

## ZONE ABONNÉS : Archives de la lettre "R&R"

### ■ Offre "Privilège" réservée aux abonnés Abondance

Trois (bonnes j'espère ;-)) nouvelles pour les abonnés Abondance :

- Le 28 mai aura lieu à Lyon un séminaire / journée d'information sur le référencement payant. Son nom : "Référencement : vers le tout payant ?". Ce séminaire est organisé conjointement par Abondance et l'association Doc'Forum.

A cette occasion, TOUS les outils de recherche francophones majeurs (Altavista, Espotting, Google France, Inktomi, Looksmart France, Lycos France, Nomade.fr, Overture, Voila et Yahoo! France) seront là pour nous expliquer leurs stratégies et leurs premiers sbilans dans ce domaine. Seront également présents tous les acteurs majeurs du référencement en France (1ère Position, @position, Ad'oc, Adcom, CVFM, E-Wave, Netbooster, Référencement Pro, Snarx et Virtuel Pub) pour nous expliquer comment elles ont intégré ces offres dans leurs propositions et leurs premiers bilans sur celles-ci.

Bonne nouvelle : Les abonnés au site Abondance bénéficient de 20% de réduction pour toute inscription ferme effectuée avant le 1er avril 2002 : 216,40 €HT au lieu de 270,50 €HT, soit 54,10 €HT (354,87 FHT, soit 424,42 FTTC) d'économie. Dépêchez-vous, passé cette date du 1er avril (ce n'est pas un poisson), la réduction ne sera plus que de 10%. Les places étant limitées, réservez au plus vite à l'adresse :

<http://abonnes.abondance.com/28mai.html>

- Le 20 mars sortira une étude Abondance très complète (plus de 100 pages) sur la recherche d'information sur le web francophone : comment les internautes francophones recherchent-ils l'information ? Sur quels outils de recherche ? Avec quels mots clés ? En minuscules, en majuscules ? Combien de pages de résultats consultent-ils sur les outils de recherche ? Etc. Cette étude a été réalisée en collaboration avec une société spécialisée dans la réalisation d'étude statistiques et a été effectuée sur la base d'une population représentative de la population française connectée à l'Internet. Là aussi, les abonnés payants au site Abondance auront droit à une remise substantielle. Vous devriez recevoir prochainement un message plus précis concernant cette étude ! Patience...

- Très bientôt, enfin, les abonnés payants au site Abondance bénéficieront d'applicatifs très pointus développés spécifiquement pour eux. Les applicatifs actuels

(<http://audit.abondance.com/>) seront alors disponibles sous deux formes : - Une version "light" accessible gratuitement.

- Une version "Pro" accessible uniquement dans la zone "abonnés", et plus poussée que les outils actuellement en ligne.

Plus d'infos très bientôt également !

Merci de votre fidélité au site Abondance !

### ■ Espotting : un acteur incontournable du CPC

La société Espotting (<http://www.espotting.fr/>) est devenue en quelques semaines un acteur incontournable du monde de la recherche d'information sur le Web francophone. Elle est aujourd'hui leader des "liens sponsorisés" en Europe et notamment sur le Web francophone grâce à ses accords avec Yahoo! France, Lycos, Hotbot, Altavista, Nomade, Copernic, Club-Internet, Netscape et bien d'autres... Pour l'instant, plus de 2,2 millions de mots clés et d'expressions ont déjà été achetés sur l'outil par plus de 400 annonceurs différents, en 11 semaines d'exploitation, multipliant ainsi par un facteur 4 les objectifs initiaux... Pour l'instant, parmi les "majors" de la recherche d'informations, seuls Voila, MSN et AOL France n'affichent pas encore les liens sponsorisés d'Espotting. AOL France devrait, lui, passer par Overture pour ce type de fonctionnalités dès que cette société s'implantera en France. Pour les deux restants (Voila et MSN), le suspense reste entier... En tout cas, le marché du CPC (achat aux enchères de mots clés et rémunération au clic), Overture (aux Etats-Unis) et Espotting (en Europe) se partagent le monde avant l'arrivée, depuis quelques jours, de Google sur ce marché.

Espotting a également signé des accords avec des sociétés de référencement comme Takezo, @position, Agence Virtuelle et Netbooster pour revendre ses mots clés de façon indirecte (stratégie différente de celle d'Overture qui travaille plutôt en direct pour sa part). D'autres accords devraient également être signés et annoncés dans les semaines qui viennent avec d'autres sociétés de référencement.

Bref, tout cela nous a poussé à aller à la rencontre des responsables de la société pour qu'ils nous expliquent de façon approfondie leur fonctionnement. En route pour un voyage dans le coeur de la bête ;-)

