette liberté et cet état d'esprit pour un salaire que peu de sociétés seraient à même de garantir aujourd'hui. Je ne crois pas non plus que le classement payant soit synonyme de plus grande qualité !

11. Il existe les adresses dmoz.fr, dmoz.ch, fr.dmoz.org et bien d'autres. Il faut bien avouer que l'on s'y perd un peu lorsqu'on veut consulter l'Open Directory et notamment sa partie francophone... Existe-t-il un "Open Directory francophone officiel" (le pendant de www.dmoz.org pour la langue française) ? Et si oui, à quelle adresse ?

AOL Netscape n'a pas souhaité créer des sous-domaines propres à chaque langue. (70 langues aujourd'hui) <http://dmoz.org/World/>

C'est donc assez naturellement que des domaines ont été créés et exploités. Certains de ces ceux-ci appartiennent à des éditeurs actifs, d'autres sont de simples utilisateurs de données, comme il y en a des centaines.

La seule adresse francophone « officielle » est <http://dmoz.org/World/Fran%e7ais/>

Les éditeurs ont créé, mettent à jour et hébergent : <http://fr.dmoz.org>

On y trouve des liens utiles comme cette page d'aide et de conseils :

<http://www.dmoz.org/special/french/www/help/helpmain.html>

12. Pourquoi le forum (<http://www.resource-zone.com>) a-t-il été créé ? Avec quelle ambition ? Euh, pourquoi ce nom, d'ailleurs ? Et pourquoi un seul S à "resource" ? ;-)

Ce projet a été inspiré par une suggestion du staff et réalisé en mars 2002 par la communauté des méta-éditeurs, dans un vrai travail d'équipe. Un méta-éditeur offre généreusement à la fois l'expertise technique et l'hébergement. Le nom de domaine lui appartenait.

Ce forum a été créé principalement pour la communication avec les webmasters et les non-éditeurs afin de répondre à leurs questions au sujet de l'ODP et de son fonctionnement. Il existe de nombreux forums externes consacrés à l'ODP, mais nous pensions utile de fournir de l'information validée de l'intérieur. Le forum est modéré par les méta-éditeurs. Le nombre de visites et les questions posées nous montrent que ce forum répond à un réel besoin. Les francophones ont leur espace réservé. Ce fil est spécifiquement destiné aux questions sur la situation d'un site particulier :

<http://www.resource-zone.com/showflat.php?Cat=&Board=french&Number=3048&page=0&view=collapsed&sb=5&o=14&fpart=1>

13. Combien de personnes travaillent-elles sur l'aspect technique de l'Open Directory (maintenance, création de nouvelles fonctions, etc.) aux niveaux mondial et francophone ? On a vu, à l'occasion d'un problème sur les caractères accentués dernièrement qu'un

***problème technique pouvait perdurer de nombreuses semaines avant d'être résolu...
N'est-ce pas un handicap ?***

Le staff AOL/Netscape actuellement impliqué dans la gestion du projet est composé de deux (2) personnes : notre rédacteur en chef, rdkeating25 et une programmeuse. C'est bien entendu insuffisant et nous aimerions tous avoir plus d'apport pour le développement et le support. Cependant, de nombreux éditeurs talentueux développent gratuitement des outils facilitant notre tâche.

Quant à la question des codages des pages et des accents, le staff effectue actuellement une migration des 70 langues vers Unicode (UTF8). Cela n'est pas sans poser de nombreux problèmes techniques. Certaines langues sont déjà passées sous Unicode, les catégories avec le plus grand nombre de sites, dont le français, migreront vraisemblablement en dernier. Nous publierons l'information lorsque ce sera effectif.

***14. Quelle est la structure "administrative" de l'ODP : éditeurs, méta-éditeurs, etc. ?
Quelles sont les interactions entre eux ?***

Vous trouverez la description des différents types d'éditeurs dans nos directives d'édition : <http://www.dmoz.org/special/french/erz/editors/editor-types.html#cateditors>

Une version en français est en préparation, elle sera disponible d'ici peu à la même adresse.

15. Sait-on officiellement qui s'occupe, qui gère administrativement et techniquement cet annuaire au niveau mondial et au niveau francophone ? Qui est responsable de l'outil ? Est-il possible de connaître facilement les méta-éditeurs et les éditeurs qui travaillent sur l'outil (en dehors de leur pseudo, bien sûr...) ? Cette pseudo-opacité (désolé du terme) n'est-elle pas à l'origine de certains malentendus (sectes, etc.) ?

Oui, nous le savons. L'annuaire appartient à AOL/Netscape. En plus des deux personnes impliqués dans la gestion quotidienne, un groupe d'une dizaine de personnes travaillant pour AOL/Netscape effectuent des tâches techniques.

Les méta-éditeurs traitent les candidatures des nouveaux éditeurs et à ce titre connaissent l'identité des éditeurs, du moins celle qu'ils ont bien voulu inscrire dans le formulaire de candidature. Le pseudo permet la confidentialité de cette information par rapport au public. La possibilité de rester anonyme pour le public garantit la tranquillité et la liberté des éditeurs. Personnellement, c'est un aspect qui m'avait complètement échappé lors de mon inscription, je n'ai d'ailleurs jamais cherché à masquer ma réelle identité. Je ferais différemment aujourd'hui, c'est-à-dire que je ne rendrais pas publique cette information. Nous demandons aux éditeurs de remplir honnêtement le formulaire d'inscription, en donnant son vrai nom. Il pourra choisir de le rendre public ou pas, c'est à lui de décider. Vous pouvez obtenir des informations sur un éditeur en cliquant sur son nom au bas d'une catégorie. Le profil de l'éditeur est affiché.

Quelques exemples de profil :

<http://dmoz.org/profiles/rdkeating25.html>

<http://dmoz.org/profiles/ettore.html>

<http://dmoz.org/profiles/waduca.html>

La liste des méta-éditeurs est publique : <http://dmoz.org/edoc/editall.html>

16. Quel est le processus pour devenir éditeur ? Cela peut-il se faire de façon anonyme (pour éviter tout problème d'ordre personnel avec des méta-éditeurs en place) ?

Quelques conseils disponibles ici: <http://www.dmoz.org/special/french/www/help/become.html> et sur cette page : <http://diokra.tripod.com/>

D'une manière générale, donnez votre réelle identité et fournissez des détails, y compris les URL de tous les sites que vous avez créés, totalement ou en partie. Indiquez également si vous avez contribué au contenu d'un site et fournissez l'URL appropriée. Si vous êtes affilié à plus de 5 sites, donnez-en 5 et indiquez la nature de votre affiliation pour chacun. Ne pas le faire peut entraîner la suppression des droits d'édition. Si vous pensez ne pas pouvoir collaborer avec les méta-éditeurs en place, alors ce projet n'est pas pour vous. La collaboration entre éditeurs est primordiale et un des éléments essentiels de notre activité.

17. Un éditeur, qui tarde un peu à mettre à jour sa rubrique, peut-il se faire "taper sur les doigts" par une autre personne appartenant à l'outil ? Idem si elle ne rajoute pas des sites pourtant "évidents" à sa catégorie ?? Existe-t-il un système de notation interne des éditeurs ?

