

## Recherche et Référencement - Numéro 27 -- Mai 2002

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

### *Au sommaire ce mois-ci :*

- > [AOL.FR / Google : Le duel](#)
- > [Description du nouvel outil de recherche d'AOL France : AOL.FR](#)
- > [Cybion, créateur de l'annuaire d'AOL.FR](#)
- > [Exalead : une technologie d'avance](#)
- > [Un point sur Inktomi](#)
- > [Lancement d'Overture France](#)
- > [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](#)

Le contenu de cette lettre est accessible sur la zone "Abonnés" du site Abondance, à l'adresse : <http://abonnes.abondance.com/archives/acturech/0205.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](mailto:oa@abondance.com)

**AOL.FR / Google : Le duel !**[Retour au sommaire de la lettre](#)

Il n'a échappé à personne qu'AOL France a récemment lancé son nouvel outil de recherche, baptisé AOL.FR. Après l'avoir essayé, nous avons été séduits par cette nouvelle approche de la recherche d'information. Et, lorsqu'on parle "recherche d'infos", aujourd'hui, on pense tout de suite "Google". De là à imaginer un "duel au sommet" entre les deux outils, il n'y avait qu'un pas que nous tentons de franchir ce mois-ci...

L'idée est donc de comparer, de façon quantitative, qualitative et fonctionnelle les deux outils pour deux populations différentes d'internautes : les débutants et les confirmés. Nous avons donc à nouveau utilisé la méthodologie qui nous a permis de comparer la plupart des outils de recherche francophones dans une étude sortie en septembre dernier, intitulée "Etude comparative des outils de recherche sur le Web francophone (annuaires et moteurs)" (disponible à l'adresse <http://etudes.abondance.com/>).



Les différents tests comparatifs mis en place dans le cadre de cette étude (toutes les recherches et comparatifs ont été effectués le vendredi 10 mai 2002 pour les deux outils) sont les suivants :

- **Comparatif quantitatif** : combien les index des deux moteurs proposent-ils de pages web en langue française sur certaines requêtes ?
- **Syntaxe d'interrogation** : quelles sont les possibilités pour affiner au mieux les recherches en utilisant des opérateurs booléens adaptés (+, -, ET, OU, SAUF, etc.) ?
- Présence et qualité d'une **zone de recherche avancée** permettant de saisir simplement des requêtes plus complexes, sans connaître obligatoirement les différents opérateurs booléens.
- **Qualité de la présentation des résultats proposés** : lisibilité des résultats, nombre d'informations proposées (taille, date de dernière modification, etc.), fonctionnalités connexes (cache mémoire, etc.).
- **Qualité de l'annuaire** proposé avec le moteur de recherche.
- **Délai d'affichage** des pages d'accueil et de résultat : rapidité d'affichage des pages du site : plus l'affichage est rapide, moins on perd de temps (vieux proverbe internaute).
- **Pertinence des réponses "moteur"** : qualité des 10 premières réponses proposées lorsqu'on saisit une requête.

Pour des raisons pratiques, et surtout pour comparer ce qui est comparable, ce qui n'est pas obligatoirement évident lorsque l'on étudie ces deux outils, à la fois assez proches et assez éloignés dans leur approche de la recherche d'information, nous avons pris en compte les résultats fournis par le site Google France (<http://www.google.fr/>) avec l'option "Pages francophones" cochée, puisque cela correspond à l'index des pages en langue française, comparable à celui d'AOL.FR, géré par l'outil d'Exalead.

**Comparatif quantitatif**

Pour ce test, nous avons saisi 30 mots clés sur chacun des deux outils de recherche et noté le nombre de résultats fournis à chaque fois. Les résultats sont les suivants :


Mot clé		
résultats tournoi ATP	8	6 630
fabricant amortisseur	116	575
vin colmar	1 189	2 410
nestlé environnement	221	3 850
musées paris	189 484	75 100
astérix officiel	3 842	6 140
eye tracking	124	680
photos paris	421 892	384 000
exemples business plan	148	16 600

vidéos humour	50	111 000
virus sircam	117	2 590
location studio paris	91	29 300
téléphone portable santé	2 133	15 200
MP3	267 551	663 000
recette couscous	311	6 520
jeux vidéo	91 898	778 000
warrants	12 098	65 400
sonneries GSM	53	24 300
convertisseur euro	4 034	56 100
hotmail	171 363	440 000
sexe	190 849	1 990 000
cartes voeux	13 789	47 600
fond d'écran	40 097	158 000
webcam	163 742	358 000
immobilier	576 603	439 000
référencement	201 922	153 000
euro	534 267	801 000
bourse	358 041	607 000
loftstory	15 578	16 200
yahoo	309 643	1 030 000

Pour noter ce critère, nous avons ôté du calcul quatre mots clés ou expressions pour lesquels les résultats nous semblaient par trop éloignés sur les deux outils ("résultats tournoi ATP", "vidéos humour", "location studio paris", "sonneries GSM") et regardé l'écart moyen entre les deux résultats. Résultat : AOL.FR donne, dans 4 cas, des résultats supérieurs en nombre, dans un coefficient allant de 1,10 à 2,52 (moyenne : 1,56) par rapport au nombre de liens trouvés fourni par Google (qui est, le plus souvent - rappelons-le - approximatif). Google l'emporte dans 22 cas sur 26 en proposant, en moyenne, 11,56 fois de résultats qu'AOL.FR.

Mais, si l'on additionne le nombre de résultats global (hormis les 4 requêtes exclues, voir paragraphe précédent), Google.fr renvoie en tout 8 115 965 liens, contre 3 571 052 à AOL.FR, soit 2,27 fois plus.

En tenant compte de ces différents résultats, nous avons décidé d'attribuer la note de 10 à Google.fr et de 4 à AOL.FR, la plupart du temps nettement inférieur à son concurrent sur ce critère. Encore une fois, il est très complexe, voire impossible, de vérifier si les nombres fournis par les deux outils de recherche sont exacts ou vraiment TRES approximatifs...

		
<b>Critère quantitatif</b>	<b>4</b>	<b>10</b>

### Syntaxe d'interrogation

Pour ce test, nous avons évalué les syntaxes d'interrogation disponibles dès le formulaire de la page d'accueil (donc pas dans la recherche avancée, qui est prise en compte plus loin dans ce comparatif). Exemple :

- Signe + pour l'opérateur ET
- Signe - pour l'opérateur SAUF
- Guillemets pour la recherche des expressions
- Recherche uniquement sur les titres, le nom de domaine, les balises Meta, etc.

Pour chacun des deux outils, 4 notes étaient possibles : 10 lorsque la syntaxe proposée nous a paru complète, riche et efficace, 6 lorsqu'elle nous a semblé intéressante mais assez limitée (juste ce qu'il faut, sans plus), 3 lorsqu'elle nous a semblé insuffisante et 0 lorsqu'elle était inexistante. Voici les résultats :

		
<b>Syntaxe d'interrogation</b>	<b>3</b>	<b>6</b>

La syntaxe de Google est assez complète, même si elle mériterait d'être encore plus efficace pour arriver à celle d'un Altavista, par exemple. Elle est quand même proche de ce qui se fait de mieux à l'heure actuelle, même si l'on peut noter quelques insuffisances. Côté AOL.FR, il n'existe pas d'aide en ligne, donc difficile de savoir ce qui marche ou pas. Le "+" (ET) semble fonctionner, le "-" (SAUF) aussi, les guillemets également, mais rien n'est indiqué officiellement. Il ne semble pas y avoir de fonctionnalité de recherche sur certaines zones des pages (titre, texte, nom de domaine, etc.) ni sur certaines fonctionnalités spécifiques (nombre de liens vers une page, etc.). Cette absence de signalisation est certainement due au parti-pris imaginé par AOL.FR de trouver une information par approche successive, sans connaître de syntaxe complexe, mais l'utilisateur expérimenté du Web aime bien aller plus vite en utilisant ce type de fonctionnalités...

### Recherche avancée

Nous avons ici tenté de noter la présence d'une zone de recherche avancée sur le site. Celle-ci permet, par l'utilisation de menus déroulants, boutons radio et champs de saisie, de formuler des requêtes complexes sans réellement connaître la syntaxe des opérateurs booléens. D'autre part, elle peut également proposer d'autres fonctionnalités (recherche linguistique, sur les dates de dernière modification des documents, investigation sur des images ou des formats de fichiers précis, etc.).

Comme dans le cas précédent, pour chacun des outils, 4 notes étaient possibles : 10 lorsque la recherche avancée proposée nous a paru très fonctionnelle et efficace, 6 lorsqu'elle nous a semblé intéressante mais, disons, "moyenne" (ce qu'il faut, sans plus), 3 lorsqu'elle nous a semblé insuffisante et 0 lorsqu'elle était inexistante.

Voici les résultats :

		
<b>Recherche avancée</b>	<b>0</b>	<b>6</b>

La recherche avancée de Google est correcte, avec quelques fonctionnalités intéressantes, mais n'est pas ce que l'on fait de mieux sur le Web. AOL.FR ne propose pas de recherche avancée.



### Qualité de la présentation des résultats

Pour ce critère, nous avons pensé qu'il était important de noter deux choses :

- Dans un premier temps, la présentation des informations sur les liens affichés : titre, résumé, url, date de dernière modification, taille, etc.
- Mais également les fonctionnalités connexes qui ont également une certaine importance à nos yeux : cache mémoire (Google), mots clés connexes (AOL.FR), etc.

Comme auparavant, pour chacun des outils, 4 notes étaient possibles : 10 lorsque la présentation des résultats proposée nous a paru très fonctionnelle et efficace, 8 lorsqu'elle nous a semblé intéressante mais n'arrivant cependant pas au niveau des meilleurs, 6 lorsqu'elle nous a semblé "classique" et 3 lorsqu'elle nous a semblé insuffisante.