## Base de données en ligne et Suggestions

~~~~~

Tout d'abord, un point important : un nouveau site web devrait voir le jour pour Espotting UK d'ici la fin du mois de mars. Quelques semaines plus tard, ce sera au tour du site français d'être "relooké". Premier objectif : apparaître clairement comme un fournisseur de positionnement publicitaire et non pas comme un moteur de recherche à part entière (difficile en effet de donner l'impression à ses clients que l'on est concurrent, GoTo avait eu le même problème et cela avait été l'une de ses motivations pour changer de nom et devenir Overture). Autre objectif du nouveau site : fournir beaucoup plus d'informations en ligne, des guides d'achat, et également un nouvel outil de suggestion de mots clés : le "Search Term Generator".

Cet outil est très important : il sera assez semblable à ce que propose déjà Overture US (<http://inventory.overture.com/d/searchinventory/suggestion/>) : lorsque vous tapez un mot clé, l'outil vous indiquera les expressions contenant ce terme. Exemple : vous tapez "voiture" et l'outil vous proposera "location voiture", "achat voiture", "parking voiture", "location parking voiture", etc. Pour chaque expression, le nombre de fois où celle-ci a été demandée sur les quatre dernières semaines est également donné. Les expressions proposées seront affichées de deux couleurs différentes selon qu'elles sont déjà "achetées" (une enchère est proposée dessus) ou non. Il sera alors facile de choisir si l'on veut la prendre en compte ou non. Des boîtes à cocher permettront d'opter pour telle ou telle expression, puis deux choix seront proposés : générer un fichier Excel ou inclure les expressions dans son compte client pour créer une nouvelle campagne.

Mais ce n'est pas tout, car l'outil en ligne devrait également proposer des mots clés connexes. Exemple : vous tapez, là encore, "voiture", et il vous sera proposé "automobile", "constructeur", "amortisseur", etc. Dans un premier temps, cette proposition se ferait sur la base d'un thésaurus de termes et d'expressions, mais il est tout à fait possible que, par la suite, les propositions soient le fruit d'un outil automatique qui génère des mots clés connexes à la volée. Pour avoir une idée de ce type d'outils, vous pouvez aller voir le site d'Exalead (<http://www.exalead.com>) qui fait cela très bien. Ceci dit, Exalead ne sera pas obligatoirement l'outil utilisé par Espotting, ne me faites pas dire ce que je n'ai pas dit ;-)

L'outil devrait être en ligne rapidement (d'ici la fin du mois de mars) pour le site anglais. Pour la France, il faudra attendre un peu plus, et la stratégie de diffusion n'est pas encore définie en ce qui le concerne : sera-t-il ouvert à tous (comme en Grande-Bretagne), réservé aux clients ayant déjà acheté au moins un mot clé sur Espotting ou aux partenaires (par exemple les sociétés de référencement) ? Pour l'instant, rien n'est fixé sur ce point...

## Conditions syntaxiques

~~~~~

Lorsqu'un webmaster va sur le site d'Espotting, il a la possibilité de créer une "campagne" pour un mot clé ou une expression donné. Une "campagne", dans le "jargon" d'Espotting, équivaut à un ensemble "titre + descriptif + url" qui seront associés au mot clé "acheté" (i.e. pour lequel on est prêt à proposer une enchère).

Signalons, tout d'abord, quelques particularités syntaxiques importantes lors de l'achat du mot clé :

- Casse des mots : sans importance. Si vous achetez le mot clé "moteur", vous achetez également, de façon transparente, "MOTEUR" et "Moteur".

- Accents : sans importance. Si vous achetez "électricité", vous achetez également "electricite" ou "electricité" sans coûts supplémentaires.

- Pluriels : pas pris en compte. Si vous désirez acheter "pâtisserie" et "pâtisseries", vous devez mettre une enchère sur les deux mots qui sont considérés comme différents par Espotting. Si vous achetez "chien", vous n'achetez donc pas par défaut "chiens".

- Association de mots clés : pas pris en compte. Si vous achetez "location", votre annonce ne sortira pas sur "location voiture". De même, si vous avez acheté de façon isolée les mots clés "location" et "voiture", vous n'apparaîtrez pas sur "location voiture" si vous n'avez pas demandé de façon expresse cette suite de termes.

- Acronymes : les points sont enlevés. Exemple : "c.g.t." est équivalent à "cgt". En revanche, "c.g.t" n'est pas équivalent à "c g t" puisque les espaces sont considérés comme des séparateurs. En revanche, "saint-etienne" sera équivalent à "saint etienne".

- "Stop words", "skip words" ou "mots clés vides" : pas pris en compte. Par exemple, si vous avez acheté l'expression "cd madonna", votre annonce sortira également sur l'expression "cd de Madonna". Voici la liste des "mots vides" qui ne sont pas pris en

compte par Espotting : à

À

â

Â

ä

Ä

è

È

é

É

ê

Ê

ë

Ë

ï

Î

î

Î

ô

Ô

ù

Ù

ü

Ü

û

Û

ü

Ü

ç

Ç

^

"

,

,

,

!

"

(

)

,

-

.

:

;

?

+

a propos de

a travers

a

achats

achete

achete

acheter

afficher

afin de

ai

aimerai

aimerais

alle

aller

alors

apres

arrive

arriver

assez

au sujet de

au travers de

au

aucun

aucune

aussi

autre

autres

aux

avant  
avec  
avoir  
beaucoup  
besoin  
bien que  
ca  
ce  
cela  
cependant  
certain  
certaine  
certaines  
certains  
cette  
cherche  
chercher  
chez  
chose  
choses  
comme  
comment  
comprend  
comprendre  
concernant

connaitre  
constitue  
constituer  
correspond  
correspondent  
correspondre  
d  
dans  
davantage  
de l information  
de la part  
de l'information  
de  
decouvrir  
decouvrez  
deja  
dernier  
derniere  
dernieres  
derniers  
des  
desquelles  
desquels  
detecte  
detecter  
devoir  
devraient  
devrait  
dire  
dis  
dit  
dite  
dites  
donc  
donne moi  
donnez moi  
dont  
du  
duquel  
elle  
elles  
en ligne  
en tant que  
en  
entre  
envoye  
envoyer  
es  
est