Comme nous n'avons pas d'obligation de rendement ni de quota de sites à fournir, nous ne tapons personne ! Par contre le contrôle qualité existe, il est effectué par les co-éditeurs. Les cas d'abus sont signalés aux métas-éditeurs qui interviennent alors auprès de ceux qui ne respecteraient pas les directives d'édition, pouvant aller jusqu'à la suppression des droits d'édition. Il n'y a pas de classement interne des éditeurs. Des défis sont parfois lancés entre groupe d'éditeurs par rapport au nombre de sites classés. Une compétition amicale entre World/Français, World/Español et World/Italiano est actuellement en cours, dans un très bon climat.

18. Comment faire pour que son site soit évalué avec les meilleures chances d'acceptation ? Donc, dans un premier temps, comment faire pour que son site soit évalué dans les meilleurs délais (et non pas "oublié" comme cela arrive parfois, notamment si la catégorie n'a pas d'éditeur) et, dans un deuxième temps, comment faire en sorte qu'il soit accepté ? Y a-t-il des trucs et astuces à connaître ?

Le meilleur conseil que je puisse donner est de suivre la notice de soumission. Trop souvent, nous trouvons des sites du type :

www.monsite.com Titre : MON SUPER SITE. Description : Venez sur mon site, vous y trouverez tout sur le domaine xyz , suivi d'une liste de mots-clés.

Comparé à un titre et une description, nous n'hésitons pas une minute : nous traitons d'abord le site correctement soumis. Le choix de la catégorie est aussi primordial. Soumettre votre site dans les catégories du sommet de la hiérarchie va ralentir son classement. Un éditeur avec des droits étendus va devoir le transférer dans la catégorie adéquate. Ceci peut prendre plusieurs mois. Trop de sites en français sont soumis dans la partie en anglais. Soumettez votre site dans la bonne langue accélère aussi le traitement.

<http://dmoz.org/special/french/guidelines/site-specific.html#multi-lingual>

Consultez la description de la catégorie. Elle apporte des conseils sur le choix du meilleur emplacement.

Par exemple pour la catégorie Achats : <http://dmoz.org/World/Fran%e7ais/Achats/desc.html>

ou logiciels <http://dmoz.org/World/Fran%e7ais/Informatique/Logiciels/desc.html>

J'adresse un appel aux référenceurs : SVP, fournissez-nous de meilleures descriptions, choisissez les bonnes catégories pour vos sites, ne soumettez plus des sites en français dans la partie en anglais (sauf si version bilingue complète) : ils pourront ainsi être classés plus rapidement. Nous voyons encore trop d'horreurs provenant des référenceurs professionnels. Ah, si nous pouvions publier quelques perles édifiantes !

19. L'ODP francophone manque-t-il d'éditeurs ? Combien en faudrait-il pour qu'un travail "presque parfait" (la perfection n'étant pas de ce monde) puisse être réalisé ? Comment proposer sa candidature ? Quelles sont les branches de l'arborescence qui en ont le plus besoin ?

Oui, nous manquons de bons éditeurs, motivés, passionnés par un domaine et prêts à consacrer du temps pour non seulement classer les sites en attente, mais aussi pour enrichir l'annuaire avec leurs propres recherches. Idéalement, chaque catégorie devrait avoir un éditeur, soit plus de 9 000 ;) Nous pourrions aisément doubler le nombre d'éditeurs actuels.

Pour proposer sa candidature, il suffit de cliquer sur le lien "devenir éditeur" présent sur la plupart des catégories. Choisissez une catégorie avec moins de 100 sites pour un début, vous n'avez aucune chance d'être accepté pour une grosse catégorie. Voici quelques idées :

http://dmoz.org/Test/Help_Wanted/World/Fran%e7ais/Suggestions/

Un domaine vous passionne, vous en êtes l'expert, vous souhaitez partager les sites sur le sujet : devenez éditeur ! Vous participerez ainsi à la construction du plus grand annuaire fait par des humains pour la satisfaction de l'utilisateur final.

20. Globalement, comment jugez-vous la qualité de l'Open Directory et de son contenu francophone au jour d'aujourd'hui ? Quel est l'avenir de l'ODP à votre avis ?

Je dirais qu'elle est inégale et en amélioration constante ! Cela dépend des catégories, de la présence ou non d'un éditeur. En général, nous essayons de maintenir un haut niveau de qualité,

ce n'est pas toujours possible. De gros efforts ont été faits ces derniers mois, de nouvelles catégories s'ouvrent chaque jour.

Pour n'en citer que quelques-unes : l'Afrique, le Québec, l'adoption, les sports, l'archéologie et j'en oublie. Un coup de cœur pour la partie de l'annuaire réservé aux enfants et adolescents, dans laquelle un effort tout particulier est fait dans le choix de sites non commerciaux faits spécifiquement pour eux.

http://dmoz.org/Kids_and_Teens/International/Fran%e7ais/

http://dmoz.org/Kids_and_Teens/International/Fran%e7ais/desc.html

Des éditeurs du monde entier partagent la même passion : participer à la construction du plus grand annuaire du web fait par des humains, dans le but de faciliter les recherches de sites pertinents pour l'utilisateur final. Au moment où l'internet marchand prend de plus en plus d'importance, il est essentiel de maintenir cette source de données dans l'esprit du logiciel libre.




Merci Michelle !

Les ressources en ligne pour tout savoir sur Google

[Retour au sommaire de la lettre](#)

Google, vous connaissez, bien sûr... Mais, finalement, êtes-vous sûr de tout savoir sur ce moteur ? Non, bien sûr... Il existe pourtant de très nombreuses sources d'informations disponibles en ligne et très intéressantes à lire pour mieux comprendre le fonctionnement de ce leader incontesté des outils de recherche actuels.

Nous avons fait quelques recherches sur le Web et nous avons classé les ressources identifiées par grandes rubriques. Nous espérons que vous trouverez ici des sources d'information que vous ne connaissiez pas et que cela vous sera utile pour mieux comprendre le fonctionnement de ce moteur. A vos lunettes, il y aura beaucoup à lire prochainement sur vos écrans...

PS 1 : le pictogramme   signifie que le site est disponible en français, le pictogramme  indique que le contenu n'est proposé qu'en anglais.

PS 2 : Il est possible que nous n'ayons pas listé tous les sites parlant de Google et proposant, surtout, une information pointue sur son fonctionnement. Si vous en connaissez d'autres, faites-nous-le savoir !!


Sites généralistes

Sont classés dans cette catégorie les sites entièrement (ou presque complètement) consacrés à Google : historique, fonctionnement, trucs et astuces, etc.

• Google 

<http://www.google.com/webmasters/>

A tout seigneur, tout honneur, commençons par le site de Google lui-même... Sa zone "webmasters" est très complète et propose de nombreuses informations. A lire avant toute chose...

• Google France  

<http://www.google.fr/intl/fr/about.html>

La version française de l'aide en ligne. Plutôt décevante. Beaucoup moins complète que la version anglaise. Lire quand même la FAQ si la langue de Shakespeare vous rebute :

<http://www.google.fr/intl/fr/faq.html>

• Abondance  

<http://outils.abondance.com/google.html>

Vous ne croyiez pas qu'on allait l'oublier ?? ;-)

• Dsi Info  

<http://www.dsi-info.ca/moteurs-de-recherche/Google/fiche-technique.html>

Description très complète de Google.

