

Il n'a échappé à personne qu'AltaVista a récemment lancé une nouvelle version de son outil de recherche (voir article de ce mois). De façon assez logique, nous avons eu envie de tester ce nouvel outil sous la forme d'un "duel au sommet" avec le site Google pour avoir une idée plus précise des améliorations qui avaient été portées, au nouvel algorithme de pertinence notamment, à cet AltaVista "new look".

L'idée est donc de comparer (comme nous l'avions déjà fait précédemment entre AOL.FR et Google.fr au mois de mai dernier), 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 2001, 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 mercredi 13 novembre 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 notamment pour comparer ce qui est comparable, nous avons pris en compte les résultats fournis par le site Google France (<http://www.google.fr/>) avec l'option "Pages francophones" cochée, et celles d'AltaVista France (<http://fr.altavista.com/>) avec l'option "Rechercher : France" pour tester l'index local réservé aux internautes francophones. La difficulté, ici, est que l'option choisie est présentée par défaut sur AltaVista France mais demande à l'internaute de choisir l'option en cliquant sur le bouton radio adéquat sur Google. Il nous a quand même semblé plus important d'effectuer des tests amenant à des résultats réellement comparables plutôt que d'utiliser les pages d'accueil des deux moteurs "telles quelles".

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	2 064	8 200
fabricant amortisseur	133	700
vin colmar	1 290	3 440
nestlé environnement	1 253	4 980
musées paris	31 185	90 300
astérix officiel	1 134	5 910

eye tracking	166	790
photos paris	219 020	650 000
exemples business plan	510	17 700
vidéos humour	21 131	215 000
virus sircam	1 108	3 200
location studio paris	240	52 600
téléphone portable santé	5 025	20 900
MP3	264 579	984 000
recette couscous	448	12 200
jeux vidéo	423 763	1 380 000
warrants	7 002	105 000
sonneries GSM	4 564	49 700
convertisseur euro	2 998	53 200
hotmail	58 642	493 000
sexe	409 019	5 110 000
cartes voeux	8 542	80 800
fond d'écran	20 137	108 000
webcam	231 204	985 000
immobilier	686 166	722 000
référencement	176 874	199 000
euro	323 177	1 240 000
bourse	737 231	845 000
loftstory	7 893	42 900
yahoo	106 619	1 670 000

Pour noter ce critère, nous avons ôté du calcul trois mots clés ou expressions pour lesquels les résultats nous semblaient par trop éloignés sur les deux outils ("location studio paris", "exemples business plan" et "recette couscous") et regardé l'écart moyen entre les deux résultats. Résultat : Google donne, dans tous les cas, des résultats supérieurs en nombre à ceux d'AltaVista, ce qui n'est pas surprenant puisque Google annonce un index global de 3 milliards de pages, alors que celui d'AltaVista dépasse légèrement le milliard. Il est donc logique que cet état de fait se concrétise dans les chiffres d'un test quantitatif. Les résultats quantitatifs de Google sont supérieurs à ceux d'Altavista dans un coefficient allant de 1,15 à 15,66 (moyenne : 7). Mais, si l'on additionne le nombre de résultats global (hormis les 3 requêtes exclues, voir paragraphe précédent), Google.fr renvoie en tout 15 071 020 liens, contre 3 751 919 à Altavista France, soit 4,02 fois plus.

En tenant compte de ces différents résultats, nous avons décidé d'attribuer la note de 10 à Google.fr et de 3 à AltaVista France, 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...

		
Critère quantitatif	3	10

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.

Voici les résultats :

		
Syntaxe d'interrogation	10	8

La syntaxe de Google est assez complète, même si elle mériterait d'être encore plus efficace pour arriver au niveau de celle d'Altavista. Côté AltaVista, il s'agit presque d'un lieu commun que de dire que sa syntaxe avancée (host:, url:, filetype:, domain:, link:), l'une des premières à avoir été mise en ligne sur le Web, est certainement ce qui se fait de mieux dans le domaine... Même si on pourrait imaginer d'autres fonctionnalités encore (recherche uniquement dans les balises Meta par exemple), on est ici proche de la perfection, car toutes les fonctionnalités "efficaces" de recherche sont présentes.

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.).

