

Le nouveau moteur de recherche "Multimedia" d'AOL.FR

Vous le savez certainement si vous êtes un fervent lecteur du site Abondance (<http://actu.abondance.com/2003-30/aol.html> et <http://actu.abondance.com/2003-31/aol.html>), le site français d'AOL.FR a mis en place de nouvelles fonctionnalités de recherche multimédia. Disponible à l'adresse <http://recherche.aol.fr/>, cette nouvelle interface donne, bien sûr, accès au Web francophone et mondial, mais également à des recherches d'images, de fichiers audio et vidéo. La technologie utilisée pour la recherche d'images est celle de Google, alors que c'est Exalead qui fournit le système de recherche audio et vidéo. Nous avons donc eu envie de comparer cet outil à ceux qui font référence dans le monde de la recherche multimedia sur Internet, c'est-à-dire Google (pour l'image) et AltaVista (pour l'audio et la vidéo).













Comparaison des outils de recherche

En effet, Google avait gagné le comparatif des outils de recherche d'images que nous avons effectué au mois d'avril 2003 dans la lettre Recherche & Référencement. Nous avons donc appliqué les mêmes critères quantitatifs, qualitatifs et fonctionnels à ce duel entre la recherche "images" d'AOL.FR et celle de Google. Même si Google fournit les résultats aux deux outils, il nous a semblé intéressant de vérifier que les images fournies étaient sensiblement similaires dans les deux cas et qu'il n'était pas utile d'utiliser Google "en natif" lorsqu'on était sur le site AOL.FR. Pour ce qui est des recherches audio et vidéo, nous avons pris comme référent AltaVista, qui est le plus souvent considéré par les observateurs du Web comme le moteur le plus efficace aujourd'hui pour ce type de fichiers. Pour mémoire, rappelons que Google ne propose pas (encore) ce type de recherche.

Etudions donc de plus près la fonction "Multimedia" d'AOL.FR comparée à celles de Google et AltaVista. Pour chacun d'entre eux, nous avons essayé d'indiquer des critères quantitatifs, qualitatifs et fonctionnels pour tenter de mieux les comparer. Si vous êtes un fervent lecteur de la lettre R&R, cette méthodologie ne doit pas vous être étrangère... :-)

Comparatif quantitatif

Dans un premier temps, nous avons tapé 12 mots clés sur chaque outil de recherche et avons noté le nombre de résultats proposé par chaque outil :

Recherche d'images			Recherche de vidéos			Recherche de sons		
Moteur :			Moteur :			Moteur :		
Technologie fournie par :			Technologie fournie par :			Technologie fournie par :		
vercingétorix	499	563	clinton	999	124	clinton	3 451	866
notre-dame	7 770	7 890	lorie	3	9	lorie	76	51
chien	18 000	21 900	beatles	52	55	beatles	5 383	1 181
lorie	3 000	3 420	vivendi	5	29	cigogne	1	40
renault	56 300	62 000	tennis	908	198	tennis	1 543	258
chirac	4 120	5 040	humour	57	893	humour	1 009	3 438
logo apple	5 290	5 520	guignols	2	56	guignols	568	149
montgolfière	2 090	2 340	lynnyrd skynnyrd	1	3	lynnyrd skynnyrd	271	64
engrenage	411	450	"taxi 3"	1	3	"le grand bleu"	8	11
samy naceri	99	128	disney	99	128	disney	2 413	815
porquerolles	2 480	2 600	alيزée	466	173	alيزée	45	91
louis XVI	3 140	3 520	iraq	1 270	183	noah	1 264	189
Moyenne :	17,80	20	Moyenne :	15,79	20	Moyenne :	20	13,20

Recherche d'images : remarque étonnante : AOL.FR renvoie quasi systématiquement entre 5 et 15% de résultats de plus que Google... Ce qui explique la meilleure note obtenue par AOL.FR à nos tests. AOL.FR - via la technologie Google - pourrait donc proposer plus d'images que Google sur son propre site ? Etrange.













Recherche de vidéos : AOL.FR remporte le match contre AltaVista, même si on aurait pu s'attendre à des résultats plus en faveur de la technologie française (Exalead) pour des mots clés à consonnance très francophone comme Alizée ou Lorie.

Recherche de sons : Dans ce cas, c'est AltaVista qui l'emporte. De façon étonnante, ce moteur emporte le duel sur des mots clés comme "guignols" ou "Lorie" (mots clés typiquement français pour ce type de recherche), sur lesquels on aurait plutôt attendu une victoire d'AOL.FR.