• Search Engine Watch 

<http://www.searchenginewatch.com/subscribers/google.html>

La "bible anglophone" des outils de recherche propose surtout un article intitulé "How Google works", accessible uniquement aux abonnés payants du site.

• Search Engine Showdown 

<http://www.searchengineshowdown.com/features/google/review.html>

Description de Google par Greg R. Notess, l'un des "gourous" des outils de recherche aux Etats-Unis.

• About.com 

<http://websearch.about.com/cs/google/>

Description de Google par le site About.com, l'une des références américaines.

• WebRankInfo  


<http://www.webrankinfo.com/>

Le site se veut généraliste sur les outils de recherche, mais Google y est largement traité et son fonctionnement expliqué.

- Google Weblog 

<http://google.blogspot.com/>

Toute l'actualité sur Google. Très complet et très à jour. Indispensable.

- Forums 

<http://www.webmasterworld.com/forum3/>

<http://www.ihelpyouservices.com/forums/forumdisplay.php?s=&forumid=11>

Deux forums de discussion entièrement dédiés à Google.

- Webmasterworld Google Knowledge Database 

<http://www.webmasterworld.com/forum3/2829.htm>

Une énorme FAQ sur Google (accessible aux abonnés uniquement mais inscription gratuite).

- Google Groups 

<http://groups.google.com/groups?hl=en&lr=&ie=UTF-8&group=google.public>

Les forums Usenet spécifiquement dédiés à Google.

- The Technology Behind Google 

<http://searchenginewatch.com/searchday/02/sd0812-googletech.html>


Une vidéo de plus d'une heure qui explique comment est techniquement structuré le moteur de recherche.

PageRank


Sont classés dans cette catégories les sites, rubriques ou pages qui expliquent plus particulièrement l'algorithme de pertinence de Google : comment il fonctionne et comment optimiser votre site en fonction de ses critères.

- The anatomy of a large-scale hypertextual web search engine 


<http://www-db.stanford.edu/~backrub/google.html>

- The PageRank Citation Ranking: Bringing Order to the Web 

<http://citeseer.nj.nec.com/page98pagerank.html>

- Dynamic Data Mining: Exploring Large Rule Spaces by Sampling 

<http://dbpubs.stanford.edu:8090/pub/1999-68>


- Extracting Patterns and Relations from the World Wide Web 

<http://dbpubs.stanford.edu:8090/pub/1999-65>

Mieux vaut s'adresser à Dieu qu'à ses saints : ce sont Larry Page et Sergey Brin, ses deux créateurs, qui nous expliquent, dans ces documents, le fonctionnement du moteur !

- Efficient computation of Pagerank 

<http://dbpubs.stanford.edu:8090/pub/1999-31>

- Topic-Sensitive PageRank 

<http://dbpubs.stanford.edu:8090/pub/2002-6>

Deux documents qui tentent d'expliquer le fonctionnement du PageRank et des différentes manières de procéder pour l'améliorer.

- Google's PageRank Calculator 


http://webworkshop.net/pagerank_calculator.php3

Un outil assez complexe qui permettrait de calculer en temps réel le PageRank d'une page web.

- US Patent 

<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=5&f=G&l=50&co1=AND&d=ft00&s1=google&OS=google&RS=google>

Le brevet qui protège l'algorithme et le mode de fonctionnement de Google.

- Comment est calculé l'Indice de popularité sur les moteurs de recherche 


<http://abonnes.abondance.com/archives/acturech/0202.html>

Un article de cette lettre (février 2002) plus particulièrement porté sur le PageRank.

- The Google Page Rank Algorithm and How It Works 

<http://www.iprcom.com/papers/pagerank/>

Un long article (aspirine parfois nécessaire) qui décrit en détail l'algorithme.

- Google Search Engine Ranking Algorithm Analysis 

<http://www.pwqsoft.com/search-engine-ranking.htm>

Tests effectués sur plusieurs requêtes pour tenter de mieux comprendre l'algorithme PageRank.

- A Survey of Google's PageRank 


<http://pr.efactory.de/>

Etude très approfondie sur l'algorithme de Google.

Google Dance

La "Google Dance" désigne la procédure de mise à jour des index de Google. Le moteur utilise actuellement trois index, disponibles aux adresses <http://www.google.com/>, <http://www2.google.com/> et <http://www3.google.com/>. Chaque mois, environ entre le 20 et le 26 du mois, ces trois index sont mis à contribution par Google pendant une petite semaine afin de mettre à jour l'index principal, présent à l'adresse <http://www.google.com/>. Les deux autres (www2 et www3) sont des fichiers de travail, qui ne servent qu'à la mise à jour. Pendant celle-ci, les résultats fournis par les 3 serveurs sont différents, ce qui permet de se rendre compte qu'une nouvelle mise à jour est en cours. A noter que le serveur www1.google.com n'est plus utilisé et qu'il existait également un www4.google.com qui semble également avoir disparu pour l'instant... Le processus de mise à jour de l'index est effectué sur 3 serveurs car le calcul du PageRank est itératif (voir l'article " Comment est calculé l'Indice de popularité sur les moteurs de recherche" dans notre lettre de février 2002, disponible à l'adresse <http://abonnes.abondance.com/archives/acturech/0202.html>). Les deux serveurs de calcul www2 et www3 (l'un est-il "backup" de l'autre en cas de panne ? Probable...) "moulinent" jusqu'à ce que l'index soit complété et que le calcul du PageRank soit effectué (i.e. l'algorithme itératif a convergé). Une fois ce travail finalisé, le nouvel index est basculé sur le serveur "réel", <http://www.google.com/>.

De nombreux outils permettent d'effectuer des interrogations sur les 3 index, afin de s'apercevoir des mises à jour éventuellement en cours :

- WebRankInfo 

<http://www.webrankinfo.com/outils/googledance.php>

Un outil classique mais très fiable, en français.

- Google Tool (1) 

<http://google.fergusons.dk/>

Un des classiques et des plus anciens.

- Google Tool (2) 

<http://www.voide.be/cgi-bin/googletool.pl>

A noter que ce site permet également des interrogations sur d'autres serveurs de Google : www-sj, www-dc, www-ab, www-va et www-ex.

- Ranks 

<http://www.ranks.nl/tools/compare.html>

A noter que cet outil permet également de vérifier les différences entre les résultats fournis par Google directement et ceux proposés par Google via Yahoo!.

- My Google-Dance-Machine

<http://google-dance.miniunternehmen.de/>

Très complet, très paramétrable, mais... en Allemand !

- Dates 

<http://www.linktree.info/googleupdate.php>

<http://www.webmasterworld.com/forum3/2657.htm>

Mois et jours des dernières mises à jour de Google. Exemples pour 2002 : 21 août, 26 juillet, 23 juin, 24 mai, 26 avril, 6 avril, 20 février, 25 janvier. Le deuxième lien vous donne même la position de la lune lors des mises à jour... ;-)

Outils de recherche

Ici, nous avons listé des outils de recherche utilisant les fonctionnalités, simples ou avancées, de Google pour proposer de nouvelles fonctions de recherche sur le Web, originales. Vous trouverez de nombreux autres outils dans notre article du mois de juillet sur les applications API de Google (<http://abonnes.abondance.com/archives/acturech/0207.html>)

- Fagan Finder 

<http://www.faganfinder.com/engines/google.shtml>

Un applicatif qui permet d'effectuer des recherches sur Google en proposant des dates "normales" ;-), converties en dates juliennes (utilisées par Google) "à la volée".

