

**François Bourdoncle (Exalead) : "Exalead va bientôt proposer des innovations importantes dans la recherche d'informations"**

[Retour au sommaire de la lettre](#)

<b>Domaine :</b>	<b>Recherche</b>	Référencement
<b>Niveau :</b>	<b>Pour tous</b>	Avancé

*François Bourdoncle est le co-fondateur d'Exalead, entreprise spécialisée dans la recherche intranet, qui propose également un moteur Web qui traite 1 000 requêtes par seconde pour 100 millions d'internautes chaque mois. Il a répondu à nos questions sur la nouvelle version du moteur Web Exalead, qui vient de voir le jour, et nous donne sa vision sur l'avenir de la recherche d'information sur le web. Passionnant...*

**François Bourdoncle** est l'un des pionniers en France du marché des moteurs de recherche. Directeur de la Stratégie d'Exalead (<http://www.exalead.fr/search>), il co-fonde la société en 2000 pour révolutionner le monde de la recherche en entreprise en proposant une plate-forme technologique qui centralise toutes les données de l'entreprise, et permet aux utilisateurs un accès unifié à toutes ces informations.

Chercheur à l'origine, François Bourdoncle est un vrai passionné de technologie : on lui doit le développement de plusieurs langages de programmation, Jazz, un langage dédié aux circuits électroniques et surtout Exascript, un langage orienté-objet XML basé sur la technologie de langage Java.

C'est au laboratoire de recherche de Digital à Paris puis à Palo Alto, en Californie, où il travaille pour le LiveTopics Project d'Alta Vista, qu'il rencontre Patrice Bertin avec qui il fonde Exalead.

X-Mines, Docteur en Informatique, Titulaire d'un Master en Sciences Informatiques de l'Ecole Normale Supérieure, François Bourdoncle est un ancien professeur associé à l'Ecole Polytechnique et à l'Ecole Normale Supérieure. Il est également l'auteur de plusieurs études techniques et publications, et intervient régulièrement dans des conférences spécialisées.

Il a bien voulu répondre à nos questions sur les projets en cours du moteur de recherche Exalead et sa nouvelle version, disponible en ligne depuis quelques jours...



***François Bourdoncle, bonjour et merci de bien vouloir répondre à nos questions. Exalead vient de mettre en ligne une nouvelle version de son moteur de recherche web (<http://actu.abondance.com/2009/07/nouvelle-version-prevue-pour-exalead.html>). Pourquoi cette nouvelle mouture ?***

Salut Olivier ;) Cela fait longtemps que nous n'avons pas fondamentalement changé le look et l'ergonomie de notre moteur Web, et le *design* commençait à vieillir un peu. De plus, nous souhaitons faire évoluer les fonctionnalités et l'ergonomie du site, et cela devenait de plus en plus difficile dans l'ancienne architecture. Et enfin, les dernières améliorations de Google et le lancement de Bing imposaient un rafraîchissement de notre interface, de manière à continuer à offrir des fonctionnalisés innovantes servies par un design et une ergonomie irréprochables. A ce sujet, nous avons décidé d'aligner le design de notre site grand public avec celui de nos solutions entreprises, car nous croyons beaucoup à la convergence des outils grand public et entreprise, surtout en termes d'interface utilisateur. Nous faisons donc bénéficier nos utilisateurs grand public des dernières avancées professionnelles, et à nos utilisateurs professionnels, les dernières avancées en terme d'ergonomie Web 2.0.

### **Quelles sont les différences fondamentales entre Google et Exalead ?**

Tu as sûrement remarqué que Google (avec ses "Search Options" et sa "Wonder Wheel") et Bing (avec sa colonne d'affinage des recherches), se sont appropriés le concept de "zapette" que nous avons introduit sur les moteurs de recherche Web il y a presque 9 ans. Cela est un signe fort que ces technologies de recherche sophistiquées sont maintenant jugées mûres pour être utilisées par un plus vaste public, qui devient avec les années de plus en plus sophistiqué dans son usage des moteurs de recherche.

C'est donc pour nous une très grande satisfaction, car Google et Microsoft sont en train d'éduquer le marché comme nous n'aurions nous même pas pu rêver de le faire seuls. Mais c'est aussi, bien entendu, une menace potentielle, dans la mesure où les innovations que nous avons introduites (et pour certaines, brevetées) il y a quelques années se banalisent progressivement. Nous sommes toutefois sereins sur notre marché entreprise car d'une part nous conservons une longueur d'avance sur Google et Microsoft, et que d'autre part ces derniers se font actuellement la guerre sur Internet et ont montré dans le passé qu'ils innovaient beaucoup plus lentement sur leurs offres professionnelles que sur leur site grand public.