Rappelons cependant que l'outil de recherche multimédia d'AOL.Fr est encore très récent (quelques semaines) et que son index va certainement s'étoffer fortement dans les semaines qui viennent...

Comparatif qualitatif

Pour comparer de façon qualitative les deux outils, nous avons tapé les mêmes 12 requêtes sur chacun d'eux. Puis, nous avons évalué les 20 premières réponses (les 20 premiers documents proposés) et noté combien, parmi elles, étaient pertinentes. La notion de pertinence pouvait être assez générale, en fonction du mot clé. Pour l'expression "samy naceri", par exemple, nous n'avons pris en compte que les images de l'acteur, pas celle de l'affiche du film "Taxi".













Recherche d'images			Recherche de vidéos			Recherche de sons		
Moteur :			Moteur :			Moteur :		
Technologie fournie par :			Technologie fournie par :			Technologie fournie par :		
vercingétorix	20	20	clinton	20	9	clinton	20	20
notre-dame	18	18	lorie	0	1	lorie	15	7
chien	18	17	beatles	19	6	beatles	17	18
lorie	19	16	vivendi	2	2	cigogne	1	2
renault	20	20	tennis	20	8	tennis	14	17
chirac	20	20	humour	20	20	humour	20	18
logo apple	18	16	guignols	2	14	guignols	20	10
montgolfière	20	20	lynrd skynyrd	1	0	lynrd skynyrd	20	9
engrenage	19	19	"taxi 3"	0	1	"le grand bleu"	6	7
samy naceri	2	3	disney	2	8	disney	18	12
porquerolles	20	18	alizée	0	16	alizée	18	11
louis XVI	17	16	iraq	20	17	yannick noah	3	10
Moyenne :	19,44	19,19	Moyenne :	16,40	20	Moyenne :	20	18,41

Recherche d'images : Assez logiquement, les deux outils sont d'une pertinence quasi équivalente. Google bat parfois AOL.FR d'une courte tête, mais, en règle générale, les deux moteurs sont très très proches. Plus de 90% des images retournées par les deux sites sont d'ailleurs les mêmes, et quasiment dans le même classement... Très peu de résultats différent de l'un à l'autre. Les pages de résultats d'AOL.FR et de Google sont pratiquement identiques (mais pas complètement). On s'y attendait un peu puisque les sources sont les mêmes.

Recherche de vidéos : les deux outils ne sont pas si éloignés que leurs scores respectifs pourraient nous le laisser envisager. Chacun gagne 5 duels. Mais les écarts sont plus grands lorsque AOL.FR gagne, ce qui explique la note finale en sa faveur.

Recherche de sons : là aussi, les deux outils sont très proches, mais le classement montre une légère préférence pour AltaVista, qui gagne 6 duels contre 5 à son challenger.

Comparatif fonctionnel

Recherche d'images			Recherche de vidéos			Recherche de sons		
Moteur :	 		Moteur :	 		Moteur :	 	
Technologie fournie par :	 		Technologie fournie par :	 		Technologie fournie par :	 	
Syntaxe avancée	OR, -, guillemets, filetype: ???		Syntaxe avancée	+, -, guillemets ???		Syntaxe avancée	+, -, guillemets ???	
Recherche avancée / filtres	Taille, format, couleur, domaine, filtre parental (US uniquement)	Filtre parental	Recherche avancée / filtres	Format, durée, source	Non	Recherche avancée / filtres	Format, durée, source	Non
Informations données	nom, taille, poids, url	nom, taille, poids, url	Informations données	nom, taille, durée	nom, url, résumé	Informations données	Titre, auteur, mono / stéréo, durée, url, autres fichiers de la page	Titre, url, résumé, autres fichiers de la page
Moyenne :	18	12	Moyenne :	15	14	Moyenne :	14	16

Recherche d'images : Sur le plan fonctionnel, AOL.FR reste un peu en deça de Google. Pas de syntaxe avancée (OU, ET, guillemets, recherche par type de fichiers - gif ou jpeg - etc.), ou en tout cas aucune indication sur celle-ci sur le site, seule une zone de recherche avancée globale à la recherche web sans spécificité "images" étant proposée. Bref, Google est plus avancé que son concurrent sur ce plan, notamment avec sa recherche avancée

(http://images.google.fr/advanced_image_search?svnum=20&hl=fr&lr=&ie=UTF-8&newwindow=1) très intéressante (recherche de fonds d'écran, de fichiers en couleurs ou noir et blanc, etc.).

L'orientation très "grand public" d'AOL.FR explique peut-être cette absence de fonctions de recherche avancée. Avantage certain, cependant, d'AOL.FR : le filtre familial, disponible en français, ce qui n'est pas le cas de Google.