Voici les résultats :

		
<b>Qualité de la présentation des résultats</b>	<b>10</b>	<b>6</b>

La présentation des résultats d'AOL.FR nous semble aujourd'hui proche de ce qui se fait de mieux sur le Web francophone : les catégories d'annuaires et les mots clés connexes, calculés selon des modes statistiques, sont le plus souvent très pertinents. Google souffre, du coup, un peu de la comparaison, même si ses pages de résultats sont très efficaces, mais les fonctionnalités complémentaires de recherche semblent plus intéressantes sur AOL.FR. A quand la si utile fonction [cache] de Google sur AOL ?

### **Délai d'affichage des pages**

Dans ce comparatif, il nous a semblé important de noter le temps d'affichage de la page d'accueil du moteur et de ses pages de résultats. Il n'y a pas de secret : plus le site s'oriente vers une structure de portail, plus il y a d'infos à afficher (météo, actualité, horoscope, etc.), plus le délai d'affichage sera long. Ce critère prend également en compte le délai entre la saisie de la requête et l'affichage des résultats. La note de 10 correspond à des pages d'accueil et de résultats très rapides d'affichage et à un délai court entre la saisie de la requête et l'affichage du résultat. La note de 8 indique des délais "moyens", mais restant de bonne qualité. La note de 6 sanctionne des temps d'affichage plutôt "limites". Et la note de 3 sert à noter des délais anormalement longs. Voici les résultats :

		
<b>Délai d'affichage</b>	<b>10</b>	<b>10</b>

Nous n'avons pas voulu départager les deux outils de recherche, même si nous aurions préféré une page d'accueil plus "efficace" pour AOL.FR (et même celle spécifiquement allouée à la recherche, disponible à l'adresse <http://recherche.aol.fr/>), mais les performances des deux outils sont quand même sensiblement égales sur ce point. Les deux sont vraiment très rapides...

### **Qualité de l'annuaire associé**

Les deux outils utilisent un annuaire de sites : L'Open Directory pour Google, un annuaire interne, créé sur la base de l'annuaire de Looksmart.fr pour AOL.FR. Le but ici est de noter la qualité (nombre et qualité des sites présentés, etc.) de cet annuaire, et ce pour la zone francophone, bien sûr. Comme auparavant, pour chacun des outils, 4 notes étaient possibles : 10 lorsque l'annuaire nous semble excellent, 8 lorsqu'il est intéressant mais qu'il peut encore être perfectionné, 6 lorsqu'il est "classique" sans plus et 3 lorsqu'il nous a semblé insuffisant. Voici les résultats :

		
<b>Qualité de l'annuaire associé</b>	<b>8</b>	<b>3</b>

L'annuaire d'AOL.FR est de très bonne qualité même si, quantitativement parlant, il n'arrive pas aux 160 000 sites proposés par les annuaires majeurs en France. L'Open Directory, pour sa partie francophone, nous semble, malheureusement, encore un peu "léger" et surtout "inégal" (son mode de fonctionnement particulier, basé sur le bénévolat des documentaliste qui y travaillent) en est certainement la cause).

**Pertinence des réponses**

Bien entendu, ce que désire l'utilisateur d'un moteur de recherche, c'est bien avant toute chose la pertinence des résultats obtenus, ce qui est assez logique...

Pour évaluer ce paramètre, vingt recherches dans différents domaines - loisirs (sport, recettes), business (recherche de viticulteur, technologies d'eye-tracking, business plan), informatique (virus) ou autres (téléphones portable, musées, etc.) - ont été choisies. Voici la liste des mots clés et expressions qui ont servi à notre test :

<b>Demande</b>	<b>Mots clés</b>
Je recherche les résultats des tournois de tennis de l'ATP Tour	résultats tournoi ATP
Je recherche un fabricant d'amortisseur	fabricant amortisseur
Je recherche un viticulteur à Colmar en Alsace	vin colmar
Je recherche la politique de Nestlé en termes d'environnement	nestlé environnement
Je recherche la liste des musées de Paris	musées paris
Je recherche le site officiel d'Astérix	astérix officiel
Je recherche des infos sur la façon dont fonctionne les technologie d'eye-tracking	eye tracking
Je recherche des photos de Paris	photos paris
Je recherche des exemples de business plans pour créer mon entreprise	exemples business plan
Je recherche des vidéos amusantes	vidéos humour
Je recherche la meilleure façon de lutter contre le virus Sircam	virus sircam
Je cherche à louer un studio sur Paris	location studio paris
Je désire avoir des informations sur les risques du téléphone portable pour la santé	téléphone portable santé
Je recherche des fichiers MP3	MP3
Je recherche la recette du couscous	recette couscous
Je recherche l'actualité des jeux vidéos	jeux vidéo
Je cherche ce que signifie le terme "warrants" dans le domaine de la bourse	warrants
Je recherche des logos et sonneries pour GSM	sonneries GSM
Je recherche un outil qui permette de convertir facilement des francs en euros	convertisseur euro
Je recherche le site Hotmail	hotmail

Pour chaque recherche et donc chaque moteur, les mots clés indiqués dans la colonne de droite ont été saisis dans le formulaire de recherche et les 10 premiers liens proposés ont été inspectés et notés. Le barème suivant a été appliqué à chacun des liens évalués :

- Page pertinente par rapport au sujet recherché : **+4 points**
- Bonus si cette page pertinente se trouve dans les trois premiers résultats proposés : **+ 2 points**
- Page pertinente par rapport au sujet recherché, mais pas en français : **+2 points**
- Page contenant un lien vers un site ou une autre page pertinent(e) : **+ 2 points**
- Erreur 404 (page inexistante) : **-4 points**
- Page se trouvant à une nouvelle adresse (message du type : "ce site a changé d'emplacement, pour le trouver, cliquez ici") : **-1 point**
- Doublet (page de contenu identique à une autre proposée dans les 10 premiers résultats, même si elle ne se trouve pas à la même adresse) : **-1 point**
- Page hors sujet : **-2 points**. Ce sont des pages qui n'ont réellement et sans ambiguïté possible aucun rapport avec la requête demandée.
- Page "banale" : **0 point**. Ce sont des pages qui traitent bien, de façon globale, du sujet demandé, mais qui ne sont pas assez précises ou ne contiennent pas assez d'informations par rapport à la demande. Elles sont globalement intéressantes, mais pas assez pour répondre à la question.

Dans tous les cas, nous avons saisi les mots clés dans le formulaire de recherche de la page d'accueil des moteurs sans prendre en compte aucune option proposée (menus déroulants, boutons radio, etc.). Nous avons utilisé l'interface de saisie telle quelle, comme le font une

immense majorité d'internautes. Seule l'option "Pages francophones" a été prise en compte sur Google.fr.

Le moteur qui a eu le meilleur total a reçu la note relative de 10. Une règle de trois nous a permis d'obtenir la notation de son concurrent.

Voici les résultats :

		
<b>Pertinence des réponses</b>	<b>8</b>	<b>10</b>

Rappelons que la note maximale, pour une recherche donnée, est de 46 (10 fois 4 pour 10 liens pertinents + 3 fois le bonus de 2 pour un lien pertinent dans les 3 premiers liens). Google, sur ce test, a obtenu une note moyenne de 24,65 / 46. AOL.FR n'est pas très loin avec 19,7 / 46.

### Récapitulatif

Une fois tous les comparatifs des pages précédentes effectués, il est maintenant temps de calculer les classements généraux. Voici les tableaux récapitulatifs obtenus :

		
<b>Critère quantitatif</b>	<b>4</b>	<b>10</b>
<b>Syntaxe d'interrogation</b>	<b>3</b>	<b>6</b>
<b>Recherche avancée</b>	<b>0</b>	<b>6</b>
<b>Qualité de la présentation des résultats</b>	<b>10</b>	<b>6</b>
<b>Délai d'affichage</b>	<b>10</b>	<b>10</b>
<b>Qualité de l'annuaire associé</b>	<b>8</b>	<b>3</b>
<b>Pertinence des réponses</b>	<b>8</b>	<b>10</b>

Pour effectuer les classements généraux, nous avons tenu compte de deux populations différentes :

• **Le débutant sur l'Internet** : il ne saisit qu'un ou deux mots clés sur l'outil de recherche et n'utilise quasiment jamais les fonctions ou les pages de recherche avancée. Ce qui l'intéresse, c'est un outil de recherche rapide, simple à utiliser, le plus exhaustif possible mais surtout le plus pertinent possible lorsqu'on saisit un ou deux mots clés. Cette notion de pertinence est, sans conteste, le critère numéro 1 de choix de l'outil de recherche. Pour tenir compte de ce profil, les coefficients suivants ont été appliqués par rapport aux tests effectués précédemment :

<b>Critère</b>	<b>Coefficient</b>
Critère quantitatif	2
Syntaxe d'interrogation	1
Recherche avancée	1

Présentation des résultats	3
Délai d'affichage des pages	2
Qualité de l'annuaire associé	5
Pertinence des réponses moteur	5

• **Le spécialiste** : il souhaite aller beaucoup plus loin, utilise de façon beaucoup plus approfondie toutes les fonctionnalités de l'outil. Il apprécie la syntaxe et les interfaces de recherche avancée, ainsi que les fonctionnalités connexes proposées sur la page de résultats. Si ce qui l'intéresse avant tout, c'est la pertinence des réponses, bien sûr, il n'oublie pas cependant de prendre en compte toutes les fonctionnalités que lui propose le moteur pour arriver au résultat le plus efficace. L'exhaustivité et la durée d'affichage des pages sont des critères moins déterminant pour lui, même s'ils restent importants, bien entendu. Les coefficients utilisés sont donc :

Critère	Coefficient
Critère quantitatif	2
Syntaxe d'interrogation	3
Recherche avancée	2
Présentation des résultats	2
Délai d'affichage des pages	2
Qualité de l'annuaire associé	5
Pertinence des réponses moteur	5

Les classements obtenus en multipliant les notes des moteurs par les coefficients adéquats sont proposés ci-dessous.