A nous donc de continuer à innover et de garder notre avance technologique. C'est d'ailleurs pour cela que nous consacrons plus de 30% de nos dépenses à la R&D et que nous participons à plusieurs programmes de recherche de l'Agence Nationale de la Recherche, de la Commission Européenne, ou encore au programme de recherche Quaero sur l'indexation multimédia.



*Fonction d'auto-complétion sur la page d'accueil de la nouvelle version du moteur d'Exalead*

### **Après avoir beaucoup travaillé sur la recherche intranet, Exalead a tenté l'aventure du moteur Web "à la Google" avant de, semble-t-il, se recentrer à nouveau sur son métier "cœur de cible" initial. Pourquoi ? La nouvelle version du moteur Web est-elle une nouvelle tentative de marcher sur les plates-bandes de Google ?**

Nous n'avons en réalité jamais changé de stratégie. Le site exalead.com est essentiel pour nous pour plusieurs raisons. J'ai déjà évoqué l'importance pour nous d'être un acteur plein et entier de la recherche grand public pour disposer du savoir-faire en termes d'interface utilisateur nous permettant de proposer à nos clients professionnels ce qui se fait du mieux sur Internet. D'ailleurs, la moitié de nos revenus proviennent de clients professionnels qui ont une activité sur Internet, pour lesquels ce savoir-faire est critique, et si l'on compte le nombre de visiteurs uniques des sites équipés de solutions Exalead, on arrive à 100 millions de visiteurs

uniques par mois, soit environs 1 000 requêtes par seconde, et près de 10% des 11.000 requêtes par seconde de google.com, selon mes informations. C'est donc considérable. Avec un trafic de près de 1 million de visiteurs uniques par mois, le site exalead.com est donc loin d'un google.com, ou même d'un Bing, mais 1 million de personnes utilisant notre site tous les mois, c'est déjà pas mal. Il faut se souvenir que nous n'avons pas 100 millions de dollars à mettre sur la table en publicité chaque année pour développer le trafic, et que le trafic sur notre site est du "vrai" trafic (nous ne faisons quasiment pas de référencement pour booster le trafic de manière artificielle).

Nous espérons donc que le *redesign* du site, avec un look plus moderne et fonctionnel, va fidéliser et élargir notre audience. En particulier, nous avons travaillé à une intégration des liens sponsorisés beaucoup moins intrusive, car l'intégration précédente faisait que certains utilisateurs confondaient les deux... et se plaignaient d'un manque de pertinence du moteur (évidemment, une publicité n'est pas aussi pertinente qu'un résultat organique) ! Nous avons beaucoup travaillé ces dernières années sur la pertinence et je pense que nous n'avons pas à rougir de cette dernière aujourd'hui. Le travail de refonte de l'interface permettra j'espère aux utilisateurs de s'en rendre mieux compte.

### **Un index de 8 milliards de pages, soit au moins 4 à 5 fois plus petit que celui de Google, est-il suffisant actuellement en termes quantitatifs ?**

Tu as raison, c'est devenu un peu "limite", c'est pour cela que nous comptons passer à 16 milliards de pages d'ici la fin de l'année. Toutefois, nous sommes compétitifs sur les pages en français par exemple, qui sont sur-représentées dans notre index, et essayons également de nous concentrer sur les pages des pays importants pour nous, comme les contenus en anglais. Du coup, nous sommes faibles sur les contenus asiatiques par exemple, de manière à optimiser le référencement des contenus qui importent à notre public. Au final, je pense que même si nous allons avoir la moitié du nombre de pages de Google à la fin de l'année, le contenu important pour notre public ne devrait pas nous échapper.