Voici les résultats :

		
Recherche avancée	10	9

A ce niveau, les deux outils proposent des zones de recherche avancée tout à fait conformes à ce que l'on peut attendre de leaders du domaine. Altavista en propose même deux ("Plus de précisions" et "Recherche avancée"). On peut estimer que les deux adversaires sont ici presque ex-aequo. Nous avons quand même donné un point d'avance à AltaVista car il propose deux zones complémentaires et également pour sa fonctionnalité de recherche par dates, plus complète et efficace.

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), option Prisma (AltaVista), etc.

Voici les résultats :

		
Qualité de la présentation des résultats	8	9

Les pages de résultats de Google sont très simples et efficaces. La fonction [Cache], notamment, est aujourd'hui inégalée sur le Web. Les pages sont dépouillées et on a rapidement accès aux infos proposées. Google est connu pour la simplicité de ses pages, ce qui en fait l'une de ses forces. Le grand "plus" de la nouvelle version d'AltaVista reste la fonction "Prisma" qui propose de nouveaux mots clés de recherche. Une fonctionnalité proche de ce que propose Exalead sur AOL.FR, qui se révèle très utile, pertinente et efficace. Le panorama est cependant un peu obscurci par les bandeaux publicitaires qui, s'ils sont absents de la page d'accueil, "envahissent" un peu les pages résultats : 1 "skyscraper", 2 "bannières", etc. Sans oublier les liens de partenaires en bas de page (Kelkoo, rencontres, informatique, jeux, Tiscali, etc.) et 6 liens Espotting (3 en haut de page, 3 en bas). Bon, il faut bien vivre ;-), mais il est vrai que la publicité est un peu envahissante sur cette page de résultats. Dommage... Ceci dit, la nouvelle présentation des liens est beaucoup plus claire que l'ancienne sur AltaVista, ce qui est plutôt un bon point.

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.



Voici les résultats :

		
Délai d'affichage	8	10

Si Google nous a semblé conforme à sa réputation de moteur ultra-rapide lors de nos tests, AltaVista a semblé un peu plus "long à la détente" pour afficher sa page de résultats. Peut-être est-ce également dû à un ralentissement dû à l'annonce de la nouvelle version, générant de nombreuses nouvelles connexions... En règle générale, les temps de réponse d'AltaVista étaient cependant assez bons, sans être pourtant à la hauteur de ceux de Google...

Qualité de l'annuaire associé

Les deux outils utilisent un annuaire de sites : L'Open Directory pour Google, Looksmart France pour AltaVista. 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. Voici les résultats :

		
Qualité de l'annuaire associé	5	6

L'annuaire n'est pas la force principale de ces deux outils. AltaVista utilise Looksmart France, annuaire qui n'a plus d'équipe (commerciale) en France et qui n'est plus maintenu que "du bout du clavier" par Looksmart pour qui la France n'est semble-t-il plus une priorité. Aucune soumission payante ou gratuite n'est d'ailleurs plus possible sur cet outil. L'Open Directory utilisé par Google, pour sa partie francophone, nous semble, lui, encore un peu "inégal" (son mode de fonctionnement particulier, basé sur le bénévolat des documentalistes qui y travaillent en est certainement la cause) même si l'annuaire a dernièrement dépassé les 90 000 sites francophones, ce qui reste un chiffre intéressant.

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 viticultriceur, technologies d'eye-tracking, business plan), informatique (virus) ou autres (téléphones portable, musées, etc.) - ont été choisies. Certains diront que 20 requêtes, c'est trop peu pour un tel test. Et ils auront raison :-). Mais il s'agit d'une première approche, car il n'est pas sûr que même 100 requêtes soient suffisantes pour bien juger de la pertinence des réponses d'un moteur de recherche. Mais l'essentiel, ici, est surtout de comparer les deux moteurs l'un par rapport à l'autre et non la pertinence globale de l'un, pris de façon isolée. Voici la liste des mots clés et expressions qui ont servi à notre test :

Demande	Mots clés
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 viticultriceur 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.

Voici les résultats :

		
Pertinence des réponses	4	6,5

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 28 / 46. AltaVista a obtenu la moyenne de 17,85 / 46.

