

Similitudes et différences entre les différents "moteurs YST" : AltaVista, AllTheWeb, Lycos et Yahoo!

[Retour au sommaire de la lettre](#)

La technologie YST de Yahoo! équipe aujourd'hui de nombreux moteurs de recherche. Mais tous affichent-ils la même information pour un même mot clé ? Cet article tente de faire ressortir les différences entre les différentes façons d'afficher les "résultats YST" par AltaVista, AllTheWeb, MSN et les autres. Et elles sont nombreuses !

Vous le savez certainement, les technologies des moteurs de recherche AltaVista et AllTheWeb ont aujourd'hui vécu et ont été remplacées, sur ces deux sites, par la technologie YST de Yahoo!, qui équipe également les sites de Yahoo! depuis plusieurs semaines. Il en est de même des clients "search" d'Inktomi, comme Lycos, MSN et HotBot. Tous ces portails proposent donc, aujourd'hui, des résultats de recherche issus de la technologie YST. Mais tous ces portails, pour un même mot clé, affichent-ils exactement le même résultat ? C'est ce que nous avons cherché à savoir en comparant les pages de résultats de ces différents outils...

Moteurs de recherche anglophones

Dans un premier temps, nous avons comparé les résultats des principaux outils utilisant YST dans leur version anglophone : Yahoo! (<http://search.yahoo.com/>), bien sûr, mais également AltaVista (<http://www.altavista.com/>), AllTheWeb (<http://www.alltheweb.com/>), MSN (<http://search.msn.com/>) et HotBot (<http://www.hotbot.com/>).

On peut noter que le site américain de Lycos, qui affichait il y a un temps les liens issus de la technologie AllTheWeb, propose maintenant les résultats de Looksmart et de son moteur Wisenut. Nous ne l'avons donc pas inclus dans ce comparatif. Sur HotBot.com, les résultats YST s'obtiennent au travers de l'option "résultats fournis par HotBot" sur la page d'accueil.

Premier comparatif : **tous ces outils de recherche affichent-ils de la même façon les résultats proposés ?** Réponse :

Affichage des résultats :

					
Titre	Si le site est dans l'annuaire : titre issu de l'annuaire Si le site est absent de l'annuaire : titre de la page (120 caractères max environ)	Titre de la page (120 caractères max environ)	Titre de la page (120 caractères max environ)	Si le site est dans l'annuaire de Looksmart : titre issu de l'annuaire Si le site est absent de l'annuaire : titre de la page (120 caractères max environ)	Titre de la page (70 caractères max environ)
Descriptif	Si le site est dans l'annuaire : descriptif issu de l'annuaire Si le site est absent de l'annuaire : contenu de la balise "meta description" (si disponible - 200 caractères max environ) ET "snippet" (extrait de la page contenant le mot demandé - 300 caractères max environ)	Contenu de la balise "meta description" (si disponible - 200 caractères max environ) ET "snippet" (extrait de la page contenant le mot demandé - 300 caractères max environ)	Contenu de la balise "meta description" (si disponible - 200 caractères max environ) ET "snippet" (extrait de la page contenant le mot demandé - 300 caractères max environ)	Si le site est dans l'annuaire de Looksmart : descriptif issu de l'annuaire Si le site est absent de l'annuaire : contenu de la balise "meta description" (si disponible - 200 caractères max environ) ET "snippet" (extrait de la page contenant le mot demandé - 300 caractères max environ)	Contenu de la balise "meta description" (si disponible - 200 caractères max environ) ET "snippet" (extrait de la page contenant le mot demandé - 200 caractères max environ)

				environ)	
URL	URL de la page (100 caractères max)	URL de la page (86 caractères max)	URL de la page (86 caractères max)	URL de la page (70 caractères max)	URL de la page (75 caractères max)
Taille de la page	OUI	NON	OUI	NON	OUI
Date de dernière mise à jour de la page	NON	NON	NON	NON	OUI
Traduction	OUI	OUI	NON	NON	NON
Version en cache	OUI	NON	NON	NON	NON
Accès aux autres pages du site (clustering)	OUI ("more pages from this site")	OUI ("more pages from... ")	OUI ("more hits from... ")	NON	NON
Catégorie de l'annuaire	OUI (si le site est présent dans l'annuaire)	NON	NON	NON	NON
Code bourse	OUI , à droite du titre (NYSE, Nasdaq si la société est cotée)	OUI, en début de page	NON	NON	NON
Possibilité d'ouvrir le lien dans une fenêtre différente	OUI	NON	NON	NON	NON
Nombre de résultats par page (par défaut)	20	10	10	15	10
"Related searches"	OUI (4 plus lien "more...")	OUI (10 max)	OUI (7 plus lien "more...")	NON	NON

Conclusion : s'ils utilisent tous l'index d'YST, les outils de recherche ne traitent pas de la même façon les résultats proposés :

- Titre : il peut être issu de la page web directement ou d'un annuaire comme ceux de Yahoo! ou de Looksmart.
 - Descriptif : de la même façon, il peut être extrait de la page ou d'un annuaire. Il peut être plus ou long, donc plus ou moins précis et descriptif.
 - URL : elle peut également être plus ou moins longue, donc plus ou moins explicite.
 - Taille de la page : seuls Yahoo!, AllTheWeb et HotBot l'affichent.
 - Date de dernière mise à jour de la page : seul HotBot propose cette information !
 - Traduction de la page : seuls Yahoo! et AltaVista proposent cette option.
 - Version en cache : uniquement disponible sur Yahoo!.
- Etc.