The screenshot shows the Exalead search engine interface. At the top, there's a search bar with the text "moteur de recherche" and a "Rechercher" button. Below the search bar, there are navigation links for "Web", "Images", "Vidéos", "Wikipédia", and "Plus". The main content area displays search results for "moteur de recherche". The first result is "Le Moteur Recherche-Web" with a description and a link to "www.recherche-web.com". Other results include "Google", "CiteSearch: Moteur de recherche Internet", "banned interdit verboden prohibido vietato proibido", and "Mirti, annuaire et moteur de recherche". On the right side, there are filters for "Type de site" (Blog, Forum), "Multimédia" (Audio, Vidéo), "Type de fichier" (pdf), "Termes associés" (Entrez un mot, Mots clef, Version beta, Live Search, Marques déposées), and "Langues" (Français (84%), Anglais (4%), Autres (2%)). There are also two pie charts: one for languages and one for countries.

Page de résultats "Web" de la nouvelle version du moteur d'Exalead

**Sur le marché de la recherche web, est-il encore possible de concurrencer Google en 2009, sachant que Yahoo! et Microsoft, malgré leurs moyens parfois énormes, n'ont pas réellement transformé leurs essais jusqu'à maintenant ?**

Notre objectif n'est pas d'être en frontal avec Google, comme peuvent l'être Bing ou Yahoo, sur le marché "très grand public". Par exemple, Bing a choisi de cibler très clairement le "consommateur" américain. Notre public est un public plus sophistiqué qui cherche des outils d'un genre un peu différent. Tu noteras que malgré un budget de communication très réduit, et aucun référencement visant à gonfler artificiellement notre trafic (98% de notre trafic est naturel), nous possédons tout de même une audience que bien des sites de destination français aimeraient posséder, alors qu'ils dépensent parfois des petites fortunes pour se faire connaître ou référencer. C'est bien la preuve qu'il y a la place pour un moteur de recherche plus sophistiqué dans le paysage de la recherche sur Internet. De plus, il n'est pas exclu que nous décidions de nous démarquer d'avantage dans le futur... Mais je ne peux pas t'en dire plus pour l'instant (même si tu seras le premier prévenu, bien entendu!)

**Enfin, la solution n'est-elle pas dans un changement radical au niveau de l'interface utilisateur ? Une bonne pertinence doublée d'une interface rompant réellement avec les habitudes des liens par pages présentés de façon linéaire ? Ou la différence se fera-t-elle sur les fonctionnalités offertes à l'utilisateur ? En d'autres termes, est-il possible de percer dans ce domaine en "ressemblant à un moteur classique" ? Ne faut-il pas "révolutionner" le domaine pour faire bouger les choses ?**

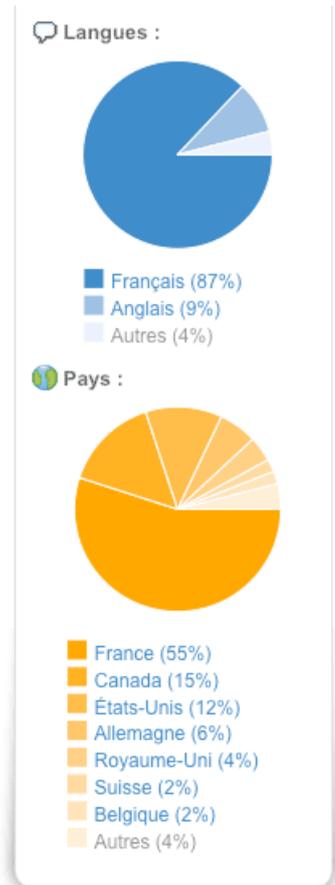
L'interface est l'une des pistes en effet, et nos efforts dans le monde de l'entreprise à cet égard sont très intéressants. Par exemple, nous nous rendons de plus en plus compte que les outils "analytiques" (par exemple *Business intelligence*) permettant par exemple d'extraire des tendances, de tracer des courbes ou des "camemberts" sont très importants. Tu auras sûrement remarqué les deux nouveaux camemberts de la zapette pour affiner par pays ou par langue (*voir ci-contre*). C'est une tendance qui nous semble importante...

Une autre tendance (très) lourde est la prise en compte des approches du Web Sémantique, qui même si c'est très largement une utopie (le fait que le monde entier se mette d'accord sur une modélisation unique de tout le savoir humain), commence à arriver... dans l'entreprise. Les technologies en question sont ce que l'on appelle les "*Linked Data*", à savoir, comment extraire des "faits" (par exemple, les noms de personnes, de sociétés, de technologies, de produits, etc.) et les relations entre ces faits (par exemple le fait qu'une société A ait racheté une autre société B). Une fois que l'on est en mesure de structurer une partie des données *a priori* non structurées que l'on indexe, on peut imaginer des interfaces utilisateur radicalement différentes. C'est précisément l'idée de notre produit à venir CloudView 360, qui permet la mise en oeuvre industrielle de ces technologies sémantiques.