- GoogleMail 

<http://capescience.capeclear.com/google/>

Possibilité d'envoyer des requêtes et de recevoir des résultats de Google par mail.

- Proximité 

<http://www.staggeration.com/cgi-bin/gaps.cgi>


Recherche les pages contenant des mots clés proches les uns des autres (fenêtre de proximité programmable de 1 à 3 termes)

- Cartographie 

<http://search.mapstan.net/>

<http://www.touchgraph.com/TGGoogleBrowser.html>

Cartographie des résultats de Google.

- Recherches quotidiennes 

<http://www.researchbuzz.com/toolbox/goofresh.shtml>

Cet applicatif effectue des recherches sur le mot clé demandé, mais uniquement sur les pages datant d'aujourd'hui, d'hier, des 7 ou 30 derniers jours.

- SearchTool 

<http://www.frysianfools.com/ggsearch/>

Logiciel qui permet d'effectuer de façon simple des recherches sur Google, à l'aide d'une fenêtre indépendante.

Les grognons

Sont classés dans cette catégorie les sites qui expliquent pourquoi ils n'aiment pas Google ou qui font part de leurs démêlés avec le moteur. On y apprend souvent beaucoup de choses...

- GoogleWatch 

<http://www.google-watch.org/>

Daniel Brandt, un webmaster de 54 ans, explique sur son site que l'algorithme PageRank de Google est "anti-démocratique" et "tyrannique". D'autre part, l'utilisation des cookies par le moteur de recherche a le don de le mettre en rogne. Bref, il n'aime pas Google et le fait savoir !

- Google Adwords Happening 

http://www.iterature.com/adwords/index_fr.html

Un webmaster s'est vu refuser des descriptifs pour ses "Adwords Select". Il s'en explique en ligne.

- Censurez Google ! 

<http://www.bowmansbrigade.com/google1.htm>

Une pétition pour censurer Google qui filtrerait certains liens ayant trait aux armes de toutes sortes dans ses "Adwords Selects".



Pour le Fun

- GooglePlex 
<http://www.google.com/plex/>
<http://www.aaronsw.com/photos/google/>
<http://www.google.com/googledance2002/>
Une plongée dans le "GooglePlex", le siège social de Google à Mountain View (Californie).
- Elgoog 
<http://www.alltooflat.com/geeky/elgoog/>
Un Google à l'envers.
- GoogleWhack 
<http://www.googlewhack.com/>
Concours : il faut chercher les expressions qui donnent 1 seul résultat sur Google !
- Logos (1) 
<http://www.google.com/holidaylogos.html>
Les logos que Google a utilisés pour commémorer un évènement sur sa page d'accueil.
- Logos (2) 
<http://www.google.com/customlogos.html>
Logos proposés à Google par des internautes, notamment lors de son lancement.
- Logos (3) 
<http://www.somethingawful.com/article.php?id=382-2>
Des logos Google détournés...
- Pâques 
http://www.google.com/Easter/feature_easter.html
Un petit jeu de Pâques proposé par Google : utilisez votre souris pour attraper les œufs portant les lettres G O O G L et E... Difficulté : il faut les attraper dans le bon ordre...
- Pigeon Rank 
<http://www.google.com/technology/pigeonrank.html>
L'algorithme secret de Google, sorti spécifiquement pour le 1er avril 2002...
- Musée 
<http://www-db.stanford.edu/pub/voy/museum/pictures/display/0-4-Google.htm>
<http://www-db.stanford.edu/pub/voy/museum/pictures/display/GoogleBG.jpg>
Les premières images de la salle machine de Google (1996)
- Parodie 
<http://www.paddedwallsfilms.com/Loogie.htm>
Parodie de Google. Son slogan : "Search about 10 good sites and about 10,000,000 crappy ones!"
- Nostalgie 
http://web.archive.org/web/*/http://www.google.com/
Les différents "looks" de Google depuis novembre 1998.
Allez, voici le premier (vous souvenez-vous ?):



Reportage photo : Espotting Londres

[Retour au sommaire de la lettre](#)

Internet et le Web, ce ne sont pas que du protocole HTTP et des mails. Ce sont aussi des hommes et des femmes (voire des animaux, hein, Lycos ?). Au travers de cette nouvelle rubrique, nous allons tenter de vous faire visiter les locaux et les ambiances de travail des acteurs majeurs de la recherche d'information en France et en Europe (voire dans le monde avec une petite visite du Googleplex, le siège social de Google, d'ici quelques mois...). Bref, de vous montrer autre chose, plus "humain", que ce que nous traitons toute l'année dans cette lettre. Pour commencer, nous parlerons d'Espotting Londres. N'hésitez pas à nous envoyer vos propositions !

Le siège social d'Espotting (<http://www.espotting.com/>) se situe à Londres. Nous avons eu le plaisir de passer leur faire un petit bonjour. Voici quelques photos qui vous feront mieux connaître l'ambiance de travail au sein de cette société, spécialiste du positionnement publicitaire.



Les locaux d'Espotting Londres se trouvent dans le quartier pakistanais, au centre de la capitale anglaise, non loin de la city.



Ils occupent le 5eme étage d'un ancienne brasserie reconvertie en bureaux.



A l'entrée, un dôme, récupéré de l'ancienne brasserie, surplombe l'accueil, histoire de rappeler les ambiances des pubs du coin...



Par la fenêtre, une cheminée rappelle l'ancienne brasserie (qui, hélas, ne fonctionne plus...)



Au premier plan, ambiance studieuse au pôle Service Client anglais. Ici, les bureaux paysagers sont de mise...



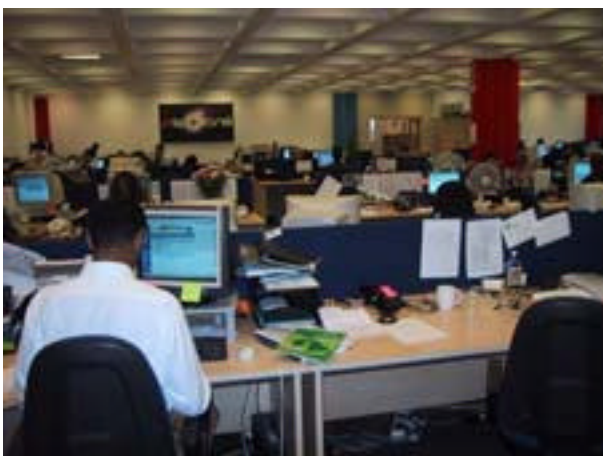
Jusqu'au fond vous avez l'équipe technique qui développe la technologie maison qui gère plus d'un milliard de requêtes / mois.



La chaîne hi-fi est là pour rappeler de façon incessante et continue les enseignements de la Direction... Big Brother is watching you... Non, it's a joke... Elle propose de la musique douce pour épanouir complètement les employés de la société :-).