<p>• <b>Population "Utilisateur débutant" :</b></p> <p><b>1. AOL.FR : 141 points</b>  <b>2. Google.fr : 135 points</b></p>
<p>• <b>Population "Utilisateur avancé" :</b></p> <p><b>1. Google.fr : 147 points</b>  <b>2. AOL.FR : 137 points</b></p>

Le résultat est assez surprenant pour un outil aussi récent que celui d'AOL.FR : pour une population d'internautes débutants, AOL.FR semble plus pertinent ou, en tout cas, globalement plus efficace que Google. La raison : un annuaire plus pointu, les options de présentation des résultats perfectionnées, qui contrebalancent une pertinence légèrement moindre des résultats sur la première page de résultats et une absence de possibilités de recherche avancée. Google prend néanmoins sa revanche pour les utilisateurs plus chevronnés. Les fonctionnalités de recherche avancée font, ici, la différence...



## Description du nouvel outil de recherche d'AOL France : AOL.FR

[Retour au sommaire de la lettre](#)

Le nouvel outil de recherche d'AOL France est donc sorti. Le mois dernier, nous avons publié une interview des chefs de projet AOL pour ce nouvel outil. Ils nous ont reçu, il y a peu, chez eux, pour nous donner plus d'informations...

### Historique

Pour revenir à la genèse de ce nouvel outil, il faut revenir 18 mois en arrière. A cette date, soit vers le mois de septembre 2000, la décision est prise au niveau d'AOL France de créer un nouvel outil de recherche dont le coeur serait un annuaire de recherche. L'Open Directory francophone, qui appartient indirectement à AOL, ne satisfait pas entièrement les besoins francophones et ne sera dès lors pas utilisé. Le contenu brut de l'annuaire LookSmart France est alors acheté en septembre 2000. Pendant plus de six mois, un très important travail de refonte de ces données de base va être entrepris afin de mettre cet annuaire aux normes éditoriales d'AOL France. Le but est de proposer des sites plus qualifiés. Un long travail d'analyse des requêtes le plus souvent demandées sur le site d'AOL donne des priorités pour rebâtir la taxonomie de l'outil. Des branches entières de l'arborescence sont modifiées, voire changées d'emplacement. Des sites sont enlevés, d'autres rajoutés, certaines descriptions sont modifiées, etc. Il en est de même pour les catégories. Bref, d'une base de travail qualifiée de correcte par AOL, une équipe de documentalistes pilotée par Cybion (voir interview dans cette lettre) recrée un nouvel annuaire, sous le contrôle d'AOL. Six mois et quelques pour un remaniement énorme de l'outil de base...

En parallèle de la création de ce nouvel annuaire, des réflexions sont menées pour l'aspect "moteur". Plusieurs technologies sont testées. Le cahier des charges peut se résumer à : rapidité, pertinence, sur mesure. Seul Exalead apportait des réelles nouveautés sur le troisième point... Les chefs de projet chez AOL étaient, par ailleurs, des "aficionados" de l'Altavista de la "grande époque". Ils connaissaient donc bien la fonction "Refine" (noms de code : Cow9 ou LiveTopics) qui avait été développée par un certain François Bourdoncle de l'Ecole des Mines de Paris pour Altavista à Palo Alto (Californie). Lorsqu'AOL approche Exalead, ils discutent avec... François Bourdoncle (voir interview dans cette lettre)... Une rencontre, des discussions, l'intime conviction que la technologie développée par Exalead tient la route, et le partenariat est conclu après toute une série de benchmark technique très pointue et plusieurs prototypes développés conjointement avec les développeurs d'AOL.

De nombreux "focus group" (enquêtes qualitatives) sont mis en place, en interne ou avec des abonnés au service en ligne. Puis un très gros travail de préparation pour prendre en compte les retours issus de ces séances de travail. Finalement, l'outil est lancé à la mi-avril 2002 avec de grandes ambitions : devenir l'un des outils de recherche leaders du marché francophone, en attendant plus... ;-)

### Fonctionnalités

Exalead amène surtout deux fonctionnalités étonnantes et innovantes : la proposition de mots clés et expressions connexes, et la proposition de catégories pertinentes dans l'annuaire (voir interview de François Bourdoncle dans cette lettre). Cette recherche "par approches successives" permet de définir des thèmes de recherche différents, des orientations diverses pour la recherche. Exemple parmi tant d'autres : le mot clé "sucre" qui peut désigner une ville de Bolivie ou ce que l'on met chaque matin dans son café. L'outil va proposer immédiatement les deux possibilités : "Pays du monde / Amérique / Bolivie" et "Cuisine et gastronomie / Cuisine et recettes / Guides de recettes". L'approche est donc différente de celle d'un moteur de recherche classique, sur lequel on va essayer, dès la première requête, d'obtenir les pages qui nous intéressent, alors que sur AOL.FR, on va effectuer une recherche en plusieurs "passes", en approximations successives, un peu comme une navigation au près, pour les fanas de voile... Ne dit-on pas que l'intelligence n'est pas de savoir où l'on veut aller, mais de réfléchir à la meilleure façon d'y aller ? ;-)

Notos également que la syntaxe de recherche d'Exalead s'applique à l'outil d'AOL. Si vous désirez la connaître, tapez n'importe quoi (en tout cas une requête qui ne donne pas de résultats) sur le site d'Exalead, l'outil vous donnera une page d'aide décrivant la syntaxe. Exemple : <http://www.exalead.com/cgi/exalead?v=0&fr=&q=gfghkfgggg%24&p=Demo>

### **Index du moteur de recherche**

L'index du moteur est en fait centré autour de l'annuaire. Si les liens de soumission d'un site dans les catégories de cet annuaire ont disparu, ils devraient refaire leur apparition sous peu, un peu de patience ;-)

Aujourd'hui, l'index du moteur (qui contient 50 millions de pages) est rafraîchi une fois par mois, mais une partie (une dizaine de millions) est rafraîchie une fois par semaine. En règle générale, le but ultime des chefs de projet est de mettre en place un rafraîchissement "idéal" des pages de l'index proportionnel au taux de mise à jour des sites indexés : si un site est mis à jour, en moyenne, tous les mois, le but sera de l'indexer tous les mois. Si le délai entre deux modifications est de 24 heures, l'indexation sera quotidienne, etc.

L'index du moteur ne contient pas que des pages en français. En effet, il a été décidé de privilégier, bien sûr, les pages en langue française, mais de fournir également des pages en anglais, ou dans d'autres langues, pouvant intéresser les internautes francophones. Exemple type : le site d'Airbus qui est en anglais mais dont la cible n'est pas exclusivement anglophone. Exalead a donc développé des systèmes de tri, notamment basés sur les liens, pour bâtir une partie de l'index dans d'autres langues. On trouve ainsi des pages en anglais, mais également en espagnol ou en allemand... C'est également le cas pour des sites officiels d'artistes (le site officiel de Madonna par exemple) ou d'actualité. Si ces sites sont très "pointés" depuis des pages en français, cela doit signifier qu'ils intéressent les internautes français. Ils sont donc intégrés dans l'index.

Pour ce qui est du crawl, il n'existe pas, a priori, de critères bloquants :

- Les pages dynamiques (asp, cfm, php...) ne posent pas de problèmes, qu'il y ait passage de paramètres ou non.
- le format de fichiers ne pose pas de problème non plus. L'outil peut tout à fait indexer des fichiers PDF ou Word, même s'ils ne sont pas encore intégrés dans l'index pour l'instant.
- Les liens Javascript ne sont cependant pas suivis par le crawler.
- Sites conçus avec des frames : seules les pages mères sont affichées (celles qui contiennent la balise Frameset). Les pages filles (pages de contenus) servent cependant au calcul de l'indice de popularité (les liens qu'elles contiennent sont pris en compte dans ce calcul). Certaines frames, comme celles qui servent à la navigation, ne sont pas indexées si leur contenu est jugé trop peu intéressant.

Les champs pris en compte sont les suivants :

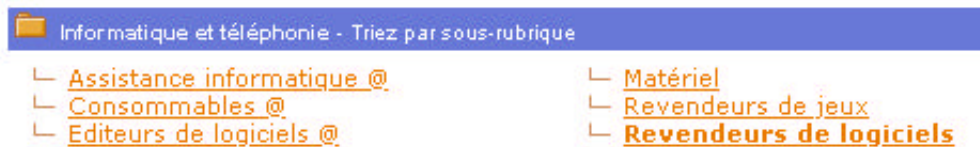
- Titre : OUI
- Balises Meta : NON
- Texte visible : OUI
- Commentaires : NON
- Options ALT des balises IMG : OUI
- Nombre d'occurrence du mot demandé : OUI, mais dans un nombre limité
- URL : NON
- Date de mise à jour : OUI (plus le document est récent, meilleur est son classement)

Au niveau des critères de pertinence, le fonctionnement est assez "classique" : le titre et le texte visible sont, bien sûr, très importants, l'indice de popularité, basé sur un calcul quantitatif et qualitatif (très proche du système utilisé par Google), etc. Rien de "révolutionnaire" de ce côté-là (mais on ne peut pas révolutionner de tous les côtés non plus ;-))...

Au niveau de la technologie, tous les serveurs sont localisés en France, chez KPNQwest et administrés par AOL.