Je pense que la convergence de ces deux tendances va dans un futur relativement proche révolutionner la manière dont les gens conçoivent et utilisent les moteurs de recherche.

**La recherche temps réel de type Twitter est-elle promise à un bel avenir ?**

Bien sûr, quelle question ! Je te rappelle que nous indexons déjà le Web en continu avant que Google n'en fasse autant (te souviens-tu des Google Dance, où Google remplaçait à intervalle régulier son index "pas frais" par un nouveau, "tout frais" ?). Bien sûr, indexer le Web en temps-réel (pour autant que ça ait vraiment un sens) est beaucoup plus difficile que d'indexer un petit corpus comme celui de Twitter, mais c'est le sens de l'histoire, indubitablement, et cela fait longtemps que nous développons et affinons les technologies nécessaires.



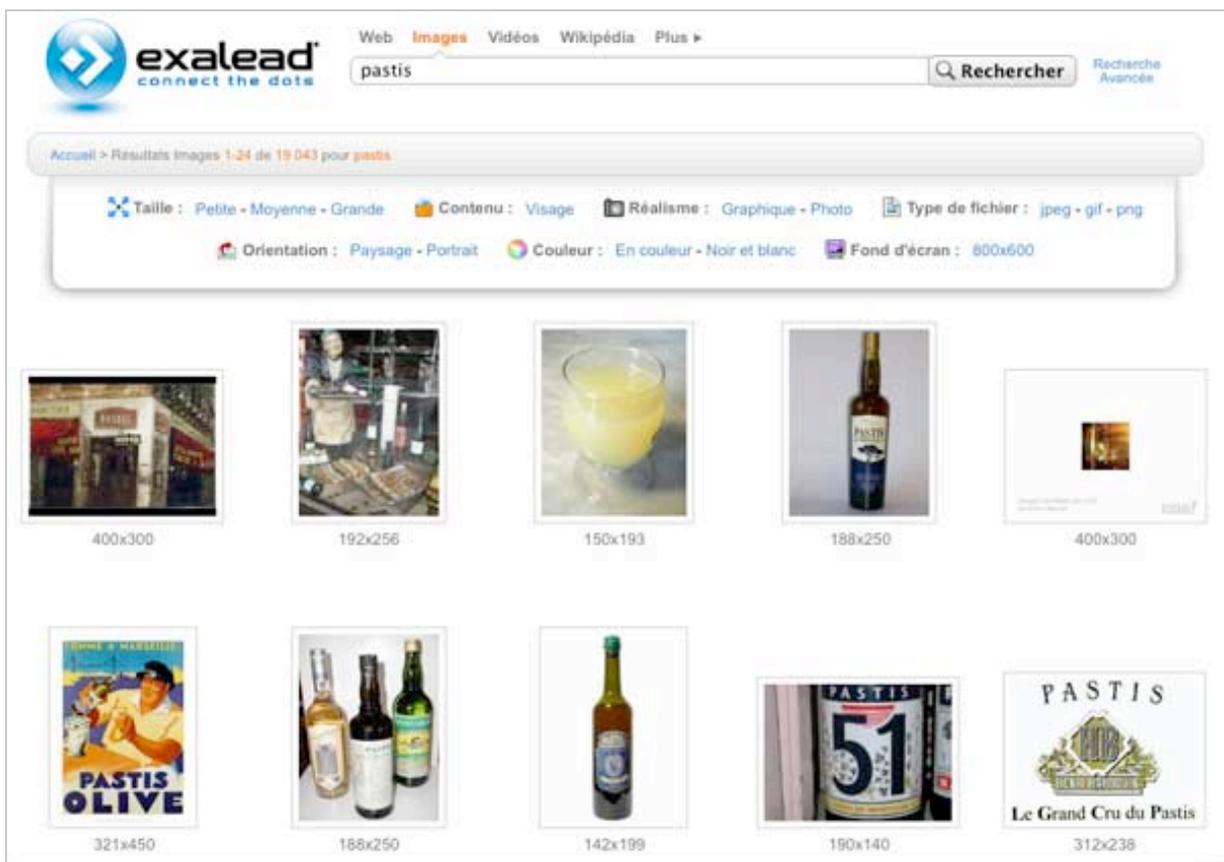
### **Quel est l'avenir de la recherche d'information sur le Web et des moteurs de recherche dans les années qui viennent ?**