Selon les options désirées, il sera donc intéressant de choisir un outil de recherche ou un autre en fonction de vos besoins... Par exemple, on peut estimer que le descriptif de Yahoo!, issu de l'annuaire, est trop généraliste pour des recherche pointues. Ce résumé peut être intéressant pour des requêtes généralistes, mais beaucoup moins pour des mots clés très précis pour lesquels l'affichage de "snippets" peuvent apporter rapidement plus d'information quant au contenu du document recherché... La combinaison entre la balise Meta "description" et un snippet, comme l'affichent AltaVista et AllTheWeb, étant certainement la meilleure solution à nos yeux.

Deuxième comparatif : **les résultats fournis sont-ils identiques ?** Pour le savoir, nous avons saisi cinq mots clés de recherche sur chacun des outils et nous avons comparé ainsi les résultats :

- Combien de résultats sont identiques : ils sont présents dans la liste de résultats (10 premiers liens) et selon le même classement que sur Yahoo! ?
- Combien de résultats sont similaires : ils sont présents dans la liste de résultats (10 premiers liens) mais à un classement différent de celui proposé par Yahoo!.
- Combien de résultats sont différents : présents dans les 10 premiers liens de l'outil testé mais absents des 10 premiers liens de Yahoo!. Autre possibilité : résultat présent dans Yahoo!, mais pas dans l'outil testé.

Pour tous les outils, nous avons coché l'option la plus large proposée ("Worldwide" sur AltaVista, "Any language" sur AllTheWeb, etc.) afin d'obtenir des pages de résultats les plus proches et donc plus facilement comparables.

Comparatif des résultats :

Mot clé de recherche					
ibm thinkpad	Résultats identiques	8	4	6	5
	Résultats classés différemment	0	1	2	2
	Résultats ajoutés	2	5	2	3
dental insurance	Résultats identiques	10	4	4 (résultats identiques à ceux d'ATW)	4 (résultats identiques à ceux d'ATW)
	Résultats classés différemment		1	1	1
	Résultats différents		5	5	5
UK search engine	Résultats identiques	10	3	3	3 (résultats identiques à ceux de MSN)
	Résultats classés différemment		3	4	4
	Résultats différents		4	3	3
related searches	Résultats identiques	6	4	4 (résultats proches de ceux d'ATW)	4 (résultats proches de ceux de MSN)
	Résultats classés différemment	2	1	2	2
	Résultats différents	2	5	4	4
brad pitt	Résultats identiques	2	2	1	1 (résultats identiques à ceux de MSN)
	Résultats classés différemment	2	3	2	2
	Résultats différents	6	5	7	7

Le résultat est éloquent : s'il est évident que les résultats proviennent du même index de par la "zone de recouvrement" des liens proposés, il est également évident qu'ils ne sont pas identiques d'un outil de recherche à l'autre. Plusieurs hypothèses peuvent alors être évoquées :

- Utilisation de plusieurs index différents mais de contenus proches pour Yahoo! et ses "clients".
- Utilisation du même index mais avec des "réglages" différents (voir fin de l'article).

En tout état de cause, les référenceurs verront rapidement que, si un site web est bien positionné sur Yahoo! US, il n'en n'est pas obligatoirement de même sur AltaVista et ses "confrères"...

Enfin il convenait de se poser une troisième question, sur **la fraîcheur des résultats** : est-elle identique sur tous les outils de recherche ?

Pour cela, nous avons testé la date de dernière modification de la page d'accueil de grands sites américains d'actualité, fournie par les outils de recherche. Si HotBot nous l'affiche automatiquement, nous avons dû explorer le cache (pour Yahoo!) ou au travers de requêtes adéquates pour les autres. Exemple : *EDT "CNN.com"* pour le site de CNN ou *"washingtonpost.com" may* pour le site du Washington Post. Voici les résultats identifiés :

Fraîcheur des résultats :






					
CNN.com	9 mai 2004	9 mai 2004	9 mai 2004	9 mai 2004	9 mai 2004
News.com	8 mai 2004	8 mai 2004	8 mai 2004	8 mai 2004	9 mai 2004
Washingtonpost.com	9 mai 2004	9 mai 2004	9 mai 2004	9 mai 2004	9 mai 2004
Internetnews.com	7 mai 2004	7 mai 2004	7 mai 2004	7 mai 2004	8 mai 2004
news.yahoo.com	6 mai 2004	6 mai 2004	6 mai 2004	9 mai 2004	9 mai 2004