### **Annuaire de sites web**

L'annuaire de sites actuel a donc été conçu par AOL et réalisé par Cybion (voir interview dans cette lettre). On peut s'y "promener" de façon simple en faisant du "browsing", en descendant donc l'arborescence de catégories en sous-catégories. On s'aperçoit rapidement que, lors de cette exploration, certaines catégories apparaissent en gras et pas d'autres. Exemple, ici, pour "Revendeurs de logiciels" :



Il s'agit d'une fonctionnalité développée spécifiquement par AOL en récupérant de la réponse retournée par Exalead le plus fort indice de popularité et afficher ensuite en gras ces catégories. Pour bien comprendre comment fonctionne l'outil, il faut préciser un point important : dans l'index du moteur, chaque page web dispose d'un "drapeau" ("flag") qui indique si cette page est extraite d'un site présent dans l'annuaire ou non. Pour chacune des 50 millions de pages de l'index, il est donc possible de savoir si elle est extraite d'un site de l'annuaire. On peut ainsi calculer de façon "simple" (pardon...) l'indice de popularité de chaque site de l'annuaire et donc les catégories de l'annuaire qui proposent les sites les plus "populaires". Ces catégories sont alors affichées en gras. Il en est de même lors d'une recherche par mots clés : si vous tapez le mot clé "warrants", vous obtenez la proposition de catégorie suivante :



A ce niveau, si vous cliquez sur "Bourse et marchés", vous obtenez les sous-catégories suivantes :



Dans ces sous-rubriques, sont indiquées en gras celles ("Informations financières", "Warrant") qui contiennent les sites les plus pertinents (issus de la recherche dans l'index du moteur) pour votre recherche.

A notre connaissance, AOL est le seul outil de recherche à fonctionner de cette façon sur le Web. Il faut avouer que, ainsi, on ajoute beaucoup d'"intelligence" supplémentaire à l'outil... Encore faut-il que l'internaute se familiarise avec (et comprennent) cette nouvelle façon de chercher. Sur ce point, seul l'avenir nous dira si ces innovations rencontrent leur public.

De la même façon, lorsqu'on parcourt l'arborescence de l'annuaire, les sites ne sont en fait réellement classés qu'en bout des branches de l'arborescence, lorsqu'il n'existe plus de possibilités de sous-rubrique. Exemple pour la catégorie : "Accueil > Bourse et finance > Bourse et marchés > Warrant"

Lorsqu'on parcourt l'arborescence, des sites sont pourtant présentés à chaque niveau, comme :  
Accueil > Bourse et finance  
Accueil > Bourse et finance > Bourse et marchés

Ces sites représentent, en fait, tous les sites présents dans les sous-catégories proposées au niveau final et qui disposent du meilleur indice de popularité. Ainsi, le site Boursorama, présenté de façon intermédiaire au niveau "Bourse et finance > Bourse et marchés", est en fait classé au

niveau "Bourse et finance > Bourse et marchés > Informations financières", mais comme il dispose d'un fort indice de popularité, il est également affiché sur les niveaux antérieurs. Qu'on ne vienne pas nous dire que cet outil ne fourmille pas d'innovations originales et intéressantes... ;-)

Enfin, notons que le chantier d'affichage des liens "Overture" est également en cours et devrait se mettre en place cet été. Le cahier des charges d'AOL est très strict et précis : les liens devront être clairement identifiés comme étant des liens promotionnels et devront être extrêmement pertinents. Ils devraient se situer avant les rubriques, donc en première position dans la page de résultat.

Pour terminer, on peut noter que les pages de résultats d'AOL.FR proposent des "widgets", des zones publicitaires (notamment pour Amazon.fr) à droite de la page de résultats. Ces "widgets", sont techniquement gérés et développés par AOL et proposent des liens qui sont vus le plus pertinents possibles grâce aux remontées statistiques d'Exalead . Ici aussi, il a été tenté d'inclure de l'"intelligence" pour ne pas proposer un simple lien vers Amazon, par exemple, mais bien un lien très précis amenant un "plus" à l'internaute. Le mot clé "madonna", par exemple, identifiera que l'internaute recherche un CD (message "CD trouvés") et pointera vers une page d'Amazon proposant de très nombreux disques de la chanteuse. Une recherche identique sur Yahoo.fr, par exemple, propose un lien assez similaire sur la page de résultats du portail, mais qui pointe vers une page d'Amazon beaucoup plus généraliste...

Au niveau de l'utilisation du service, un petit mois après son lancement, la répartition de l'utilisation du nouvel outil de recherche est la suivante :

- Recherche par Mots-clés brute : 70 %
- Fonctionnalité "Affiner avec les mots" : 15 %
- Utilisation des rubriques de l'annuaire : 15 %

Ces chiffres sont à interpréter avec prudence, car nous n'avons pas assez de recul pour en tirer des conclusions qui pourraient s'avérer trop hâtives : découverte des nouvelles fonctionnalités, adoption par les utilisateurs, période fin avril et début mai avec congés et élections, etc. Bref, il faut encore attendre un peu pour avoir à notre disposition une vision plus juste de la façon dont les visiteurs du site utilisent l'outil...

Le lancement de la nouvelle recherche a, d'autre part, permis une augmentation significative de l'utilisation de la recherche sur AOL.FR et sur la partie réservée aux abonnés. Cette augmentation est de l'ordre de 30 % et continue à progresser. Selon le baromètre Xiti / 1ère Position ([http://www.1ere-position.com/outils\\_de\\_recherche/barometre/](http://www.1ere-position.com/outils_de_recherche/barometre/)), qui mesure le trafic généré sur un site web par les outils de recherche, AOL serait également passé en avril de la 8ème à la 6ème place avec une augmentation de 1,9% du trafic généré par rapport au mois précédent.

## Conclusion

En conclusion, on peut dire que ce nouvel outil de recherche est "bien né" : il propose une alternative intéressante à Google et son évolution dans les six mois à venir (de très nombreux projets sont encore dans les cartons et seront intégrés au fur et à mesure) sera intéressante à suivre. Bravo donc à AOL et à ses partenaires qui nous proposent un outil de recherche innovant, original et qui fera date !

## **Cybion, créateur de l'annuaire d'AOL.FR**

[Retour au sommaire de la lettre](#)

AOL.FR a fait appel à la société Cybion pour créer son annuaire de sites web. Une bonne occasion pour chercher à mieux connaître cette entreprise, présente depuis fort longtemps sur le marché de la recherche d'information sur le Web. Interview des deux chefs de projets qui ont travaillé sur l'annuaire d'AOL.FR...

### ***- Bonjour, pouvez-vous vous présenter brièvement ainsi que nous en dire plus sur l'activité de la société Cybion ?***

Je m'appelle Guillaume Lory, et je suis directeur technique de Cybion. Diplômé de l'ISTIA (Institut des Sciences et Techniques de l'Ingénieur d'Angers) et doctorant auprès de l'Université Paris X, je m'occupe plus particulièrement des outils nécessaires au maintien de la base de données des sites et de la taxonomie. Pour ce faire, un logiciel spécifique a été développé pour permettre à l'équipe de cyberdocumentalistes d'administrer la taxonomie et d'y indexer les sites. Cette base de données nous sert enfin à exporter les fichiers RDF qui sont régulièrement envoyés à AOL pour être intégrés à l'annuaire.

Je m'appelle Bénédicte Noblet, et je suis chargée d'études chez Cybion depuis octobre 1999 et responsable du projet AOL depuis novembre 2000. J'ai une Maîtrise en sciences de l'information et de la documentation et un DUT (année spéciale) en Documentation d'entreprise. Dans le cadre de la collaboration de Cybion avec AOL, je m'occupe plus précisément de l'optimisation de la taxonomie, du contrôle qualité de l'indexation de sites web, ainsi que de l'encadrement de l'équipe. Je rencontre régulièrement le chef de projet "Annuaire" chez AOL pour mettre au point l'optimisation de la base de données.

Cybion est la première société européenne spécialisée dans la recherche d'informations et la veille sur Internet. L'expérience de la veille est capitalisée au sein de l'entreprise, grâce aux revues de presse, aux études sectorielles, thématiques ou d'image, et aux séminaires de formation que nous effectuons régulièrement. Ainsi Cybion développe depuis 1996 des services qui vont des recherches ponctuelles d'informations jusqu'à la mise en place de cellules de veille (externalisées ou directement chez le client), en passant par la mise en place de dispositifs d'alerte pour gérer toute situation de crise ou la personnalisation d'annuaires Web pour Intranet ou site portail.

### ***- Comment s'est passé l'historique de la création de l'annuaire d'AOL.FR ?***

AOL a fait l'acquisition de la base de données Looksmart. Nous avons ensuite travaillé à son optimisation avant la mise en ligne de mai 2001 : réorganisation de pans entiers de l'arborescence, changement d'intitulé de catégories, etc.

Depuis cette première mise en ligne, l'annuaire est sans cesse amélioré pour permettre à l'internaute de naviguer simplement et de trouver l'information qu'il recherche.

Avec la mise en ligne du nouveau moteur AOL.FR en avril 2002, la synergie est totale entre annuaire et moteur.

### ***- Combien l'annuaire d'AOL propose-t-il de sites aujourd'hui ? Et de catégories ?***

Environ 90.000 sites et 12.000 catégories.

### ***- Quel est le taux de progression ? Combien ajoutez-vous de nouveaux sites chaque mois par exemple ?***

Le travail consiste à optimiser constamment la base de données. Des sites sont supprimés car ils ne sont plus mis à jour ou leur qualité a baissé.

Pour l'annuaire, le but n'est pas d'avoir un grand nombre de sites mais d'avoir les sites incontournables sur un thème précis. Le qualitatif est préféré au quantitatif.

### ***- Par rapport à l'annuaire de Looksmart qui a été l'annuaire sur lequel vous avez basé votre travail initial, l'avez-vous sensiblement modifié dans sa structure et son contenu ou êtes-vous partis de cette base pour "simplement" :-> y rajouter des sites ?***

La taxonomie a évolué en fonction des demandes précises d'AOL et également en fonction de l'actualité : développement ou réorganisation de certaines parties de l'arborescence comme pour la musique ou la méta-catégorie Famille et santé, création de catégories sur Loft Story ou l'élection présidentielle.