Recherche de vidéos : La recherche avancée manque cruellement à AOL.FR pour cet outil. On aurait également apprécié l'indication de la taille des fichiers ou de la durée de la vidéo, afin d'éviter des téléchargements longs et fastidieux (mais, bonne nouvelle, cette indication est prévue dans la prochaine version de l'outil). Léger avantage à AltaVista à ce niveau.

Recherche de sons : Comme auparavant, AOL.FR pâtit d'absence de fonctionnalités de recherche avancée, notamment au niveau du format (wav, mp3, avi, etc.). Mais la lisibilité de ses pages de résultats, notamment grâce au résumé fourni à droite du fichier, est globalement meilleure.

Récapitulatif

Recherche d'images			Recherche de vidéos			Recherche de sons		
Moteur :	 		Moteur :	 		Moteur :	 	
Technologie fournie par :	 		Technologie fournie par :	 		Technologie fournie par :	 	
Quantitatif	17,80	20	Quantitatif	15,79	20	Quantitatif	20	13,20
Qualitatif	19,44	19,19	Qualitatif	16,40	20	Qualitatif	20	18,41
Fonctionnel	18	12	Fonctionnel	15	14	Fonctionnel	14	16
Moyenne :	18,67	17,60	Moyenne :	15,90	18,50	Moyenne :	18,50	16,50
	La pertinence, la recherche	La	Les "Plus"	La recherche avancée,	L'affichage très clair.		La pertinence, la	L'affichage

	avancée, la recherche sur des mots clés anglais.	
Les "Moins"		L'absence de possibilité de recherche avancée

	l'indication de la durée.	
Les "Moins"	La pertinence parfois moyenne	L'absence de possibilité de recherche avancée

	recherche avancée, l'indication de la durée.	
Les "Moins"		L'absence de possibilité de recherche avancée

Pour calculer la note finale, nous avons donné un coefficient 1 à la note des comparatifs quantitatifs et fonctionnel et un coefficient 2 à la note du comparatif qualitatif, car celui-ci nous semble le plus important.

Google sort donc vainqueur de très peu du comparatif "recherche d'images", mais AOL.FR semble très proche de son challenger. On peut donc utiliser indifféremment Google ou AOL.FR, les résultats seront, de façon logique (mais encore fallait-il le vérifier), quasi identiques. On aurait cependant apprécié, chez AOL.FR (la remarque est valable pour les 3 types de requêtes), une interface et une url spécifiques à la recherche d'images, ce qui éviterait d'utiliser obligatoirement et systématiquement la souris pour cliquer sur l'icone "images" afin de lancer la recherche. L'utilisation de la touche "Entrée" fait parfois gagner bien du temps, ce qui ne semble pas possible sur cet outil. Pêché de jeunesse ? C'est pourtant à ce type de petits détails que s'attachent souvent les internautes, surtout s'ils sont gros consommateurs de ce type d'outil...

Pour ce qui est des autres types de fichiers multimedia, les deux outils semblent finalement assez proches, avec un léger avantage à AltaVista pour la recherche de sons, AOL.FR prenant sa revanche pour la recherche de vidéos. Ce résultat est assez remarquable pour le moteur d'AOL.FR, qui n'a que quelques semaines d'existence, car n'oublions pas qu'AltaVista dispose d'un index mondial, ce qui est loin d'être négligeable pour des requêtes identiques en français et en anglais comme "Clinton", "Beatles" ou "tennis". La recherche en mode contextuel développée par Exalead (voir interview ci-après), et qui est l'une des grandes innovations de cette technologie, fait que, même si le nombre de pages de l'index de départ est assez réduit, le nombre de résultats retournés par AOL.FR est souvent assez important.

Cependant, si nous avons apprécié la très bonne lisibilité des pages de résultats d'AOL.FR (l'une des meilleures du Web certainement à l'heure actuelle), nous avons souvent préféré la pertinence, notamment pour les fichiers sons, d'AltaVista. Comme souvent, un "mix" des deux outils, prenant chez l'autre ce qu'il a de meilleur, serait parfait :-). Mais il y a fort à parier que lorsque AOL.FR aura "musclé" son index multimedia, cet outil prendra un avantage non négligeable par rapport à son concurrent. Affaire à suivre...

Interview de François Bourdoncle, Exalead

Pour en savoir plus sur le nouvel outil de recherche multimedia d'AOL.FR, nous avons posé quelques questions à François Bourdoncle, PDG de la société Exalead, qui a fourni la technologie de ce nouveau moteur de recherche.