est-ce-qu  
est-ce-que  
et  
etes  
etre  
eu  
eux  
fait  
faite  
faits  
il  
ils  
info  
information  
informations  
infos  
interessant  
interessante  
interessantes  
interessants  
interessant  
interessante  
interessantes  
interessants  
j  
je commande  
je ne peux pas  
je peux  
je reserve  
je  
l  
la  
laquelle  
le  
lequel  
les  
lesquelles  
lesquels  
leur  
leurs  
localise  
localiser  
lui  
m  
ma  
mais  
malgre  
me  
mes  
mettre  
moi  
mon  
montre  
montree  
montrees  
montrent  
montres  
ni  
nos  
notre  
nous  
obtenir  
obtenir des informations  
obtenu  
obtenue  
obtenues  
obtenus  
obtiens  
obtient  
on  
on line  
online  
on-line  
or  
ou

par  
parle  
parlent  
parler  
personne  
peu  
peut etre  
plus  
pour  
pourquoi  
pourrai  
pourraient  
pourrais  
pourrait  
pourtant  
prendre  
probleme  
problemes  
procure  
procurer  
qu  
quand  
quant a  
que  
quel  
quelle  
quelles  
quelqu un  
quels  
qu'est  
qui  
quoi  
recherche  
rechercher  
recherches  
regarde  
regarder  
renseignement  
renseignements  
s  
sa  
saisir  
savoir  
se  
sera  
ses  
si  
signifie  
signifient  
signifier  
site  
sites  
solution  
solutions  
sommets  
son  
sont  
souhaiterai  
souhaiterais  
sous  
suis  
suivre  
sur  
t  
ta  
te  
tel  
telle  
telles  
tels  
tes  
ton  
tous  
tout le monde  
toutefois

toutes  
tres  
trop  
trouve  
trouver  
tu  
un  
une  
vendre  
venir  
veulent  
veut  
veux  
vos  
votre  
voudrai  
voudrais  
vouler  
voulez  
vouloir  
vous  
vous

Notons également que le titre de l'annonce ne doit pas faire plus de 40 caractères, le descriptif est limité à 180 caractères (espaces compris) et l'url à 160 caractères. Enfin, un mot clé ou une expression ne doivent pas contenir plus de 64 caractères, espaces compris.