Il s'agit d'être réactif par rapport aux requêtes des internautes. Si un mot clé semble émerger, alors la création d'une catégorie correspondante avec des sites pertinents est essentielle.

L'annuaire AOL tel qu'il est aujourd'hui n'a donc plus grand chose à voir avec la version initiale de Looksmart ;-)

***- Effectuez-vous des "ménages" (nettoyage de la base de données : modification de descriptifs existants, élimination de sites obsolètes, etc.) sur certaines catégories ? Selon quels critères ? Selon quelle périodicité ?***

La base de données se compose de 12 méta-catégories, et en une année, chacune d'entre-elles sera revue : suppression des sites obsolètes, modifications des chroniques ou des titres si nécessaire.

L'actualité peut également être un critère pour revoir une catégorie : par exemple pour la Saint Valentin ou le tournoi de Roland Garros.

***- Combien de personnes travaillent-elles sur cet annuaire, chez vous ? Quel est leur profil ? Sont-elles spécialisées chacune sur une thématique donnée ou sont-elles "multi-domaines" ?***

Une équipe de 5 personnes est complètement dédiée au projet AOL. Elle comprend 4 cyberdocumentalistes ainsi qu'une personne pour l'encadrement et le travail de taxonomie. Guillaume Lory apporte son soutien pour les aspects techniques. Les documentalistes sont issus d'une formation universitaire généraliste (sciences humaines, ethnographie, ...), complétée par un diplôme en sciences de l'information. Chaque cyberdocumentaliste supervise 3 méta-catégories et veille à leur optimisation tant au niveau de la taxonomie que de la sélection de sites pertinents. Néanmoins, chacun est amené à travailler sur l'intégralité de la base.

***- Existe-t-il une syntaxe d'interrogation spécifique de l'annuaire (par exemple, recherche uniquement sur l'url, le titre ou la description) ?***

Même si une syntaxe est utilisable, le moteur sélectionne de lui-même les opérateurs booléens les mieux adaptés à la recherche en cours. Ainsi une recherche portant sur Poivre d'Arvor donnera d'abord les résultats sur le présentateur (utilisation de la chaîne de caractère) avant de donner les résultats sur le «poivre» (utilisation du booléen OR). Ainsi pour le grand public, il n'y a pas de question à se poser.

***- Comment peut-on référencer un site dans l'annuaire d'AOL.FR ? Quels sont les délais d'évaluation ? Quels sont vos critères d'acceptation et de refus ?***

L'option pour suggérer un site a été supprimée temporairement mais elle sera prochainement rétablie. Nous appliquons le cahier des charges élaboré par AOL : qualité du site et de son information, ergonomie, bon fonctionnement des liens, mise à jour, adéquation avec le type d'information recherché par les internautes. Le délai d'évaluation est variable.

***- Quel est le pourcentage des sites pris en compte lors de soumissions volontaires de la part des webmasters par rapport à ceux intégrés par l'intermédiaire d'un processus de veille par les documentalistes ?***

Seul 1/3 des sites suggérés est conservé. Les 2/3 des sites nouveaux indexés proviennent de la veille menée par les documentalistes. De par son expérience en recherche d'information et en veille, Cybion est à même de détecter régulièrement les nouveaux sites de référence et de les intégrer à l'annuaire AOL. A ce sujet, nous avons mis en place une cellule de veille à base d'agents intelligents pour faciliter ce travail de surveillance et d'identification de nouveaux sites.

***- Combien de soumissions receviez-vous lorsque cette possibilité était donnée aux webmasters sur l'annuaire ?***

Nous recevons 250 soumissions de sites par semaine en moyenne.

**- Pourquoi avoir ôté les liens de soumission sur l'annuaire ? Quand cette possibilité sera-t-elle remise en place ?**

Cybion ne gère que le contenu de l'annuaire. Cette modification de l'interface est du ressort d'AOL.

**- A priori, il n'existe pas de projets de soumission payante sur l'annuaire d'AOL.FR. Quelle est votre vision de, votre avis sur, ce type de prestations ?**

Encore une fois, Cybion ne gère que le contenu de l'annuaire. Si un système de soumission payante est ajouté par AOL sur son annuaire, nous intégrerons de la même manière ces sites (NDLR : a priori, aucun projet allant dans ce sens n'est pour l'instant dans les cartons d'AOL.FR).

**- Travaillez-vous, chez Cybion, avec d'autres partenaires pour créer de tels annuaires, généralistes ou spécialisés ?**

L'équipe de 5 personnes travaille exclusivement pour l'annuaire AOL. Néanmoins, Cybion est présent sur les marchés des annuaires en proposant son expertise et ses outils à ses clients pour la création d'annuaires relatifs à un domaine restreint et spécialisé comme l'électricité, le textile ou les cosmétiques. Cybion a également créé et développe un annuaire consacré au monde de l'intelligence économique sur le site [www.veille.com](http://www.veille.com), et un autre dédié au monde des agents intelligents [www.agentland.fr](http://www.agentland.fr).

**- Quels sont les projets pour 2002 sur l'annuaire d'AOL.FR ?**

De notre point de vue, nous allons continuer à réviser l'intégralité des méta-catégories de l'annuaire en optimisant la synergie avec le moteur Exalead. L'intégration de nouveaux sites continuera avec toujours le souci de privilégier la qualité par rapport à la quantité. Cybion apportera d'autre part son savoir-faire dans le domaine des agents intelligents pour automatiser certaines des tâches de gestion de cet annuaire. Nous avons d'ores et déjà développé un outil qui nous permet de vérifier l'état des sites présents dans l'annuaire. Les sites inaccessibles ou non valides sont ainsi régulièrement listés pour être révisés ultérieurement. Enfin, lors de l'actualisation des catégories, nous utilisons des agents de recherche pour identifier rapidement les sites incontournables à ajouter.

**Merci, Bénédicte Noblet et Guillaume Lory !!**

**Exalead : une technologie d'avance**[Retour au sommaire de la lettre](#)

Si Cybion a été choisie pour la partie "annuaire", c'est la société Exalead qui a pris en charge la partie "moteur" de l'outil de recherche d'AOL.FR Nous avons déjà donné pas mal d'infos sur cette technologie dans cette lettre (voir la lettre R&R du mois d'octobre 2001 à l'adresse <http://abonnes.abondance.com/archives/acturech/0110.html>). En complément de ces renseignements, nous avons rencontré François Bourdoncle, l'un de ses dirigeants, qui nous en dit plus sur ses produits et sa stratégie....



**- François Bourdoncle, bonjour. Pouvez-vous vous présenter à nos lecteurs ?**

Vous en saurez plus ici, sur cette superbe page ;-) : <http://www.exalead.com/Francois.Bourdoncle/>, mais disons que je suis PDG de la société Exalead ainsi que Professeur associé à l'École Polytechnique. J'ai travaillé en 1993 et 1994 sur le moteur de recherche Altavista pour lequel j'ai créé la fonction "Refine" (noms de code "Cow9" ou "Live Topics") qui permettait d'affiner une recherche en proposant plusieurs mots clés connexes par analyse statistique des pages proposées pour la requête.

**- Quelles sont les fonctionnalités innovantes de votre technologie par rapport à celles existant déjà sur le marché ?**

L'ensemble de notre technologie est basée sur des algorithmes statistiques. Cela a des avantages et des inconvénients. Des inconvénients car, sur le papier, cela marche moins bien que des algorithmes strictement linguistiques. Des avantages parce que ça fonctionne sur des corpus qui peuvent être immenses et/ou hétérogènes. D'autre part, les algorithmes statistiques permettent de proposer des fonctions "intelligentes" que les systèmes linguistiques ne peuvent proposer. Nous proposons pas mal de fonctions nouvelles et originales que nous essayons d'optimiser chaque jour un peu plus :

- La lemmatisation, qui permet de trouver la racine des mots. En fait, on ne fait pas de la lemmatisation "pure" (personne n'en fait vraiment sur le Web), mais les options prises vont plus loin qu'une simple troncature. Exemple : "vaches folles" va trouver "vache folle" et inversement. "chiennes" va trouver "chien" et "chiens" et inversement, etc. Le moteur lemmatise en fait les requêtes dites "compliquées" (comportant au moins deux mots). Les requêtes sur un seul terme ne sont pas lemmatisées.
- D'autre part, on reconnaît, par des analyses statistiques, les requêtes qui sont en fait des expressions (on fait du "guillemet automatique"). Si, statistiquement, l'expression "vache folle" est très souvent présente sur le Web, on sait que quelqu'un qui demande ces deux mots clés séparés par un espace demande l'expression telle quelle (comme s'il mettait des guillemets) et on va lui fournir des pages qui contiennent les deux mots l'un à côté de l'autre et dans cet ordre, et non pas "vache" dans le titre et "folle" en bas de page... Le système connaît donc, en amont et grâce à des analyses statistiques, les "bonnes phrases" du Web. Cela permet de gagner beaucoup de temps lors de la recherche d'information... En fait, on recherche le "bon" groupe nominal dans la requête pour définir le "sens" de la demande : que désire la personne qui tape ses mots clés ?
- Le classement des résultats, en lui-même, s'effectue selon des techniques assez "classiques" sur le Web. On peut dire que, de ce côté, on est assez proche d'une approche à la Google, avec un calcul de l'indice de popularité basé à la fois sur un mode quantitatif et qualitatif (la "qualité" des liens pointant sur un site est importante), mais la plupart des moteurs majeurs orientent également leurs travaux dans cette direction au jourd'hui... En revanche, nous sommes plus innovants dans la lutte anti-spam : une personne chez nous s'occupe de cela à temps plein. Le but est de définir, de façon statistique, ce qui du spam et ce qui n'en n'est pas, et de l'appliquer à notre index : si l'on sait détecter, de façon statistique, ce qui "ressemble" à du spam, on arrive à



lutter de façon plus efficace contre ce fléau en faisant ressortir assez rapidement les pages incriminées. C'est un gros boulot, car les sites de sexe, les premiers visés, mènent la vie rude aux moteurs de recherche... Mais nous intensifions notre travail dans cette voie, sans pouvoir en dire trop, comme vous vous en doutez bien... ;-)