Le pôle éditorial francophone au travail. Une équipe de 6 personnes qui vérifie l'adéquation des liens avec les règles éditoriales, la législation française, les cahiers des charges des différents affiliés... Si vous les avez eus au téléphone, votre interlocuteur (trice) est certainement sur la photo...



Le pôle commercial anglais avec de dos le meilleur commercial du mois :-). Au fond à droite, on peut apercevoir un passage qui mène aux bureaux de la Direction.

Et merci à Espotting Londres de nous avoir si bien reçus en leurs locaux :-)

Positionnement publicitaire : un point sur la situation aux USA et en France

[Retour au sommaire de la lettre](#)

Le marché du positionnement publicitaire se limite de plus en plus à deux acteurs aux Etats-Unis : Overture et Google. En Europe, Espotting vient se mêler au jeu. Pour y voir plus clair, voici un tableau récapitulatif des principaux portails et outils de recherche anglophones et francophones, avec, pour chacun d'eux, le partenaire associé. Attention, cela change très vite !!










Les infos ci-dessous sont également fournies pour les pages de résultats des fonctions "recherche d'information sur le Web" des portails proposés, suite à la saisie d'un mot clé. Les portails en question peuvent avoir signé d'autres accord avec des prestataires de liens promotionnels pour d'autres parties de leur site (annuaire, chaînes d'information, etc.).

Principaux outils de recherche et portails anglophones

Nom	Prestataire	Liens affichés	Emplacement	Noms	Commentaires
ALTAVISTA		3	En tout début de page	Sponsored Matches	D'autres liens (notamment le premier) peuvent être proposés directement par la régie publicitaire d'Altavista. Un lien "More sponsored matches" permet d'afficher plus de liens sponsorisés.
AOL		3	En tout début de page, avec la mention "Provided by a third party and not endorsed by AOL."	Sponsored Links	
ASK JEEVES		4	En tout début de page, juste après les sites présentés en "You may find this featured listing helpful:"	You may find these sponsored links helpful:	Pas en ligne actuellement (du fait du passage d'Overture à Google ?)
GOOGLE		8	A droite de la page de résultats	Sponsored links	Google affiche au maximum 8 Adwords par page de résultats
INFOSPACE	 	10	En tout début de page	Search engine : Overture found X results	Infospace vient de passer d'Overture à Google mais affiche encore les résultats d'Overture
LYCOS		3	En tout début de page	Sponsored search listings	Lycos propose également une solution "maison" avec Findwhat.com (sur la droite des pages de résultats)
MSN		3	Entre les "featured sites" (contrats publicitaires propres à MSN) et les sites web issus de l'annuaire	Sponsored sites	
YAHOO!		3	En dessous des "catégories" de l'annuaire, au dessus des sites	Sponsor matches	

			dessus des sites web proposés en positionnement naturel		
--	--	--	--	--	--

Principaux outils de recherche et portails francophones

Nom	Prestataire	Liens affichés	Emplacement	Noms	Commentaires
ALTAVISTA FRANCE		5	En tout début de page	Liens partenaires	2 liens en haute de page et 3 en bas. Un lien "Plus liens partenaires " permet d'afficher plus de liens sponsorisés.
AOL FRANCE		4	En tout début de page	Liens sponsorisés	Affichage des liens assez bizarre (parfois illisibles) sur certaines plates-formes...
CLUB INTERNET		"Full integration"	En tout début de page	Liens promotionnels - Overture	Autant de liens promotionnels affichés qu'il y a d'annonceurs
GOOGLE		8	A droite de la page de résultats	Liens commerciaux	Google affiche au maximum 8 Adwords par page de résultats
HOTBOT FRANCE	 	3	A gauche de la page	Liens sponsorisés	HotBot France devrait changer en décembre 2002 et passer d'Espotting à Overture.
LYCOS	 	3	En tout début de page	Liens sponsorisés	Lycos France devrait changer en octobre 2002 et passer d'Espotting à Overture. Les résultats d'Espotting seront cependant encore affichés dans les "chaînes" et les annuaires du portail
MSN FRANCE		3	Entre les "sites présentés" (contrats publicitaires propres à MSN) et les sites web issus de l'annuaire	Sites sponsorisés	
NOMADE.FR / TISCALI RECHERCHE	 	20	En tout début de page	Résultats Espotting	Les liens Espotting sont affichés sur le spages de résultats sur une interrogation par mots clés. Autant de liens promotionnels affichés qu'il y a d'annonceurs (à concurrence de 20 résultats). Tiscali

					(propriétaire de Nomade.fr) affiche également les résultats Overture sur les interrogations Shopping, les chaînes thématiques et l'annuaire.
WANADOO/ VOILA	<i>Pas de prestataire pour l'instant</i>	--	--	--	Voila / Wanadoo propose une offre au CPM baptisée "Pole Position"
YAHOO!		5	En tout début de page	<i>Liens promotionnels</i>	

Fichier Robots.txt : les syntaxes pour Altavista, Exalead, Fast, Google, Inktomi et Voila

[Retour au sommaire de la lettre](#)

Le fichier Robots.txt permet d'indiquer aux moteurs les zones d'un site web à indexer ou pas. Un article complet lui est consacré dans la zone gratuite du site Abondance (<http://docs.abondance.com/robots.html>). Nous ne reviendrons pas sur son utilisation et sa syntaxe générale, mais nous avons posé, à son sujet, trois questions aux moteurs de recherche majeurs sur le Web à l'heure actuelle (Altavista, Exalead, Fast, Google, Inktomi et Voila) :

1. Un fichier robots.txt est-il indispensable ou recommandé sur un site web par rapport à son indexation par votre moteur ? Que se passe-t-il si ce fichier est absent ?

2. Si on désire indiquer à TOUS les robots de ne pas indexer un répertoire (exemple "cgi-bin"), la syntaxe sera :

User-agent: *

Disallow: /cgi-bin/

Mais quelle **syntaxe** utiliser pour indiquer spécifiquement à votre robot de ne pas indexer, par exemple, le répertoire "clients" :

User-agent: ????

Disallow: /clients/

Y a-t-il plusieurs orthographes possibles (plusieurs noms de robots) ?

3. Les balises Meta "Robots" sont-elles prises en compte par votre moteur à l'heure actuelle ?

Voici les réponses des différents moteurs :



Réponses d'Altavista :

1. La présence de ce fichier est recommandée pour contrôler l'indexation du site, mais elle n'est pas indispensable. Si le fichier est absent, nous faisons le crawl. Cependant si le fichier est présent, mais nous ne pouvons pas le lire ou l'accéder (par exemple erreur 403 "Access Forbidden"), nous ne faisons pas le crawl du site.

2. Le nom du robot est "scooter", par exemple:

User-agent: scooter

Disallow: /clients/

3. Oui, par exemple:

<META NAME="ROBOTS" CONTENT="NOINDEX, NOFOLLOW">



Réponses d'Exalead :

1. Le fichiers robots.txt n'est pas indispensable, et s'il est absent ou invalide le moteur considère que tout le site est autorisé. Le fait de savoir s'il est recommandé dépendra du contenu du site : une meilleure indexation du site sera obtenue si le fichier robots.txt bloque l'accès aux pages dont le contenu n'est pas pertinent pour le moteur (pages contenant des logs, pages générées à la volée sans contenu documentaire utile par exemple). En gros le fichier robots.txt est souvent inutile pour les sites simples, et d'autant plus utile que le site est sophistiqué ou généré par des outils sophistiqués.