#### Validation d'une campagne par Espotting

~~~~~

Une fois que vous avez en main toutes ces données, vous êtes prêt à "acheter" un mot clé sur Espotting. La procédure est simple et nous vous laissons la tester (<http://www.espotting.fr/>). Il nous semble, en revanche, plus intéressant de voir ce qui se passe après, dans le délai de 48 heures entre la demande de prise en compte d'un webmaster et la validation de la part d'Espotting de la "campagne" demandée...

Une "campagne" est donc constituée, pour un mot clé donné, d'un titre, d'un descriptif et d'une url qui sont proposés par le webmaster. Espotting va alors travailler sur plusieurs plans pour valider la forme et le fond de la campagne.

Tout d'abord (c'est ici le travail de l'équipe éditoriale d'Espotting France qui est constituée de 4 personnes, bientôt 5), la première tâche est de valider le fond : le mot clé demandé est-il pertinent par rapport au site qui le demande ? L'équipe éditoriale se connecte donc sur le site en question, l'analyse en profondeur et valide cette pertinence (ou non...).

Les critères d'acceptation sont assez "classiques" (un guide traitant de façon précise de ces critères sera d'ailleurs disponible sur la future version française du site) : il s'agit en tout premier lieu de la pertinence du site par rapport au mot clé demandé. La langue est également importante : un site entièrement en anglais aura du mal à demander un terme francophone. Mais il peut y avoir des exceptions : par exemple, pour l'expression "location voiture londres", le site en anglais d'un loueur de véhicules londonien pouvant apporter une information intéressante pourrait éventuellement être accepté... Dans ce cas, une mention "site en anglais" sera ajoutée au début du descriptif. Enfin, Espotting accepte très peu de sites perso ou de sites à la charte graphique datant du paléolithique inférieur du web... ;-). On peut, en fait, considérer que la politique éditoriale d'Espotting correspond à l'aggrégation de celles de ses affiliés (Yahoo! France, Nomade.fr, etc.).

Il est important de noter que le site demandeur doit réellement traiter du domaine demandé, sans équivoque possible. Ainsi, pour le mot clé "environnement", un site entièrement consacré à ce domaine n'aura aucun problème à être "recevable". En revanche, le site d'un candidat aux élections présidentielles, qui parle de façon accessoire d'environnement à l'intérieur du programme du candidat, ne sera pas accepté, même si l'url proposée pointe sur la page qui parle du programme environnemental en question. Ce point est important et de nombreux cas sont souvent "limites", ce qui nécessite un traitement au cas par cas. Par exemple, un site comparateur de prix comme Kelkoo peut acheter le mot clé "nokia" s'il compare spécifiquement des téléphones portables Nokia et s'il fournit comme url une adresse qui pointe à tout le moins vers sa partie "téléphones portables". Idem pour le terme "Astérix", aujourd'hui acheté par trois librairies en ligne proposant un lien pointant directement vers la page proposant d'acheter les albums d'Astérix. Espotting demande, de plus, toujours à l'annonceur de créer une campagne spécifique par nom de marque, pour éviter tout problème.

A cela s'ajoute toute la partie législative, notamment en France : pas de possibilité d'acheter des noms de médicaments, etc.

Deuxième phase : validation de la forme : dans ce cas, si le mot clé est considéré comme pertinent, ce sont le titre et le descriptif qui sont acceptés ou refusés par l'équipe d'Espotting.

En cas d'acceptation globale, pas de problèmes. Sinon, les informations sont envoyées au "service client" (2 personnes chez Espotting France pour l'instant) qui contacte le webmaster ayant effectué la demande initiale et lui explique pourquoi la campagne est refusée ou pourquoi une modification est demandée. A priori, cela semble bien se passer dans l'immense majorité des cas pour l'instant.

Une fois la campagne acceptée, elle est donc validée et elle apparaît sur tous les sites affiliés d'Espotting. Le webmaster peut ensuite la modifier s'il le désire. Jusqu'à maintenant, si cela arrivait, la campagne était totalement bloquée (l'annonce n'apparaissait plus, même dans son ancienne version) en attendant la validation d'Espotting pour la nouvelle version demandée. D'ici quelques semaines, et de façon plus logique, l'ancienne version, qui a déjà été validée, continuera à s'afficher tant que la nouvelle n'a pas été validée.

#### Flux entrant et sortant ~~~~~

Cette procédure représente une première façon de travailler pour Espotting. On parle généralement de flux "entrant" dans ce cas (le webmaster propose un mot clé qui est validé ou non par Espotting).

Mais il ne s'agit pas là de l'unique façon de fonctionner de l'équipe éditoriale, loin de là. En effet, Espotting est bien souvent force de proposition pour de nouveaux mots clés connexes à l'activité du client. On parle alors de "flux sortant". Lorsqu'un gros client ou un gros prospect, notamment, demande des chiffres plus précis sur son domaine d'activité, Espotting travaille sur une liste assez complète de termes et d'expressions définissant au mieux l'activité professionnelle du client, tout en prenant garde de ne proposer que des suites de mots clés réellement saisis par les internautes (sinon, quel est l'intérêt, à la fois pour le client et pour Espotting ?). Ce document est ensuite complété avec le nombre de requête et le taux de clics générés sur les semaines précédentes, lorsque ce dernier chiffre est disponible.

Ce type de travail est effectué tous les jours pour des prospects, mais il est également effectué pour des webmasters ayant acheté des mots clés dans les jours précédents : Espotting propose donc souvent d'autres termes à ses clients, de façon spontanée. Dans ce cas, il s'agit de "flux sortant" induit par le "flux entrant".

#### Optimisation des campagnes ~~~~~

Autre métier d'Espotting : l'optimisation des campagnes existantes, domaine sur lequel un accent plus particulier est mis en ce moment. En effet, un site qui propose des loteries en ligne et une petite partie traitant d'horoscope pourrait acheter, par exemple, les mots clés "loterie" et "horoscope" mais proposer pour les deux le descriptif "devenez milliardaire grâce site Tartempion". Si ce descriptif est compréhensible pour la partie "loterie" du site, elle ne l'est pas pour l'horoscope. Dans ce cas, l'annonce ne sera pas cliquée. Ce mot clé doit donc être l'objet d'une campagne à part entière pour le même site, avec son propre titre, son propre descriptif et sa propre url. En effet, plus une campagne sera en adéquation avec le mot clé acheté, plus le taux de clic augmentera et plus le client sera satisfait. Acheter 200 mots clés ou expressions et leur associer un descriptif identique pour tous n'est pas efficace. On peut même imaginer que la situation idéale est celle qui associe une campagne spécifique à chaque mot clé ou expression achetés. Pas toujours facile, on est bien d'accord, si vous êtes acheteur de milliers de termes (eh oui, cela existe...).

#### Filtrages ~~~~~

Notons également que des travaux importants sont en cours sur les filtrages d'adresses IP. Tout comme un internaute français ne voit pas les résultats d'Overture lorsqu'il va sur Yahoo! US, car son adresse IP "française" est détectée par l'outil de recherche, il sera également possible, prochainement, pour les outils de recherche affiliés d'Espotting de demander un filtrage sur tel ou tel pays. Par exemple, on peut imaginer que si un français voit les liens sponsorisés sur Yahoo! France, ce ne sera pas le cas d'un Belge ou d'un Suisse sur le même outil de recherche avec les mêmes mots clés.

Mais l'affilié Espotting peut déjà aller assez loin actuellement en termes de filtrages : il peut demander notamment un filtre familial (pas de sites pornos venant de chez Espotting), mais également un filtre basé sur une "black list" qu'il fournit (pas de sites untel ou untel comme annonceur car ils ont essayé, par exemple, de spammer l'outil de recherche par le passé). L'affilié peut également demander qu'un concurrent n'apparaisse pas : par exemple, Nomade (Tiscali) pourrait ne pas vouloir que le site de Wanadoo apparaisse en premier sur l'expression "fournisseur d'accès internet". Autre possibilité : un



filtrage sur certains mots clés que l'affilié peut refuser, pour une raison ou pour une autre.

#### Chiffres

~~~~~

Nous avons déjà dit, au début de cet article, que plus de 2,2 millions de mots clés ou expressions avaient déjà été achetés sur Espotting par plus de 400 annonceurs. Il est intéressant de noter qu'au mois de février, seuls 25 mots clés ont générés près de 37% du CA d'Espotting France. Les voici (avec les pourcentages de chacun d'eux ramenés à 100) :

sexe (12,41%)  
referencement (9,81%)  
voyages (8,42%)  
immobilier (8,12%)  
sex (6,66%)  
jeux (6,63%)  
emploi (5,85%)  
horoscope (5,27%)  
musique (3,61%)  
webcam (3,35%)  
voyage (3,06%)  
porno (2,94%)  
automobile (2,63%)  
casino (2,58%)  
informatique (2,32%)  
astrologie (2,21%)  
sexe gratuit (2,14%)  
argus (2,05%)  
billet avion (1,61%)  
maison (1,51%)  
pornographie (1,48%)  
rencontre (1,44%)  
charme (1,33%)  
voiture (1,32%)  
dvd (1,23%)

Notons que le mot clé "référencement" représente plus un champ d'expérimentations pour les référenceurs qu'un mot clé réellement demandé par les internautes sur le Web... Les intéressés comprendront l'allusion ;-))