- Nous avons également tenté d'innover dans le crawl, en ne nous basant pas sur des critères strictement linguistiques pour créer un index axé vers les internautes francophones. Il nous semble, en effet, que l'utilisateur d'un outil de recherche ciblé sur les internautes francophones n'est pas obligatoirement et à 100% intéressé par des pages strictement en langue française, mais que des sites en anglais, notamment, peuvent l'intéresser si il n'existe pas d'équivalent en français. Le site d'Airbus (<http://www.airbus.com/>) en est un bon exemple. Il est entièrement en anglais mais reste une bonne source d'information pour les utilisateurs français ou francophones du réseau. Idem pour le site de CNN. Il nous semble donc intéressant de proposer un index majoritairement francophone, mais avec une frange de sites connexes en d'autres langues. La sélection s'effectue par analyse des liens sortants des pages en français. Si de très nombreuses pages en français pointent vers un même site en anglais, il y a de fortes chances que cette source d'information soit pertinente, même pour un public francophone..

- Nous avons également intégré un système de correction orthographique et de phonétisation, mais elle est disponible uniquement dans notre offre "market". Ces possibilités n'existent donc pas sur les sites Exalead.com et AOL.FR.

- La synthèse des résultats sur la page d'accueil, avec les catégories des annuaires et les mots clés connexes, sont de purs résultats des algorithmes statistiques. Vous en avez parlé dans votre article de présentation de l'outil AOL.FR (NDLR : voir plus haut) Pour l'instant, seuls des sites comme Teoma ou Wisenut se sont aventurés sur ce terrain, mais il semble qu'ils travaillent sur des corpus (des masses de pages à analyser) plus faibles, ce qui rend des résultats moins fiables à mon avis. Mais je ne me suis pas penché de façon très fine sur les technologies déployées par ces moteurs...

Pour conclure, je dirais que notre force est de proposer une recherche assistée (catégories des annuaires + mots clés connexes) associée à de nouvelles possibilités intéressantes de navigation ("browsing") à l'intérieur des catégories des annuaires, ce qui est devenu très complexe sur des outils comme Yahoo! qui proposent plus de 35 000 rubriques !

***- Par rapport à la "démo" de votre technologie, présente sur le site Exalead.com, quelles particularités ont-elles été rajoutées pour le site AOL.FR ?***

La première différence est la base de départ : sur le site d'Exalead, nous avons pris en compte l'Open Directory. Sur AOL.FR, nous partons et centrons le système sur l'annuaire créé par Cybion sous la supervision d'AOL. Nous analysons également les liens autour des sites de l'annuaire, qui est véritablement le coeur du dispositif.

Les autres différences sont le fruits de développements AOL qui ont fait suite à d'études consommateurs sur les attentes des internautes :

Le fait de proposer, dans les catégories de l'annuaires, certains intitulés en gras pour indiquer les rubriques qui contiennent les sites les plus "populaires" ou celles qui contiennent les sites correspondant le mieux à une recherche par mots clés (NDLR : voir article de présentation de l'outil AOL.FR).

L'interface est également différente : sur le site d'Exalead, nous utilisons la colonne de gauche pour les catégories et les mots clés, alors que sur le site AOL.FR, ces informations sont présentées de façon horizontale, en début de page. Avantage : c'est plus visible, plus lisible. Inconvénient : la place est plus limitée que lorsqu'on utilise une colonne... Ainsi, seules trois catégories, avec leur chemin d'accès, sont proposées sur AOL.FR (ce chemin d'accès n'est d'ailleurs pas affiché sur le site d'Exalead). Nous proposons également plus de mots clés connexes sur AOL.FR (5 termes max sur Exalead, 8 sur AOL.FR).

Enfin, le site d'AOL.FR présente des "widgets", petites publicités à droite de l'écran, auxquelles AOL a désiré ajouter un peu d'intelligence en affinant le degré de compréhension de la requête de l'utilisateur.

Tout cela fait d'AOL.FR un outil spécifique est assez différent de notre "démon" présente sur le site d'Exalead.

***- Sur quels marchés vous positionnez-vous aujourd'hui ?***

Notre cœur d'activité se situe autour de technologies permettant aux entreprises de créer un moteur de recherche "perso" pour des coûts les plus faibles possibles, même si nous nous positionnons plutôt sur un créneau qui vise le haut de gamme. Nous visons notamment les grosses entreprises qui ont une cellule de veille. Nous pouvons leur vendre notre crawl francophone ainsi que des mises à jour périodiques. Libre à l'entreprise de traiter cette masse de données pour ses besoins spécifiques. Autre possibilité : livrer un crawler configurable qui permet de créer un index ne prenant en compte qu'un certain nombre de sites, sur un marché identifié par exemple : concurrents, sites d'actualité, etc. On peut extrapoler cette notion à la création d'index thématique : le système crawl uniquement des sites sur l'industrie, ou la santé, ou la justice, ou les sites scientifiques, ou la défense, etc. Autre possibilité : explorer l'index actuel et en ressortir spécifiquement les pages traitant d'un domaine donné. Il existe également beaucoup de traitements possibles dans l'analyse des liens du Web. Mais nous pouvons également créer des outils de recherche d'image, de fichiers PDF, Powerpoint, Excel ou Word : la technologie est tout à fait ouverte à l'indexation de fichiers autres qu'au format HTML. Bien entendu, notre technologie intéresse également les grandes entreprises, pour leurs portails Internet ou intranet. En fait, le spectre des applications est assez étendu, tout en restant dans le domaine de la recherche d'information...

Nous essayons notamment de travailler autour de la fusion des sources hétérogènes qui est le cœur de notre business : créer une plate-forme industrielle d'accès à des sources d'infos hétérogènes, voilà notre tâche au quotidien... Nous travaillons pour des clients qui ont de très nombreuses bases de données structurées différemment pour homogénéiser l'ensemble. Nous "aspérons" les données, harmonisons en XML le tout et mettons en place des interfaces pour naviguer, soit en utilisant des données structurelles (titres des documents, dossiers, etc.) ou des données sémantiques (comme la catégorisation automatique). Nous travaillons sur une unité et une harmonisation de la navigation. Cela a l'air simple, mais c'est un gros travail, que notre plateforme de développement Exascript permet de mener à bien.

***- Comment jugez-vous le marché de la recherche d'information sur le Web à l'heure actuelle ?***

Difficile, sans aucun doute. Il y a de moins en moins d'acteurs dans le domaine des moteurs de recherche (Altavista est en perdition, Inktomi semble se tourner vers des marchés orientés sur les technologies de cache et les intranets), de moins en moins de portails qui peuvent payer de telles technologies qui coûtent, d'autre part, de plus en plus cher. Aujourd'hui, Google est leader, mais qu'en sera-t-il demain ? La "Googlisation" du Web serait-elle une bonne chose ? Est-elle envisageable ? Quel est le modèle économique de sociétés comme Google ? Est-il fiable à moyen terme ? De nombreuses questions se posent. Ce qui est sûr, c'est que le marché se restreint à quelques entreprises. Nous tentons de tirer notre épingle du jeu au niveau francophone et observons, pour l'instant, les marchés européens et mondial. Mais chaque chose en son temps... ;-)

***Merci, François Bourdoncle !!***

## Un point sur Inktomi

[Retour au sommaire de la lettre](#)

Inktomi équipe le site AOL.FR pour ses recherches sur le "Web mondial". Nous en avons profité pour poser quelques questions à Ron Verheijen, "Business Development Manager Web Search" chez Inktomi Europe, à propos de cet outil de recherche, ce "grossiste en urls" qui semble plus se positionner aujourd'hui sur des technologies de "cache serveur" et des applicatifs intranet que sur le marché des outils de recherche Internet...



### ***- Quelle est la taille actuelle de l'index d'Inktomi ?***

Plusieurs réponses sont possibles : notre "Webmap" (notre connaissance du Web, constituée par notre crawl de base) est d'environ 3 milliards de pages web. Il est en phase de croissance. Après traitement (détection du spam, dédoublonnage, etc.), nous arrivons à un index d'environ 600 millions de documents, qui atteindra bientôt le chiffre de 700 millions, avec pour objectif à moyen terme d'augmenter cette capacité de façon beaucoup plus importante...

### ***- Combien de "clusters" (index) utilisez-vous aux Etats-Unis et en Europe ? Sont-ils différents ?***

Nos clusters sont localisés à plusieurs endroits aux Etats-Unis (côte est et ouest), et il existe également plusieurs emplacements avec des serveurs proxy (cache). Nous utilisons également des clusters privés pour certains clients. Des "IDPP (Inktomi Data Protocol Proxies)" sont également utilisés afin de récupérer des données de plusieurs bases de données pour les uniformiser et fournir une liste de résultats uniques. Il existe également des IDPP en Europe.

### ***- Est-ce que ces clusters sont divisés en plusieurs parties ?***

Tout d'abord, nous utilisons plusieurs types de clusters (d'index) : recherche web "classique", urls soumises au travers de notre offre de référencement payant, "objets multimédia", etc. Nous utilisons également des clusters privés pour certains clients. Chacun de ces index dispose d'un crawler spécifique. L'index européen utilise des technologies qui permettent d'effectuer une identification régionale (détection de la langue ou du nom de domaine). Nous utilisons également de façon assez forte l'analyse de liens pour parachever ce travail de distinction des index régionaux.