2. Le robot Exalead répond aujourd'hui au doux nom de "NG" (il s'annonce par "User-Agent: NG/1.0"). Il honore donc les directives préfixées par "User-agent: NG" ou "User-agent: NG/*". Dans la prochaine version ça sera "Exabot".

3. Ce n'est pas supporté par la version qui est actuellement en production, mais c'est prévu dans la prochaine, qui devrait sortir courant septembre.



Réponses de Fast :

1. Nous n'avons besoin d'un fichier Robots.txt que si vous désirez que le robot ne crawle pas certaines parties de votre site. Vous n'êtes pénalisé en aucune façon si le fichier robots.txt n'est pas présent.

2. Utilisez le " User-agent" fourni dans vos logs si vous désirez faire une différenciation entre certains de nos crawlers. Sinon, cette syntaxe convient à tous nos crawlers :
User-Agent: fast
Disallow: /clients/

3. Oui, nous prenons en compte les options "nofollow", "noindex", "noimageindex" et "none".



Réponses de Google :

1. Le fichier n'est pas obligatoire. S'il existe, nous le prenons en compte. S'il n'existe pas, nous indexons le site.

***NDLR :** pour info, indiquons qu'auparavant, lorsque l'accès au fichier Robots.txt d'un site web générait une erreur 403 ("access forbidden"), Google n'indexait aucune page du site. Dorénavant, dans ce cas-là, le moteur fait comme si ce fichier n'existait pas et indexe les pages.*

2. La bonne syntaxe est :
User-Agent: Googlebot
Disallow: /clients/

3. Oui, Google supporte la balise Meta "Robots", mais également les balises "Googlebot" qui lui sont propres. Voir :
<http://www.google.com/remove.html>
<http://www.google.com/webmasters/3.html#B1>



Réponses d'Inktomi :

1. Le fichier n'est pas nécessaire. Par défaut, comme la plupart des autres robots, nous estimons que le fait qu'un fichier Robots.txt n'existe pas signifie que nous pouvons crawler le site dans son ensemble.

2. La bonne syntaxe est :
User-agent: slurp
Disallow: /clients/

3. Oui, pas de problèmes, nous prenons en compte la balise Meta "Robots". Voir :

<http://www.inktomi.com/slurp.html>



Réponses de Voilà :

1. L'absence d'un tel fichier ne pose pas de problème. Nous indexons, dans ce cas, le site. Pour info, le fichier robots.txt est absent sur 93% des sites.

2. La syntaxe est :

User-agent: VoilaBot

Disallow: /clients/

A noter que le robot change très bientôt mais son nom devrait être le même (seul son numéro de version sera modifié).

3. Oui, nous suivons les directives "noindex" et "nofollow". Il y a des cas qui ne marchent pas, par ex: A -> B dans A il y a un nofollow, donc on indexe pas B. Par contre dans une page externe, si quelqu'un pointe sur B, et on cawle B.

Et un grand Merci à Mark Barlow (Altavista), Patrice Bertin (Exalead), Kristian Aune (Fast), Nate Tyler (Google), Ron Verheijen (Inktomi) et Pierre Aubert (Voila) d'avoir bien voulu nous donner ces indications.

Bruits et chuchotements

[Retour au sommaire de la lettre](#)

Une rubrique qui regroupe tous les bruits et rumeurs dans le (petit) monde des outils de recherche mondiaux et francophones. Rien n'est obligatoirement vérifié, mais toutes les infos sont données... de source sûre ;-)

-> Voici les dernières infos de "Radio-Moquette" ;-)) sur le choix de Yahoo! pour son prochain prestataire "moteur" : dans un premier temps, il semblait que Yahoo! voulait clairement changer de partenaire "moteur" et que le contrat avec Fast était en fait déjà signé depuis le début de l'été. Fast travaillait d'ailleurs depuis quelques semaines sur certaines fonctionnalités qu'il ne proposait pas encore et que Yahoo! avait demandées (sur l'indice de popularité entre autres). Fast avait jusqu'à septembre pour "faire ses preuves" et proposer les fonctions demandées. Mais il semblerait qu'une rumeur sur Inktomi soit parvenue à certaines oreilles en fin d'été. Inktomi serait-il choisi si Fast n'arrive pas à répondre aux attentes de Yahoo! ?

Ceci dit, d'autre part, Danny Sullivan (<http://www.searchenginewatch.com/>) a noté sur Yahoo! US certaines nouvelles fonctions (voir rubrique "en bref" ci-dessous) qui tendraient à prouver que Yahoo! garderait Google comme prestataire. Bref, avec tout ça, on n'est pas bien avancé... ;-)

Voici donc 3 hypothèses qui nous semblent plausibles pour ce "feuilleton de l'été" :-)

1. Yahoo! ne veut plus de Google qui lui a "chipé" trop de trafic "search" ces derniers mois. La meilleure solution de recherche, estimée par Yahoo! est Fast : Fast Gagnant ! Cette solution semble tenir la corde aujourd'hui.

2. Yahoo! ne veut toujours pas de Google pour les mêmes raisons que ci-dessus. Fast est intéressant technologiquement mais Yahoo! a la possibilité de racheter Inktomi pour un prix très intéressant. Il peut ainsi beaucoup mieux maîtriser sa solution "moteur". Inktomi Gagnant ! Cette solution, que Yahoo! envisageait fortement le mois dernier, semble moins d'actualité aujourd'hui.

3. Google ne veut pas que l'un de ses challengers (Inktomi ou Fast) soit "boosté" par un accord de partenariat avec Yahoo! Il relance le portail américain et emporte in fine le contrat. Google vainqueur ! Solution envisageable pour les "fans" des retournements de dernière minute ! A vous de choisir votre scénario ;-)

En tout cas, il semblerait que la décision soit imminente et même que Yahoo! ait déjà pris sa décision ! Normalement, l'annonce devrait se faire aux alentours du 22/23 septembre au plus tard...

A ce sujet, voici une url (merci à "Serge" sur le forum Abondance consacré au référencement) qui ressemble fort à un comparatif et un "test interne" de Google et Fast :

<http://m18.search.yahoo.com/eval-fr1/html/static/jogging.html>

-> Il serait de plus en plus question de l'installation d'Ask Jeeves en France...

-> Google Italie ferait-il dans le bandeau publicitaire de type "bannière" ? Oui, si l'on en croit cette url :

<http://www.google.com/custom?q=gaz+de+france+/secure+computing&hl=it&lr=&ie=UTF-8&cof=L>: <http://www.westcoast.com/search/header.gif>

Url trouvée par l'un de nos abonnés dans ses stats "referrers"...

-> Lors de la présentation et du lancement d'Overture France, du jeudi 12 au dimanche 15 septembre 2002 au Paris Golf & Country Club (Rueil Malmaison), Espotting, son concurrent, avait fait venir d'Angleterre un taxi londonien à ses couleurs qu'il avait garé à l'entrée du club de tennis... Hum hum...