#### Conseils

~~~~~

Pour terminer cet article, voici quelques conseils pour créer une campagne efficace sur Espotting :

- N'essayez pas de mettre trop d'infos dans votre descriptif. Soyez précis et explicatif avec le moins de mot possible. L'internaute doit comprendre en quelques secondes pourquoi l'annonce proposée correspond parfaitement au mot clé pour lequel elle s'affiche.
- Créez (lorsque cela est possible) une campagne spécifique pour chacun des mots clés ou expressions achetés, c'est souvent beaucoup plus efficace.
- Proposez (comme le fait Amazon par exemple : testez sur le mot clé "Astérix") un libellé d'url avec le nom de domaine simple (<http://www.tartempion.fr>) mais redirigez l'internaute vers une adresse spécifique et pertinent par rapport au mot clé demandé (<http://www.tartempion.fr/produits/epiderie/sel-et-poivre.html>). Vous y gagnerez à la fois en notoriété globale pour votre site et en ne perturbant pas l'internaute sur une url trop longue et complexe.

A vous de jouer maintenant pour créer vos propres campagnes !!

#### ■ Mirago : un nouveau moteur de recherche francophone

L'année 2002 voit l'arrivée d'un nouvel entrant dans le paysage des outils de recherche francophones. Mirago (<http://www.mirago.fr/>), un nouveau moteur, est actuellement disponible en version Beta et mérite que l'on s'y attarde, car il propose bon nombre de fonctionnalités originales. Ce nom (choisi car c'est un nom simple et facilement mémorisable dans la plupart des langues) ne vous est d'ailleurs pas inconnu si vous scrutez les logs de vos sites web, car son robot "butine" le Web francophone depuis plusieurs mois...

#### Historique

~~~~~

Au départ, Mirago est issu d'une société, nommée Compsoft (<http://www.compsoft.co.uk/>), basée en Grande-Bretagne et spécialisée dans les bases de données depuis 1979. De par son travail dans ce domaine, la société a eu à traiter de nombreuses problématiques tournant autour des index de données, puis, de façon naturelle, du Web à partir des années 90. L'idée vint alors aux dirigeants de l'entreprise de mettre en place un projet autour du Web et de la recherche d'information, qui devint une extension de Compsoft, puis une entreprise à part entière en 1999. Le premier projet vit le jour fin 1998 en Grande-Bretagne sous la forme d'un moteur de recherche (<http://www.mirago.co.uk/>). Cet outil remportant un succès important, il fut décidé d'étendre la "zone géographique cible" à la France (<http://www.mirago.fr/>), à l'Espagne (<http://es.mirago.com/>), où le site est, là aussi, encore en version Beta pour quelques jours, et à l'Allemagne, d'ici les mois d'avril ou mai 2002.

Actuellement, tous les sites sont gérés depuis la Grande-Bretagne, tous les serveurs se trouvent dans la banlieue de Londres et toutes les équipes techniques et éditoriales également. Cependant, cette situation pourrait être amenée à bouger, en fonction du succès rencontré par les outils dans les pays où ils s'installent actuellement, sous la forme d'une implantation, notamment commerciale (de la même façon que peuvent le faire des sociétés comme Espotting, Overture ou Google, par exemple) dans les mois qui viennent.

En Grande-Bretagne, le site Mirago traite actuellement 1,5 millions de requêtes par jour et compte 380 000 utilisateurs. Sa grande et principale originalité est qu'il propose trois approches différentes : globale, régionale et thématique. Voyons cela de plus près...

#### Recherche Globale ~~~~~

L'index de Mirago France compte à l'heure actuelle 15 millions de pages (50 millions en Grande-Bretagne et 5,5 millions en Espagne) en langue française "prélevées" sur 500 000 sites. Les serveurs sont donc situés dans la banlieue de Londres et sont connectés directement au backbone Internet anglais, ce qui fournit au moteur des accès plutôt rapides...

Sur la page d'accueil, une interface minimaliste "à la Google", propose de saisir un ou plusieurs mot(s) clé(s). L'opérateur par défaut est un ET : moteur renault est donc identique à moteur ET renault.

Un lien "Plus d'options" est proposé en haute à droite du formulaire de recherche. Il propose une interface intéressante de recherche avancée. Une recherche sur la date de dernière modification des documents (de 1 an à 1 mois) est fournie. La possibilité de chercher, bien entendu, dans le Web de façon globale, mais également dans des index plus précis (pages avec images, avec sons, vidéo, multimédia, etc.) est également proposée. L'option "mots liés" cherche les documents dans lesquels les mots de la recherche sont proches les uns des autres. Cependant les mots peuvent apparaître dans n'importe quel ordre.

Plus originale est la possibilité de classer les résultats par pertinence (tous les critères de pertinence de Mirago sont alors pris en compte : nombre d'occurrence du mot demandé dans la page, indice de popularité, etc.) ou par indice de popularité (ce critère étant pris en compte de façon strictement quantitative par Mirago, ce sont alors les pages web - contenant le mot demandé - qui sont le plus "pointées" qui apparaîtront en premier). Autres possibilités : un tri par nombre d'images (plus il y a d'images dans la page en question, mieux elle est classée) ou par date (les plus récentes seront affichées en premier).

Mirago propose une possibilité d'utiliser une syntaxe d'interrogation booléenne. Dans ce cas, utilisez l'option "Requête booléenne" du menu déroulant nommé "Rechercher" sur la page d'accueil de la recherche avancée. Cependant, peu d'opérateurs semblent disponibles : des essais avec "ET", "OU", "OR", "SAUF" n'ont pas été concluants, loin de là... Le moteur envisage de modifier les options de sa recherche booléenne, c'est la raison pour laquelle il y a peu d'informations disponibles sur la question sur le site et certainement pourquoi les tests réalisés n'ont pas donné de résultats très fameux. D'autre part, il n'est pas possible de limiter la recherche à un site ou un hôte (comme le propose Altavista, par exemple, avec son option "host:") depuis le formulaire de recherche.