Pour ce qui est des index concernant la recherche web "classique", voici la situation actuelle :

- Un index nommé "Best of the Web" (plus de 110 million de documents), remis à jour tous les 3 jours, avec un rafraîchissement complet tous les 9 jours.
- Un index nommé "Gigadoc" (plus de 500 million documents), entièrement rafraîchi tous les 30 jours.
- Un index spécifique pour l'Asie-Pacifique (plus de 65 millions de docs), rafraîchi tous les 21 jours.

### ***- Est-ce que les pages "dynamiques" (asp, cfm, php) sont acceptées dans l'index d'Inktomi ?***

Pour le crawl "classique", non. Pour les sites soumis par l'intermédiaire de notre offre "Paid inclusion" de référencement payant, cela ne pose pas de problème, sauf s'il s'agit de spam ou autre problème, point que nous contrôlons, bien sûr, de façon très fine. Vous aurez plus d'infos ici, par exemple, sur notre politique dans ce domaine :

[http://www.inktomi.com/products/web\\_search/guidelines.html](http://www.inktomi.com/products/web_search/guidelines.html).

### ***- Il n'est plus possible de soumettre une url sur HotBot US depuis plusieurs semaines. Comment les webmasters peuvent-ils soumettre un site dans l'index d'Inktomi ?***

Exact, il n'est plus possible de soumettre un site sur HotBot US. Notre expérience nous indique qu'il y a beaucoup trop de spam soumis au travers d'une telle interface gratuite. Nous encourageons les webmasters à utiliser notre offre de référencement payant ([http://www.inktomi.com/products/web\\_search/inclusion.html](http://www.inktomi.com/products/web_search/inclusion.html)), avec tous ses avantages, ce qui nous permet de bâtir de larges index de contenus pertinents sur cette base. Cependant, nous avons également une liste de sites non commerciaux que nous crawlons de nous-même. Les urls soumises par l'intermédiaire de notre offre payante se trouvent sur des serveurs et dans des index spécifiques. Les résultats venant de ces bases de données sont mixés avec ceux émanant de nos index "classiques". Même si certains sites de nos clients "portails" proposent encore des formulaires de soumission gratuite, nous ne les prenons que très peu en compte. Le référencement payant est devenu quasi prioritaire pour être sûr d'entrer dans l'index d'Inktomi, si vous ne voulez pas attendre que le spider suive un lien de façon "naturelle" vers votre site. Nous ne remettons pas en ligne de formulaire de soumission gratuite sur les sites qui n'en proposent plus aujourd'hui...

***- Quel est le premier bilan de votre offre de référencement payant ? Est-ce que les résultats correspondent à vos objectifs initiaux ?***

Ce programme fonctionne très bien. Nous en sommes très content. Les clients qui utilisent cette fonctionnalité semblent également ravis, ce qui est très important à nos yeux. Je reçois beaucoup de commentaires élogieux de la part de nos utilisateurs. La société "Position Technologies", l'un de nos revendeurs, m'a envoyé dernièrement un florilège de réactions très encourageantes de la part de leurs clients. Vous pouvez également trouver pas mal d'infos, peut-être plus objectives ;-), ici : <http://www.highrankings.com/issue001.htm> à ce sujet.

***- Avez-vous un revendeur spécifique en France pour vos offres de référencement payant ?***

Vous trouverez la liste de nos revendeurs ici : [http://www.inktomi.com/partners/search\\_resellers.html](http://www.inktomi.com/partners/search_resellers.html). Certains d'entre eux (par exemple Outrider) ont une présence européenne. Nous travaillons également sur un programme de distribution de ces offres par nos clients portails européens.

***- Inktomi a perdu plusieurs gros clients dans les mois qui viennent de s'écouler, notamment en France : Yahoo!, Nomade.fr, Lycos, Club-Internet. Fast Iorgne vers HotBot US. Looksmart fait migrer ses clients depuis la solution Inktomi vers celle de Google. AOL utilise maintenant Exalead pour son index francophone. Le marché de la recherche d'information sur le Web est-il encore un marché stratégiquement prioritaire pour Inktomi ?***

Oui, tout à fait, on peut même dire que le business d'Inktomi autour de la recherche d'information n'a jamais été autant en expansion... A la fois en revenus et en investissements. Nous proposons des technologies de plus en plus efficaces, et nous continuons à engranger de nouveaux clients, comme Bluewin.ch, le portail des PTT suisses dernièrement, marché que nous avons gagné contre les sociétés dont vous parlez dans votre question. Il est peut-être bon également de savoir que Lycos a une stratégie basée, en fait, autour de deux solutions de recherche, Fast et Inktomi. Je ne peux malheureusement pas en dire plus dans le cadre de cet interview, désolé...

***- Quels sont les projets d'Inktomi dans le domaine de la recherche d'information en 2002 ?***

Nous oeuvrons dans plusieurs directions : objectif quantitatif en créant des index de plus en plus importants, proposant de plus en plus de pages, en agrégeant plusieurs sources. Mais également une volonté d'augmenter notre pertinence, notamment dans des langues autres que l'anglais, et un travail continu sur nos algorithmes de "ranking" et nos méthodologies.

***Merci, Ron Verheijen !!***

**Note :**

En complément de cet article, notons qu'il existe encore quelques moyens de soumettre un site de façon gratuite à Inktomi, en passant notamment par les sites de ses clients, aux adresses :

<http://submitit.bcentral.com/msnsubmit.htm>

<http://hotbot.spray.se/submit.html>

<http://www.hotbot.lycos.co.uk/submit.html>

<http://www.hotbot.lycos.fr/submit.html>

Essayez également auprès des clients "search" d'Inktomi, pour voir si certains proposent une offre de soumission gratuite :

<http://www.inktomi.com/partners/oem/index.html>

Mais ces soumissions sont-elles réellement prises en compte ? C'est une autre histoire...

## Lancement d'Overture France

[Retour au sommaire de la lettre](#)

C'est désormais officiel : Overture ouvre des bureaux en France. Leader, outre-Atlantique, des liens sponsorisés sur les outils de recherche, Overture arrive en Europe sur un marché où son concurrent Espotting est déjà bien présent, notamment en France et en Allemagne. Christophe Parcot a été nommé responsable d'Overture France, une structure aujourd'hui opérationnelle dans l'Hexagone (la photo ci-contre a été prise sur le site <http://lecarnet.journaldunet.com/>). Il a accepté (merci !!) de répondre à nos questions.



### ***- Bonjour. Pouvez-vous vous présenter à nos lecteurs ?***

En bref, j'ai 33 ans, j'ai fait l'ESSEC en 90, un début chez Hachette Distribution et Services au Canada, 4 ans chez Bertelsmann au sein de la direction internationale des magazines grand public, 4 ans chez LVMH/Groupe Arnault dans des fonctions financières, puis de "business development" et enfin d'investissement Internet. Mi-99, je rejoins Pierre Besnainou et l'équipe fondatrice de Liberty Surf (employé Numéro 5!) en tant que Directeur de la Stratégie et de Développement. Fin 2001, je quitte Liberty Surf devenu le 2ème fournisseur d'accès français et repris par Tiscali, pour rejoindre Overture en Avril dernier en tant que Directeur Général de la filiale française. Ma mission est simple : lancer Overture en France avec le même succès qu'aux Etats Unis dès 1997, en Angleterre en 2001, et en Allemagne en 2002.

### ***- Quelle a été la genèse de la création d'Overture France ?***

Après les succès en Angleterre et en Allemagne, la France est une suite logique à la stratégie européenne ambitieuse d'Overture.

### ***- Quand Overture France sera-t-elle créée officiellement ?***

Overture France est déjà créée. Le lancement opérationnel aura lieu au 3ème trimestre.

### ***- Pouvez-vous nous décrire l'équipe avec laquelle vous allez commencer votre activité en France ? Combien de personnes ? Quel profil ?***

L'équipe sera principalement commerciale, avec rapidement une quinzaine de personne de haut niveau afin de garantir un développement à la fois rapide et pérenne de notre activité sur le marché français.

### ***- Quels sont les marchés les plus "chauds" en Europe aujourd'hui ?***

La France est clairement en tête de liste, avec les marchés d'Europe du Nord, l'Espagne et l'Italie.

### ***- Pourquoi arriver en France "aussi tard" (pardon !!) ?***

Trop tôt .. pour en juger. La stratégie d'Overture est d'étudier en permanence de nouveaux marchés, de valider les opportunités d'investissement sans précipitation et surtout de ne se lancer qu'une fois que les produits et systèmes sont correctement développés et adaptés aux spécificités locales. C'est le prix de la qualité. Cette approche a permis à Overture de s'imposer comme leader incontestable tant aux Etats Unis qu'en Angleterre. Notre ambition est de réaliser la même performance en France. Pour info, Overture se lance également au Japon cette année.

### ***- Ne pensez-vous pas qu'Espotting est aujourd'hui leader en France sur ce marché ?***

Pas difficile d'être leader quand on est seul ! :-)) Mais ce n'est pas le cas partout en Europe. En Grande-bretagne, Overture est devant Espotting, quels que soient les critères pris en compte (CA, nbre d'annonceurs, pénétration, rentabilité, revenu moyen par clic, etc.).

### ***- Quelles sont vos forces par rapport à la concurrence d'Espotting et de Google ?***

Nous n'avons qu'une seule obsession : la qualité et la pertinence de nos résultats notamment grâce à une équipe éditoriale basée à Paris, la rapidité et la fiabilité de nos réponses grâce à nos systèmes et infrastructures techniques à la pointe de l'innovation, et enfin la facilité d'utilisation de nos services pour nos annonceurs.