-> Selon Planet Ocean (<http://www.searchengine-news.com/>), Altavista n'indexerait plus du tout les balises Meta Keywords et que partiellement les balises Meta Description. Renseignements pris auprès du moteur de recherche, l'info est exacte : les balises Meta keywords ne sont plus du tout indexées, en revanche, les balises Meta description le sont, elles.

-> Espotting va bientôt (d'ici la fin de l'année) ouvrir des filiales dans d'autres pays européens. D'autre part, il travaille actuellement à adapter sa technologie pour afficher ses liens sur les téléphones portables et autres appareils mobiles GPRS.

-> Le moteur de recherche Wisenut (<http://www.wisenut.com/>) devrait bientôt annoncer un index bien plus important que celui qu'il utilise actuellement.

-> Le trafic des sites européens de Google régresserait depuis 2 à 3 semaines, sauf pour la France, où le trafic du site google.fr progresserait pour sa part...

-> Nouvelle version d'AOL.FR courant septembre normalement : la nouvelle mouture de la technologie Exalead proposera pas mal de nouveautés, avec une pertinence fortement accrue, un contour de crawl affiné (indexation de sites en anglais mais pouvant intéresser un public français), des outils de correction linguistique (notamment phonétique), une nouvelle présentation des keywords et catégories, etc. Bref, pas mal de nouveautés à découvrir en ligne dans les jours qui viennent.

-> Le contrat qui lie Yahoo! Europe à Espotting est renégocié en novembre 2002. Overture compte bien s'installer autour de la table. Ca va chauffer !! ;-)

En bref...

[Retour au sommaire de la lettre](#)

-> Emmanuelle Siary, responsable "développement" de Looksmart France, quitte la société Looksmart à partir de mi-septembre. Looksmart France qui, par ailleurs, n'existe plus (ça, on le savait déjà), mais ne propose également plus de référencement payant depuis début septembre. Bref, Looksmart France is dead...

-> Si vous voyez dans vos logs le robot "Nomadobot" :
Mozilla/4.0 (compatible; MSIE 5.5; Windows NT 5.0; NomadeBot 2.0)
Il s'agit bien du robot de Nomade (Tiscali Recherche pour les intimes ;-))
Il s'agit d'une application crawlant les sites pour déterminer s'ils sont valides. Ensuite, les documentalistes de Nomade vérifient les sites crawlés en 404 et les supprime de la base, afin d'en améliorer la fraîcheur et la qualité.

-> Altavista a annoncé l'indexation de plus de 3 000 nouvelles sources d'information de type "dépêches d'actualité" toutes les 15 minutes.

-> Suite au "remplacement" de la technologie Antisearch par Deepindex sur Francité, Stéphane Loesel, Directeur technique d'Antisearch, nous donne quelques indications : "La technologie de recherche AntiSearch est toujours opérationnelle sur notre site de démonstration <http://www.antisearch.net> et sera prochainement utilisée dans de nouveaux contextes indexant tout ou partie du web francophone avec encore plus de pertinence et de formats de fichiers supportés et avec des possibilités de recherche avancée innovantes. S'il semble en effet que certaines parties de Francité n'utilisent plus AntiSearch (la recherche avancée utilise toujours nos services !) ; nous n'avons pas encore d'information officielle à ce sujet car Francité achetait notre technologie chez notre partenaire Ixo et pas directement chez nous. Vous pouvez continuer à soumettre vos pages en utilisant ce formulaire : http://www.antisearch.net/URL_SUBMISSION/submit_url.php"

-> Overture refuse plus de 25% des demandes de mots clés qui lui sont soumises.

-> Sur Yahoo! US, les descriptions des sites web proposés par Yahoo! ont changé. Si le site est inscrit dans Yahoo!, c'est le résumé écrit par les documentalistes de l'annuaire qui apparaît au lieu du "Snippet" habituel (texte entourant le mot recherché dans le texte visible de la page).

-> Le site Rolist (<http://www.rolist.com>) propose des résultats en rotation pour les liens sponsorisés. Voie intéressante...

-> Lycos a instauré un nouveau programme, intitulé "Trusted Partner", afin d'assurer une certaine pertinence dans les résultats affichés, et offrir un service de qualité aux Internaute. Dans le cadre de ce programme, Lycos a décerné un premier certificat à la société de référencement NetBooster. Source : http://www.netbooster.com/netbooster_lycos.htm

-> Gilbert Wayenborgh nous a fait parvenir un bilan du moteur de recherche DeepIndex (<http://www.deepindex.com>) après 3 mois d'exploitation. Laissons-lui la parole : "Lancé en juin dernier DeepIndex a réussi à se hisser parmi les premiers outils de recherche francophone avec près de 3.000.000 d'urls indexées et un trafic en constante hausse. Cette réussite est due à plusieurs éléments dont notamment la vente en marque blanche du moteur. C'est ainsi plus de 150 portails utilisent déjà le moteur de recherche de DeepIndex dont le célèbre Francité que DeepIndex a décroché fin Juillet.

"Nous avons réussi en trois mois à valider l'outil de recherche, le modèle économique et une adhésion du public, alors que la période estivale est une des plus creuses au niveau du trafic." C'est aussi une preuve que l'internet francophone n'est pas aussi moribond. En effet la plupart des acteurs impliqués ont travaillé bénévolement à ce projet qui se trouve aujourd'hui commercialisé par DeepIndex Inc au Canada avec son fondateur Gilbert Wayenborgh, et Laurent Lucas de Networldmedia. La société DeepIndex SARL est quand à elle chargée du développement technologique.

Gilbert Wayenborgh, acteur de l'internet francophone depuis 1996, avait déjà lancé plusieurs portails auparavant, avant de se lancer contre toute attente dans le projet DeepIndex, lorsqu'un ami lui lançait "il n'y a que des moteurs de recherche américains".

"Nous avons alors appliqué l'une des bonne vieille méthode : - Louer un garage, et monter un projet entre amis -."

Hormis cette image du garage, le projet germait depuis plusieurs mois déjà dans de longues conversations avec Sebastien Halley (Francesurf.net) notamment et aussi Chris Hédé (enfin.com) sans compter Laurent Lucas (networldmedia.net) qui a déclenché la commercialisation.

A bout de trois mois le "garage" est devenu trop exigû, et il devenait indispensable de cadrer le projet amical dans un projet à plus long terme en fondant une société commune entre Networldmedia et DeepIndex SARL : DeepIndex Inc. Cette entité commune assumera dorénavant tous les frais d'exploitation et de mise en marché de Deepindex.

La validation de Francité nous a permis de voir plus loin, mais pas seulement. Nous devons valider le projet et l'application sur un portail grand public, voir si l'application gardait des temps reponses satisfaisants.

Nous avons aussi depuis établi des contacts avancés avec d'autres portails majeurs.

"Aujourd'hui nous pouvons aller sérieusement et sereinement sur le marché" pour louer DeepIndex, car en plus d'être performant, les coûts d'utilisation et d'exploitation de DeepIndex sont beaucoup moins élevés que ceux de la concurrence, près de 40% de moins."