Les requêtes sont, d'autre part, traitées de la façon suivante :

Ordre des mots : pas d'importance, paris dakar donne le même résultat que dakar paris

Traitement minuscules/MAJUSCULES : identique : IBM, ibm et Ibm sont des requêtes similaires.

Lettres accentuées : identique : electricite, électricite et électricité sont des requêtes similaires.

Il est également possible de saisir une phrase en langage naturel ("Où puis-je trouver le palmarès des Jeux Olympiques de Salt Lake City ?"), une technologie développée en interne chez Mirago (l'héritage des bases de données...) va tenter de comprendre ce que vous demandez et d'y répondre de la meilleure façon possible.

L'outil sait également détecter les pluriels (chien trouve chiens, chiens trouve chien), les verbes conjugués (aimer trouve aimé, aimé trouve aimer), etc. Bref, un certain nombre de fonctionnalités qui tentent de rendre le moteur un peu plus "intelligent" que la moyenne des outils de recherche...

Deux options sont enfin proposées sous la forme de boîtes à cocher : le fait de résumer le site (si la boîte est décochée, la description textuelle du lien n'est pas proposée sur la page de résultat) et la fonction de clustering (un seul lien affiché par site), ce qui évite qu'un ou deux sites web trustent la première page de résultats.

Notons qu'un lien "Préférences" (<http://www.mirago.fr/fr/preferences.asp>) permet de "customiser" l'utilisation du moteur par défaut (recherche simple ou avancée, région de prédilection, filtre familial, présentation des liens affichés, etc.). Très simple mais bigrement efficace...

#### Pages de résultats

~~~~~

Point intéressant sur la page de résultats : est proposé en premier, s'il existe, le site web dont le nom de domaine correspond à la requête (la demande du mot clé "abondance" propose "www.abondance.com"), avec une option "Nom de sites similaires" (abondance.net, abondance.fr, abondance.tm.fr...) si nécessaire. Simple et souvent très utile. On se demande bien pourquoi les autres outils de recherche ne le proposent pas de cette façon, d'ailleurs...

Ensuite, la présentation des résultats est assez classique (le résumé affiché correspond au contenu de la balise Meta Description, ou au début du texte visible en cas d'absence de cette balise), avec quelques originalités :

- La possibilité d'envoyer ces liens par mail (en haut de page)
- La fonction "voir liens" fournira toutes les pages qui ont mis en place un lien vers la page proposée (équivalent du "link:" sur Altavista ou Google). Simple, pratique et pas si courant que cela...

#### Critères de pertinence

~~~~~

Le moteur de recherche Mirago utilise un algorithme de pertinence assez "classique", utilisant des critères aujourd'hui pris en compte par la majorité des outils :

- Titre de la page.
- Balises Meta keywords et Description (poids faible).
- Texte visible, avec une prédominance pour les premiers caractères et la mise en exergue (gras, taille de la police utilisée, texte d'un lien, etc.).
- Le nombre d'occurrences du mot clé demandé dans la page.
- L'indice de popularité, calculé sur un plan strictement quantitatif (la page qui a le plus de liens vers elle est dotée du plus gros indice de popularité).

La densité du texte et l'indice de popularité sont aujourd'hui les deux critères de pertinence "clés" sur Mirago.

L'équipe technique du moteur a également mis au point pas mal de techniques de détection du spam (texte blanc sur fond blanc, répétitions abusives de mots, termes dans les balises meta n'ayant rien à voir avec le contenu de la page, etc.). A bon entendeur, salut... ;-)

Mirago n'a pas, a priori, de réels critères bloquants pour indexer des pages web. Il prend en compte les pages PHP, ASP, CFM, même avec passage de paramètres, le texte contenu dans les pages HTML proposant du Flash, etc. Il accepte les fichiers Robots.txt ainsi que la balise Meta "Robots". En revanche, le moteur n'indexe pour l'instant que du HTML, pas de formats comme le PDF, XLS, DOC, etc.

#### Soumission d'un site

~~~~~

Mirago propose une page spécifique pour soumettre un site, à l'adresse : <http://www.mirago.fr/fr/region/ajout.asp>

Le délai de prise en compte est d'environ 4 à 5 semaines. Mirago indique qu'il n'est pas nécessaire de soumettre plusieurs pages d'un même site, seule l'adresse de la page

d'accueil suffit. Les autres pages seront indexées "naturellement" par le spider (qui s'appelle PierreLeRobotMirago) par la suite. Le spider descend 6 niveaux d'arborescence, ou plutôt jusqu'à 6 clics à partir de la page d'accueil pour indexer les pages web internes du site à crawler, mais ce chiffre pourrait augmenter dans un avenir proche. Pour l'instant, aucun projet de référencement payant n'est en cours de réflexion... D'autre part, il n'y a pas de limite au nombre de pages soumises par jour.

Comme on peut le voir sur la page de soumission, Mirago propose de régionaliser et thématiser le site soumis, afin de prendre en compte des informations pour ses deux autres fonctionnalités, proposées sur des onglets sur la page d'accueil : "Région" et "Secteur"

#### Régionalisation - Thématisation

~~~~~

On arrive ici à l'une des grandes originalités de Mirago : un moteur de recherche (et non pas un annuaire) également orienté sur les régions françaises et un certain nombre de secteurs d'activité.