On peut noter, au vu de ce test, un certain nombre d'informations intéressantes :

- Si l'écart de notes entre les deux moteurs peut paraître important (2,5 points sur 10), il nous semble qu'en réalité, l'écart n'est pas si important. En effet, sur une immense majorité des requêtes effectuées, de nombreux liens pertinents identiques étaient proposés à la fois sur Google et AltaVista. La présence simultanée des mêmes résultats sur les 2 moteurs semble prouver qu'AltaVista a effectué ses tests de pertinence avec des écrans "Google" à côté de lui pour s'approcher au plus près de la pertinence de son concurrent. C'est de bonne guerre... Et il semble n'en être pas si loin que cela... En fait, si seuls les 3 premiers liens avaient été pris en considération, les deux moteurs auraient été plus proches l'un de l'autre. Mais Google est globalement plus pertinent que son adversaire du jour sur les liens proposés en position 4 à 10.

- En effet, sur une majorité de recherches, la différence s'est plutôt faite sur le nombre de liens non pertinents proposés. Disons que, en moyenne, Google proposait pour une requête 6 à 8 liens pertinents et AltaVista 4 à 6. La différence entre les notes des 2 moteurs s'explique clairement par ce phénomène. Qu'AltaVista ramène plus de liens pertinents dans ses pages de résultats (qui en contiennent déjà) et il aura comblé le dernier fossé qui le sépare de Google...



- D'autre part, Google a été pertinent de façon assez homogène sur les 20 requêtes. AltaVista, lui, a connu quelques manques sur 2 ou 3 requêtes avec des pages résultats assez faibles. Un manque d'homogénéité pour AltaVista qui pourrait s'effacer lorsque le moteur aura pris un peu de "bouteille" et que les bugs de jeunesse auront été effacés...

- Les deux moteurs proposent encore des erreurs 404 mais pas énormément et pas l'un plus que l'autre.

- Une tentative de spam sur AltaVista (requête "photos paris") a été détectée. Il s'agit de la seule dans ce domaine. En revanche, il nous a semblé que Google était plus sensible aux pages satellites que son concurrent. Il nous a semblé en voir s'afficher furtivement un certain nombre...

Récapitulatif

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

		
Critère quantitatif	3	10
Syntaxe d'interrogation	10	8

Recherche avancée	10	9
Qualité de la présentation des résultats	8	9
Délai d'affichage	8	10
Qualité de l'annuaire associé	5	6
Pertinence des réponses	4	6,5

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 :

Critère	Coefficient
Critère quantitatif	1
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	1
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>• Population "Utilisateur débutant" :</p> <p>1. Google.fr : 136,5 points</p>

2. Altavista.fr : 108 points

• **Population "Utilisateur avancé" :**

1. Google.fr : 152,5 points

2. Altavista.fr : 130 points

Conclusion

Google sort vainqueur de ce duel avec AltaVista, sans conteste possible. Pourtant, après avoir mené à bien tous ces tests, nous avons l'impression qu'AltaVista, avec cette nouvelle version, s'est fortement approché du leader actuel des moteurs de recherche et qu'il peut, au prix de certains changements, concurrencer de façon encore plus forte Google dans les mois qui viennent. La marge qui sépare les deux "combattants" ne nous semble pas si grande que cela. AltaVista pourrait, par exemple, prendre en compte certaines améliorations futures comme :

- La prise en compte d'un annuaire de meilleure facture et mieux maintenu que celui de Looksmart France.

- Ramener plus de liens pertinents sur une même page de résultats pour une requête donnée.

- Augmenter la taille de son index (tout en privilégiant la qualité, bien sûr ;-)) pour arriver au "standard" actuel qui oscille entre 2 et 3 milliards de documents.

- Se calmer un peu au niveau de l'affichage de publicités sur sa page de résultats, un peu trop fournie en la matière à notre goût...

- Revenir à des temps d'affichage plus rapides, ce qui sera certainement le cas lorsque la période de lancement sera terminée.

A ce moment-là, on peut prédire qu'AltaVista se joindra à Fast comme concurrent principal de Google pour contester le quasi-monopole actuel du leader incontesté. Ceci dit, il nous a sincèrement semblé, au vu des tests effectués, qu'AltaVista était bien sur la voie du renouveau. Une bonne nouvelle pour les internautes qui ont appris la recherche d'info sur le Web grâce à ce moteur !