Les nouveaux entrants dans l'annuaire des outils de recherche régionaux

[Retour au sommaire de la lettre](#)

Pour obtenir tous les sites :

<http://annuaire.abondance.com/>

-> Websaone (région Bourgogne)

<http://www.websaone.com/index.php?page=liens>

-> Annuaire web du Finistere (région Bretagne)

<http://www.annuaire29.org>

-> An Arvorig (région Bretagne)

<http://www.anarvorig.com>

-> CaledOguide (région Dom-Tom)

<http://www.caledoguide.com>

-> Web O4 (région PACA)

<http://www.france-escapade.net/annuaire.asp>

-> L'annuaire des professionnels de Saumur (région Pays de la Loire)

<http://www.i-saumur.com>

-> Annuaire de la Vendée (région Pays de la Loire)

<http://www.chouan.com>

-> Lyon Passionnement (région Rhône-Alpes)

<http://www.lyon-passionnement.com/annuaire/index.htm>

Cherchez, référencez-vous (nouveaux outils ou rappel d'outils existants)

[Retour au sommaire de la lettre](#)

-> Neptune 4

<http://www.neptune4web.com/>

Nouveau moteur de recherche (Canada)

-> Recherche Web

<http://www.recherche-web.com/recherche.shtml>

Nouveau moteur de recherche

-> Windseek

<http://www.windseek.com/>

Nouveau prestataires de liens sponsorisés (type Overture, Espotting...)

-> GateReal

<http://www.gatereal.com/>

Moteur de recherche sur le multimédia

-> DataScor

<http://www.datascor.net/phpdig/index.php>

Moteur de recherche sémantique

-> OOzap

<http://www.oozap.net>

Une "télécommande du Web"

-> Suisse

<http://www.valipage.ch/>

Nouvel annuaire suisse

-> Belgique
<http://www.b-annuaire.be>
Nouvel annuaire belge

Contenu (sites proposant du contenu ou des fonctions intéressantes)

[Retour au sommaire de la lettre](#)

-> Suisse
<http://www.suisse.t2u.com/>
Liste des outils de recherche suisses

-> Formation
<http://www.formation.info>
Annuaire spécialisé sur les thèmes de la formation et de l'éducation

-> Tutoriels
<http://www.tutoriels.com/home.php>
<http://www.tutorialfind.com/tutorials/>
Toutes les ressources pour se former en ligne (le premier site en français, le deuxième en anglais)

-> Cartes
http://www.bonweb.com/p_5_515.html
Recherchez vos destinations en saisissant des requêtes dans les cartothèques interactives.

-> Livre
<http://www.searchengine-report.co.uk/>
Search Engine Marketing : un livre (en anglais) consacré au référencement et au positionnement.

-> Portails
<http://www.webmasterworld.com/forum5/1234.htm>
<http://www.bruceclay.com/searchenginechart.pdf>
Liste des grands portails anglophones avec leur prestataires en annuaire et moteur de recherche.

-> Francophonie
<http://www.enqueteur.fr.fm>
Annuaire de la francophonie

-> Weblog
<http://www.searchengineblog.com/>
Weblog qui traite... des outils de recherche.

-> Forum
<http://www.supportforums.org/>
Forum sur l'optimisation de site web pour le référencement.

-> LLRX.com's New Search Engine Resource Center
http://www.llrx.com/search_engines.html
Des tas de ressources sur les outils de recherche, par un des sites références sur le sujet.

Outils (logiciels et sites web proposant des fonctionnalités utiles)

[Retour au sommaire de la lettre](#)

-> Métamoteur
<http://www.ithaki.net/indexf.htm>
Métamoteur qui interroge Free, Altavista, Yahoo!...

-> Annuaire
<http://www.comscripts.com/scripts/?script=804>
Annuaire de liens type Yahoo! pour hébergeur PHP 4 / MySQL

-> Outils

<http://www.searchengineworld.com/misc/tools.htm>

Quelques outils disponibles en ligne pour tester votre site web.

-> Liens courts

SnipURL

<http://snipurl.com/>

TinyURL

<http://tinyurl.com/>

MakeAShorterLink

<http://makeashorterlink.com/>

Trois sites qui proposent de créer des liens très courts de redirection vers des urls très (trop) longues.

-> MailSyst

<http://www.mailsyst.com>

Outil d'envoi de newsletter par mail.

-> Robotcop

<http://www.robotcop.org/>

Projet de nouveau fichier "Robots.txt"

Revue d'URL

[Retour au sommaire de la lettre](#)

-> Taking Advantage of Open Directory Project Submissions

http://www.searchengineguide.com/lander/2002/0904_e11.html

Comment tirer avantage d'une soumission sur l'Open Directory

-> German Search Engine Resources

<http://www.searchenginewatch.com/sereport/02/09-german.html>

Nombreuses informations sur les outils de recherche allemands.

-> An End to Metatags

<http://www.traffick.com/article.asp?aID=102>

Les balises Meta sont-elles vraiment obsolètes ?

-> PPC Versus "Free" Listings: Which is Best?

http://www.searchengineguide.com/wi/2002/0827_wi1.html

Le match entre référencement manuel et positionnement publicitaire.

-> Liens promotionnels : le statut des contrats de positionnement

<http://www.journaldunet.com/juridique/juridique020826.shtml>

Quelles règles sont applicables aux "liens promotionnels" et aux contrats de positionnement ? La loi Sapin s'applique-t-elle aux intermédiaires qui vendent des mots clefs ? Par Me Arnaud Diméglio

-> Les contrats de positionnement sur l'internet

http://www.droit-technologie.org/1_2.asp?actu_id=631&abonne_id=22

Même thème, même auteur.

-> Clotilde de Mersan : Yahoo à l'heure du payant

<http://www.neteconomie.com/perl/navig.pl/neteconomie/infos/article/20020819114455>

Clotilde de Mersan explique et commente la stratégie de Yahoo! pour les mois qui viennent.

-> Top 10 Strategies for Boosting Link Popularity

http://www.searchengineguide.com/wi/2002/0809_wi1.html

Les 10 "killer strategies" pour optimiser votre site web

-> Building a Better Search Engine

<http://www.pcworld.com/news/article/0.aid.103676.00.asp>

Les nouvelles technologies de NTT autour des outils de recherche.

-> Using Layers as an SEO Technique

http://www.searchengineguide.com/whalen/2002/0725_jw1.html

Les "layers" sont-elles utiles pour optimiser un positionnement ?

-> Pages dynamiques

http://developpeur.journaldunet.com/tutoriel/php/020407php_urlrewriting.shtml

Comment réécrire des urls dynamiques pour qu'elles soient lisibles par des moteurs de recherche.

-> Top Meta Tag Myths

<http://pandia.com/optimization/SEO-metatags.html>

Quelques questions (et quelques réponses) sur les balises Meta.

-> Preparing Your Site for the Search Directories

<http://www.pandia.com/post/014-3-directories.html>

Comment faire un site "Directory friendly"

-> 131 (Legitimate) Link Building Strategies

<http://sewatch.com/searchday/02/sd0711-linktips-long.html>

Comment optimiser son indice de popularité en 131 leçons...

-> Cloaking - Blessing or Curse? Curse, of Course!

<http://www.promotionbase.com/article/756/33>

Un article sur le cloaking : pour et contre s'expriment...

Merci pour votre lecture... Pour toute suggestion : oa@abondance.com