***- Allez-vous vous occuper du marché français ou francophone (France + Belgique + Suisse + Québec...) ?***

Notre cible prioritaire est constituée de tous les annonceurs souhaitant générer du trafic important et ciblé sur les sites destinés au marché français. Cela dit, les annonceurs du reste de la francophonie seront bien sûr les bienvenus.

***- Quels seront vos premiers portails clients en France ?***

Nous avons annoncé notre partenariat européen avec AOL. La mise en oeuvre de cet accord se déroule dans d'excellentes conditions avec les équipes d'AOL France. Je ne peux pas vous en dire plus pour l'instant, sinon que d'autres partenaires très importants suivront dans les prochaines semaines. Nous sommes très confiants quant à notre capacité à toucher dès le lancement une part très significative des internautes français.

***- Quels sont les retours que vous avez de la part d'Overture US et UK sur le marché des liens sponsorisés ?***

Overture US a annoncé dernièrement des résultats très prometteurs.

***- L'activité est-elle rentable ? Quels sont vos objectifs à ce niveau en France ?***

Le modèle unique inventé et développé par Overture aux Etats Unis puis en Europe est très rentable, et apporte une vision nouvelle de la publicité en ligne en Europe. Je ne peux pas commenter les objectifs d'Overture en France. Comme vous le savez, Overture est une société cotée au Nasdaq, ce qui présente beaucoup d'avantages, notamment pour nos partenaires et annonceurs, mais aussi quelques contraintes en matière de communication ...

***- Comment estimez-vous le marché des liens sponsorisés en France ? Y a-t-il de la place pour les 3 challengers actuels (Espotting, Google et Overture) ?***

Le potentiel est très important. Rendez-vous dans 3 ans pour répondre à la deuxième question. Une seule certitude : nous serons là, et je l'espère leader du marché en Europe.

***- Quels seront vos chantiers prioritaires pour 2002 ?***

Mettre en place les fondations (humaines et marketing) d'un développement commercial réussi pour Overture France dans les prochaines années.

***Merci, Christophe Parcot !!***

## 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 ;-)

-> Google utiliserait actuellement 12 000 serveurs, en jetant 500 machines par jour (soit 5% du parc !!!) pour cause de panne, selon les dires d'un technicien de Google dans une conférence ! Les machines en panne ne sont jamais réparées...

-> Fast (<http://www.fast.no/>) serait en train de s'installer en France. Un représentant est déjà nommé. Plus d'infos dans la lettre R&R du mois prochain, promis :-)

-> La société Iliad (Free, Free Telecom, 3617 Annu) aurait une grosse activité de crawl en ce moment : UserAgent=pompos, crawlers = pompos2.iliad.fr et atsol2.iliad.fr... A priori, un projet d'outil de recherche serait donc en train d'être mis en place chez eux ??

-> TF1, selon certains bruits, serait en train de créer un gros annuaire de sites web francophones pour équiper son site... Et d'autres ???

-> Webformance travaillerait sur un baromètre international du référencement en partenariat avec WebSideStory.

-> La version "Web mondial" du moteur de recherche de Voila.fr devrait très prochainement proposer un index de 250 millions de pages.

-> L'annonce officielle de la création de Google France devrait se faire aux alentours du 15 juin. Vous serez avertis les premiers, bien sûr :-). D'autre part, en juillet, un site sur les "Adwords" en français devrait être mis en ligne par Google avec une équipe en Europe pour traiter l'achat des mots clés...

-> Des nouvelles pages de résultat devraient bientôt (d'ici à début juin) voir le jour sur Voila, Wanadoo et Voila pour les Geeks. Plus d'infos très bientôt...

-> Le site <http://www.lebois.com/> voudrait traîner Google en justice car son site n'apparaît pas dans l'index du moteur...

## En bref...

[Retour au sommaire de la lettre](#)

-> Pour compléter notre article du mois dernier sur le référencement de sites web faits avec des frames ou du Flash : Inktomi prend en compte les pages mère et les pages filles des sites web, donc la balise "noframes" sauf s'il y a trop de pages filles imbriquées les unes dans les autres. Mais, dans une immense majorité de cas, cela ne pose pas de problème. En revanche, il ne prend pas en compte la balise "Noembed" des pages lançant des animations Flash, pour cause de spam trop souvent identifié dans ce type de configuration.

-> Alexa (<http://info.alexa.com/>) a mis en place un nouveau site web dont les résultats sont fournis par Google, mais "à la sauce Alexa"... Intéressant d'y aller faire un tour !

-> Toutes les mascottes de Kartoo sont ici ;-): <http://www.kartoo.net/fr2/pgpourvous07.html>

-> Pour que Google prenne en compte les mots transparents (Stop words), il faut mettre un plus devant le mot en question. Par exemple, une requête sur "the search engine" :

<http://www.google.com/search?hl=en&q=the+search+engine>

ne prend pas en compte le "the", qui est un "stop words", un terme trop fréquent. Le message suivant apparaît : *"the" is a very common word and was not included in your search.*

Mais une requête sur "+the search engine" prend en compte le "the" :

<http://www.google.com/search?hl=en&q=%2Bthe+search+engine&btnG=Google+Search>



Et, là ça marche... Plus d'infos ici :

<http://www.google.com/help/basics.html#stopwords>

-> En rajoutant l'expression &safe=on à une requête sur Voila, on rajoute un filtre familial. Par exemple, une requête sur le mot clé "sexe" :

<http://search.ke.voila.fr/S/voila?gb=site&kw=sexe>

donne 36 réponses du Guide et 226 565 réponses du moteur.

En rajoutant le suffixe &safe=on

<http://search.ke.voila.fr/S/voila?gb=site&kw=sexe&safe=on>

On n'obtient plus que 78501 pages web issues du moteur...

## 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/>

-> Elitel (région Bretagne)

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

-> Tout Nancy (région Lorraine)

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

-> Abcpaca.com (région PACA)

<http://www.efrance.fr/abcpaca/artisan/recherche.asp>

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

[Retour au sommaire de la lettre](#)

-> Skihoo

<http://www.skihoo.com>

Portail et outil de recherche sur la montagne.

-> Russomania

<http://www.russomania.com/annuaire/>

Annuaire de sites web russes.

-> Flipper

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

Métagoûteur + outil de recherche sur le Web invisible.

-> Facile

<http://www.facile.info/>

Nouvel annuaire francophone des sites d'information.

-> Toorista

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

Moteur de recherche de voyages.

## Contenu (sites proposant du contenu ou des fonctions intéressantes)

[Retour au sommaire de la lettre](#)

-> The Virtual Chase

<http://www.virtualchase.com/tvcalert.shtml>

Infos sur les outils de recherche du Web (en anglais)

-> Malaspina

<http://malaspina.links2go.com>

Outil de recherche utilisant une arborescence pour présenter les réponses.

-> Omnidroit.fr

<http://www.omnidroit.fr>

Site professionnel du domaine juridique.

-> Opodo

<http://opodo.fr>

Réservation de billets d'avion et location de voitures (site créé en commun par Aer Lingus, Air France, Alitalia, Austrian Airlines, British Airways, Finnair, Iberia, KLM et Lufthansa).

-> Contrôle parental

[http://www.bonweb.com/p\\_2\\_535.html](http://www.bonweb.com/p_2_535.html)

Une page qui présente la majorité des logiciels de contrôle parental disponibles aujourd'hui... Ca peut servir !

-> TLF

<http://zeus.inalf.fr>

Le Trésor de la Langue Française. Dictionnaire informatisé des XIXe et XXe siècles.

-> Logithèques francophones

[http://www.bonweb.com/p\\_1\\_757.html](http://www.bonweb.com/p_1_757.html)

Liste des principaux sites francophones de téléchargements de logiciels.

-> CyberProtec

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

Site de dépôt en ligne de documents numériques

### Outils (logiciels et sites web proposant des fonctionnalités utiles)

[Retour au sommaire de la lettre](#)

-> WebRankInfo

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

Un site entièrement consacré à la façon d'être mieux positionné sur Google.

-> Spider Food

<http://spider-food.net/>

Plein d'infos sur les spiders des moteurs de recherche principaux.

### Revue d'URL

[Retour au sommaire de la lettre](#)

-> Naissance, vie et mort d'un moteur de recherche

<http://www.infotoday.com/online/may02/OnTheNet.htm>

Un article écrit par Greg Notess, l'un des gourous du domaine, sur les outils de recherche disparus depuis quelques années.

-> Search Engine Optimization - A 10 Step Program

[http://www.netmechanic.com/news/vol5/promo\\_no8.htm](http://www.netmechanic.com/news/vol5/promo_no8.htm)

Comment optimiser son site web en 10 étapes, par Net Mechanic.

-> Power Up With New Metasearch Tools

<http://websearch.about.com/library/weekly/aa070901a.htm>

Les nouveaux métamoteurs du Web, par About Web Search

-> Search Engine Optimization Guarantees - Don't Be Fooled!

<http://c.moreover.com/click/here.pl?e36769085&e=2989>

Un article issu de la publication "Top Site Listings"

-> Ask Jeeves

<http://www.moteurzine.com/archives/moteurzine38.html#entretien>

Entretien avec Steve Berkowitz, Directeur des sites web d'Ask Jeeves.

-> Google

<http://interactif.lemonde.fr/article/0,5611,2852--271800-0,FF.html>

Article du Monde : "Google doit faire face à un renouveau de la concurrence"

-> Ask the Search Engine: Coping with Fraudulent Pay-Per-Click Traffic

<http://searchenginewatch.com/searchday/02/sd0418-ppc-fraud.html>

La fraude sur les systèmes de CPC (Espotting, Overture), cela existe également !

Merci pour votre lecture... Pour toute suggestion : [oa@abondance.com](mailto:oa@abondance.com)