Si vous cliquez sur l'onglet "Région", sur la page d'accueil, vous avez la possibilité de mettre en place un filtre sur les 22 régions françaises, depuis l'Alsace jusqu'à Rhône-Alpes (dommage, les Dom-Tom sont oubliés mais ils devraient être bientôt présents). Il est possible d'effectuer une recherche à partir d'une ville, d'un département ou d'une région en France. Les résultats de la recherche régionale sont ordonnés de manière à ce qu'apparaissent en premier les résultats situés dans la ville sélectionnée, puis les réponses "départementales" (i.e les réponses d'autres villes du département), suivies des résultats "régionaux" (sites situés dans d'autres départements de la région).

Si vous cliquez sur l'onglet "Secteur", 7 thématiques sont proposées : Finance, Juridique, Habitat, Agroalimentaire, Médecine, Tourisme et Agriculture. Exemple-type : tapez "sexe" avec la thématique "Médecine" et à part le nom de domaine correspondant ([www.sexe.com](http://www.sexe.com), il y a là un bug que Mirago devra corriger...), toutes les pages proposées ont trait au sexe, mais dans l'univers de la santé... Il existe aujourd'hui 7 thématiques, mais d'autres sont prévues à moyen terme si l'outil de recherche rencontre un franc succès en France.

On voit tout de suite l'intérêt d'un tel outil car, s'il est vrai que les annuaires régionaux et thématiques sont disponibles par dizaines, voire par centaines sur le Web francophone, les moteurs de recherche, indexant des pages web en langue française, sont beaucoup plus rares...

Comment Mirago fait-il pour régionaliser et thématiser de façon automatique les documents ? De quatre façons différentes :

- Lors de la soumission du site, on a vu (ci-dessus) que le webmaster peut donner un certain nombre d'informations permettant de préciser la zone géographique et le domaine d'activité du site ;
- Des interrogations automatiques (Whois) sur les bases de données de Network Solutions et de l'Afnic donnent également de précieuses informations ;
- Lors du crawl du site web, un algorithme tente de détecter une adresse postale et un nom de ville, de département ou de région dans le code des documents indexés.
- Une équipe éditoriale effectue également un travail similaire à la main, classant les sites dans la région et le domaine d'activité adéquat.

Bien entendu, un site peut tout à fait être enregistré dans plusieurs thématiques, voire dans plusieurs régions (cas de filiales). En Grande-Bretagne, il est même proposé un système de type "latitude / longitude", qui permet d'effectuer une recherche, par exemple sur Londres, ou sur les villes autour de Londres. Ce système pourrait être intégré sur Mirago France dans un proche avenir.

#### Business Model

~~~~~

Le site actuel est encore en version Beta pour quelques semaines, voire quelques jours. Le business model de l'activité n'est pas basé sur l'affichage de publicités (vous vous apercevrez vite qu'aucune bannière n'est visible sur les pages de l'outil), mais sur l'achat de mots clés aux enchères, comme le proposent Espotting, Overture ou Google Adwords Select, au CPC (Coût Par Clic) donc. Dans ce cadre, il se rapproche un peu d'un outil comme Misterbot (<http://www.misterbot.com/>), mais qui prendrait en compte son propre index (Misterbot utilisant l'annuaire Open Directory).

Le système d'enchères sur les mots clés n'est pas encore ouvert sur Mirago France, il devrait voir le jour dans les semaines qui viennent, à la fin de la phase Beta d'expérimentation. Mirago ne réfute pas non plus l'idée de vendre sa technologie à d'autres sociétés, notamment dans le cadre de projets intranet. En revanche, la revente

auprès d'autres sites des mots clés gérés par Mirago (et achetés auprès de l'outil par des webmasters) n'est pas à l'ordre du jour.

Pour l'instant, le site démarre petit à petit, sans objectifs commerciaux précis. Ceux-ci seront fixés de façon beaucoup plus avancée d'ici un mois ou deux, quand la phase Beta sera terminée et que la société aura une bonne vision du marché potentiel français.

#### Conclusion

~~~~~

Les nouveaux projets en termes de moteurs de recherche ne sont pas légions, qui plus est sur le Web francophone. Raison de plus pour tester Mirago, qui s'annonce comme un concurrent intéressant et innovant dans la "nouvelle vague" menée aujourd'hui par des outils comme Exalead ou Kartoo, chacun dans des domaines différents. En tout cas, il amène sa propre dose d'innovations souvent originales et basées sur le bon sens, ce qui est gage de succès. Bonne chance à eux donc !

#### ■ Voila : un point sur l'offre "Pôle Position"

L'offre "Pôle Position" de Voila (voir lettre "R&R" de septembre 2001 à l'adresse <http://abonnes.abondance.com/archives/acturech/0109.html>), qui permet de prendre en compte des fonctionnalités de positionnement publicitaire, est proposée par le moteur de recherche de France Télécom (Voila et Wanadoo) depuis le mois d'octobre 2001. Une bonne occasion de faire le point avec Francois-Xavier Husherr, Directeur Adjoint du pôle business de Wanadoo Portails, sur cette offre, pour un moteur qui reste le seul, à l'heure actuelle et parmi les outils majeurs francophones, à ne pas être passé sur le modèle du "CPC" (achat de mots clés aux enchères avec paiement uniquement sur les clics générés), et restant donc fidèle au CPM (achat au forfait et paiement au nombre de liens vus, qu'ils soient cliqués ou non).