- Pouvez-vous nous présenter la nouvelle recherche multimédia d'AOL.Fr en quelques chiffres ?

La recherche multimédia d'AOL recense aujourd'hui près d'un million et demi de fichiers audio et vidéo, dont près de 20% de fichiers vidéo, une proportion plus élevée que ce à quoi je m'attendais. Ces fichiers sont "pointés" par des documents sur le web, au travers de liens hypertexte, et le moteur vérifie leur existence avant de les référencer. L'index est mis à jour toutes les semaines. Les fichiers multimedia



sont donc identifiés par des liens existant sur eux dans des pages web rencontrées par notre crawler.

- Comment cette recherche fonctionne-t-elle ?

L'innovation que nous avons développée pour cet outil est d'utiliser une recherche "en contexte" permettant de rechercher non seulement sur les noms des fichiers multimedia mais aussi sur leurs méta-données (les "champs cachés" des fichiers multimedia comme le nom de l'auteur, etc.), mais également dans les pages qui référencent ces fichiers. Cela permet un ciblage des résultats beaucoup plus fin que ce que permettent les autres technologies, et également de beaucoup moins courir le risque de manquer un fichier multimedia. Par exemple, si vous cherchez "Natalie Dessay" vous trouverez plusieurs extraits audio dont aucun ne pourrait être retrouvé sur le nom des fichiers seuls (qui pourrait s'appeler dessay.mp3 ou ndess.mp3 par exemple). Cette technologie est donc très efficace pour dénicher des extraits musicaux autrement difficiles à trouver!

- Quels sont les points de différenciation entre cette recherche et celle proposée par Google, Fast, AltaVista et les autres ?

Essentiellement, nous prenons en compte le contexte, et cela change tout ! De plus, notre approche permet d'utiliser toute la finesse du *ranking* du moteur Exalead, mise au point depuis plus de cinq ans maintenant, pour en faire profiter la recherche multimedia.

- Est-elle limitée aux fichiers du web francophone ?

Aujourd'hui, nous nous limitons pour AOL au même contour que la recherche Web, mais la capacité du moteur à dénicher des fichiers de manière beaucoup plus "profonde" que les autres moteurs multimedia fait que notre base semble parfois plus riche que celle des moteurs mondiaux !

- La recherche multimédia repose-t-elle sur les mêmes technologies et algorithmes que la recherche "Web" d'Exalead ?

Absolument, nous utilisons [notre technologie ExaScript](#) pour "scripter" notre "crawler" et notre moteur de manière à faire du traitement intelligent de l'information lors de sa collecte et de son indexation. C'est comme cela par exemple que nous "catégorisons" les pages ou détectons les spammers.

ExaScript, qui est un langage compilé qui étend le modèle de données objet de Java pour permettre une manipulation du XML (ici, du XHTML) et qui offre des outils de filtrage par motif du type de ceux utilisés dans PERL, est particulièrement adapté à cette tâche, et permet des développements et des innovations très rapides. L'ensemble de notre plateforme est en réalité "scriptable" en ExaScript.

- Cette recherche est-elle exclusive à AOL.FR ou pensez-vous la proposer à d'autres portails francophones ou autres ?

Dans la mesure où notre plateforme est un vrai caméléon grâce à ExaScript, nous pouvons proposer des variantes de ce genre de recherche, ou d'autres innovations, à qui est lassé du "prêt à porter" en matière de moteur de recherche et est intéressé par un moteur "sur mesure" ;-)

- Quelles sont les principales innovations à venir sur Exalead.fr ou AOL.FR ?

Elles sont nombreuses, mais je préfère ne pas trop déflorer le sujet. Ce qui est clair, c'est que notre plateforme logicielle unique, utilisée à la fois pour le Web mais aussi pour les applications d'entreprise, nous permet de mettre au profit des moteurs de recherche Web des technologies professionnelles comme par exemple la mise à jour en temps réel des bases indexées, une innovation qui nous permet de faire converger les technologies des moteurs de recherche des technologies de bases de données plus traditionnelle. C'est pour nous le Graal de la recherche d'information: grand volume de données, recherche plein-texte sur des données semi-structurées, et mise à jour en temps réel. Je ne serais pas surpris de voir des innovations en ce sens ;-)
Et enfin, nos amis cruciverbistes devraient beaucoup aimer les nouvelles technologies de requête qui sont en train de sortir des labos (imaginez pouvoir savoir quels sont les mots qui commencent par "a", qui ont ensuite une lettre, puis un "d" et finissent par "u"?) !