J'ai déjà largement répondu précédemment : par la structuration d'une partie de plus en plus importante des contenus, et par la prise en compte de cette structure dans l'interaction avec les utilisateurs, un peu comme l'on fait en *Business Intelligence*. De ce point de vue, des projets de recherche comme Google Squared (<http://www.google.com/squared>) ou des prototypes comme Wolfram|Alpha (<http://www.wolframalpha.com/>) vont clairement dans cette direction, même si Google Squared semble encore assez peu "sémantique" et Wolfram|Alpha est très en retrait en terme de technologies du traitement de la langue.

### **Quels sont les objectifs du nouveau moteur d'Exalead ?**

Notre objectif est d'une part une harmonisation de nos interfaces grand public et entreprise, pour une meilleure capitalisation entre les deux. Et nous en profitons bien sûr pour remettre le site au goût du jour afin de rester dans la course. Mais la raison fondamentale est de mettre en place une nouvelle architecture informatique nous permettant dans le futur de déployer des innovations plus rapidement.

A ce propos, il me faut mentionner tout de même quelques innovations qui font partie de cette nouvelle version, comme par exemple un outil de suggestion dans la barre des requêtes, l'apparition de "related searches" sous la barre de recherche, la possibilité de configurer sa zapette comme on l'entend (cliquer sur la clef à molettes !), ou encore une recherche d'images avec une capacité "d'auto-scroll" (quand vous faites défiler la page et arrivez vers la fin, de nouvelles images s'affichent sans que vous n'ayez rien à faire). A propos de l'outil sur les suggestions, tu pourras te rendre compte qu'il va parfois plus loin que celui de nos concurrents en prenant en compte la phonétique quand il ne trouve pas de suggestions autrement.



Page de résultats "Images" de la prochaine version du moteur d'Exalead

### *Quelques nouvelles du projet Quaero peut-être pour terminer ?*

Quaero (<http://www.quaero.org/>) suit son cours de manière tout à fait satisfaisante, et les premières innovations commencent à sortir des labos après un an (seulement) d'activité. L'une des raisons du redesign de notre site est précisément de pouvoir intégrer plus facilement certaines innovations issues du programme Quaero. Par exemple, nos labos ont sorti un fabuleux outil de recherche dans les vidéo de news.

Cet outil, appelé **Voxalead News** (<http://voxaleadnews.labs.exalead.com/>) indexe en temps réel, des flux audio et vidéo de news en français, anglais, arabe, et chinois. Il résulte de l'intégration de technologies de partenaires de Quaero, en particulier la technologie de reconnaissance de la parole de VECSYS qui nous permet d'indexer la bande son des contenus vidéo, et donc de recherche dans ce qui est dit (et bientôt, écrit), dans les vidéo. Cela permet une puissance de recherche qui n'existe pas sur des sites du type YouTube ou DailyMotion qui n'indexent que les tags décrivant la vidéo, et non le contenu de la vidéo elle-même. De plus, nous passons le texte de la bande son dans nos outils d'extraction sémantique et permettons une navigation thématique dans les contenus grâce à un "nuage de tags". Par ailleurs, la page d'accueil affiche la synthèse de l'actualité de la journée, de la semaine, ou du mois. Et pour finir, *last but not least*, la vidéo est affichée avec une fonction "karaoké" qui surligne en bleu pâle le texte qui est en train d'être prononcé dans la vidéo. Et je laisse découvrir aux lecteurs toutes les fonctions cachées de ce prototype que je trouve vraiment fabuleux, comme par exemple la "time line" de la page de résultats.

Ah, et j'oubliais, il existe une version iPhone, disponible sur l'App Store d'iTunes (il suffit de chercher "voxalead"). Et en plus, c'est gratuit ;)

*Merci, François Bourdoncle, d'avoir bien voulu répondre à nos questions.*

**Interview menée par mail par Olivier Andrieu, éditeur du site Abondance.com, les 6 et 7 juillet 2009**

**Réagissez à cet article sur le blog des abonnés d'Abondance :**

<http://abonnes.abondance.com/blogpro/2009/07/francois-bourdoncle-exalead-exalead-va.html>