**Question : L'offre "Pole Position" de positionnement payant au CPM sur Voila et Wanadoo a été lancée il y a quelques mois de cela. Quel premier bilan peut-on en tirer ? Est-ce que le nombre de demandes entre dans les objectifs visés au départ ? Quel est le taux de progression depuis le lancement de la prestation ?**

Réponse : L'offre a été lancée en octobre 2001 et répond à une forte demande du marché du référencement. Nos annonceurs ont des résultats qui ont été très satisfaisants sur 2001 (en moyenne 10% de taux de clics), et investissent de façon plus conséquente en 2002. Le chiffre d'affaire paru mensuel évolue du simple au triple, ce qui est normal pour un produit en phase de croissance.

**Q : Quel est le profil-type du site profitant de cette offre ? Avez-vous quelques exemples ?**

R : L'offre Pôle Position est commercialisée par Wanadoo Régie et est proposée à ses annonceurs qui sont pour la plus grande partie des grands comptes mortars des secteurs de l'assurance, de la banque, de la grande distribution ou encore de la vente par correspondance. Beaucoup ont sur leur site un dispositif de vente en ligne et recherchent une audience qualifiée.

**Q : Quel est le budget moyen investi par vos clients dans cette offre ? Y a-t-il un minimum d'achat ?**

R : Le budget moyen dépend essentiellement des mots achetés et de l'audience mensuelle (du nombre de requêtes mensuelles) qu'ils représentent. Le seuil minimum d'achat est celui du minimum de facturation de Wanadoo Régie, qui est de 1000 €.

**Q : Aujourd'hui, la tendance serait plutôt au CPC (Overture, Espotting) plutôt qu'au CPM pour ce type de lien acheté. Quelle est votre position sur ce point ? Pensez-vous mettre en place une solution de type Espotting dans les pages de résultats de Voila et Wanadoo, comme la plupart des autres outils de recherche francophones ?**

R : A la vue du succès commercial que semble remporter ce modèle dans d'autres pays, nous étudions actuellement la possibilité de le proposer sur nos deux portails Wanadoo.fr et Voila.fr. Pour ce qui est du mode opératoire, c'est-à-dire outsourcing ou solution interne, rien n'est encore décidé pour le moment.

**Q : Quelle est votre vision aujourd'hui du marché publicitaire sur le Web francophone ? Ce marché va-t-il repartir prochainement ?**

R : Un certain nombre d'indicateurs semblent effectivement montrer que 2002 sur l'année

du redémarrage. Nous pensons pour notre part que le marché connaîtra une croissance de l'ordre de +15%.

**Q : Comptez-vous ouvrir votre base de données des mots clés demandés sur votre outil de recherche, sous la forme d'un abonnement payant pour la consultation, au cours de l'année 2002 ?**

R : Non, cela n'est pas prévu. Nous communiquons volontiers et régulièrement sur le top 100, mais ce n'est pas une ligne de développement majeure pour Wanadoo Portails.

Merci à Francois-Xavier Husherr pour ces réponses !!

## ■ Bruits et chuchotements

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 ;-)

-> Une nouvelle version du site Nomade.fr (<http://www.nomade.fr/>) devrait voir le jour d'ici au mois d'avril. MSN Search (<http://search.msn.fr/>) travaillerait également à une nouvelle interface.

-> Toujours pas de nouvelles de l'offre de soumission payante sur l'annuaire de Lycos France. A priori, il y a fort peu de chances que quelque chose apparaisse d'ici l'été 2002. Wait and see...

-> Un site web (<http://jaccuse.free.fr/>) accuse depuis peu certains éditeurs de l'Open Directory de pratiques peu courtoises, c'est le moins que l'on puisse dire... Les éditeurs mis en cause ont choisi de ne pas répondre à ce qu'ils considèrent comme une provocation. Leur réponse "officielle" à la mise en place de ce site est celle-ci : "Les causes de renvoi des éditeurs dmoz sont décrites dans les directives:

<http://dmoz.org/World/Fran%e7ais/guidelines.html#CEE>

Ni le staff dmoz, ni les metas éditeurs ne communiquent à ce sujet. Les 4 ex-éditeurs qui s'expriment sur le forum jaccuse ont tous été renvoyés pour un des motifs décrits dans les directives, directives qui sont envoyées à chaque nouvel éditeur. L'Open Directory compte aujourd'hui plus de 9 600 éditeurs actifs. 70 000 sites sont répertoriés dans World/Français par le travail bénévole de plus de 400 éditeurs francophones." Fin de citation ;-)

-> Altavista France ne serait pas dans une situation très favorable pour le début d'année 2002 et se heurterait à une forte baisse des rentrées publicitaires (ceci dit, la situation ne touche pas qu'Altavista en ce moment...).

-> L'index d'HotBot France ne semblerait plus mis à jour par Inktomi depuis deux mois.

## ■ En Bref

-> Google indexe de plus en plus les pages web se trouvant sur des serveurs sécurisés (<https://...>) et qui ne serait pas protégées par un fichier robots.txt. Si vous avez des informations se trouvant sur ce type de serveur et que vous ne désirez pas les voir indexées par les moteurs en général et Google en particulier, n'oubliez pas d'utiliser le fichier robots.txt ou la balise Meta "Robots"...

-> Google a mis en place depuis peu un moteur de recherche sur l'actualité à l'adresse <http://news.google.com/>. Il semblerait que la syntaxe de Google s'adapte également à ce moteur, et notamment les fonctionnalités suivantes :

intitle:chirac recherche le terme demandé dans le titre de la dépêche

intext:jospin recherche le terme demandé dans le texte de la dépêche

chirac site:cnn effectue une recherche du mot demandé sur la source d'information indiquée par la fonction "site:".

-> Sur Lycos France, les trois premiers résultats, lorsqu'ils existent, sont toujours issus de l'annuaire (voir les descriptions qui sont bien celles de l'annuaire), mais il n'est plus fait de distinctions graphiques entre résultats "annuaire" et résultats "moteur/Fast".