Globalement, on s'aperçoit que tous les outils fournissent la même information. Seules les données affichées par HotBot diffèrent, mais cela peut s'expliquer par le fait que la date fournie par ce moteur représente la date de dernière modification du document (date fournie par le serveur) alors que, pour les autres outils, nous avons pris en compte la date affichée sur la page web. D'où un léger décalage possible... En tout état de cause, il semble qu'il soit ici légitime de penser que, quelque soit l'outil de recherche utilisé, on aura accès à la même fraîcheur d'information...

Moteurs de recherche francophones

Dans le cadre du comparatif spécifique des outils de recherche francophones, nous avons enlevé de nos tableaux le moteur AllTheWeb, qui ne dispose pas de version strictement francophone, et ajouté Lycos France, qui proposait jusqu'à maintenant les résultats d'AllTheWeb. Sont ainsi proposés ci-dessous les comparatifs de Yahoo! France (<http://search.yahoo.fr/>), AltaVista (<http://www.altavista.fr/>), Lycos (<http://www.recherche.lycos.fr/>), MSN.FR (<http://search.msn.fr/>) et HotBot (<http://www.hotbot.fr/>).

Affichage des résultats :

					
Titre	Si le site est dans l'annuaire : titre issu de l'annuaire Si le site est absent de l'annuaire : titre de la page (120 caractères max)	Titre de la page (120 caractères max environ)	Titre de la page (78 caractères max environ)	Si le site est dans l'annuaire de MSN France : titre issu de l'annuaire Si le site est absent de l'annuaire : titre de la page (120 caractères max)	Titre de la page (120 caractères max environ)





	environ)			environ)	
Descriptif	Si le site est dans l'annuaire : descriptif issu de l'annuaire Si le site est absent de l'annuaire : contenu de la balise "meta description" (si disponible - 200 caractères max environ) ET "snippet" (extrait de la page contenant le mot demandé - 200 caractères max environ)	Contenu de la balise "meta description" (si disponible - 200 caractères max environ) ET "snippet" (extrait de la page contenant le mot demandé - 200 caractères max environ)	Contenu de la balise "meta description" (si disponible - 200 caractères max environ) ET "snippet" (extrait de la page contenant le mot demandé - 200 caractères max environ)	Si le site est dans l'annuaire de MSN France : descriptif issu de l'annuaire Si le site est absent de l'annuaire : contenu de la balise "meta description" (si disponible - 200 caractères max environ) ET "snippet" (extrait de la page contenant le mot demandé - 200 caractères max environ)	Contenu de la balise "meta description" (si disponible - 200 caractères max environ) ET "snippet" (extrait de la page contenant le mot demandé - 200 caractères max environ)
URL	URL de la page (100 caractères max)	URL de la page (86 caractères max)	URL de la page (30 caractères max)	URL de la page (70 caractères max)	URL de la page (40 caractères max)
Taille de la page	NON	NON	NON	NON	OUI
Date de dernière mise à jour de la page	NON	NON	NON	NON	OUI
Traduction	NON	NON	NON	NON	NON
Version en cache	OUI	NON	NON	NON	NON
Accès aux autres pages du site (clustering)	OUI ("Plus de résultats sur ce site")	OUI ("Plus de pages avec ...")	OUI ("Autres résultats de ce site... ")	NON	NON
Catégorie de l'annuaire	OUI (si le site est présent dans l'annuaire)	NON	NON	NON	NON
Code bourse	NON	NON	NON	NON	NON
Possibilité d'ouvrir le lien dans une fenêtre différente	OUI	NON	NON	NON	NON
Nombre de résultats par page (par défaut)	20	10	10	15	10
"Related searches"	OUI (1 + lien "suite")	OUI (10 max)	NON	NON	NON

On peut, ici, tirer les mêmes conclusions que pour l'espace anglophone : les résultats ne sont pas affichés de la même façon par les différents outils, chacun proposant plus ou moins de fonctions ou d'informations.

Chaque internaute choisira donc en fonction de ses besoins et de ses attentes l'outil qui lui convient le mieux.

Nous avons également effectué un comparatif des résultats fournis, sur des recherches "en français" pour chacun des outils :

Comparatif des résultats :

Mot clé de recherche					
apple ipod	Résultats identiques	3	3	1	1
	Résultats classés différemment	3	1	1	1
	Résultats différents	4	6	8	8
assurance vie	Résultats identiques	8	3	3	3 (résultats proches de ceux de MSN)
	Résultats classés différemment	0	4	4	3
	Résultats différents	2	3	3	4
moteur de recherche	Résultats identiques	9	3	3	2
	Résultats classés différemment	0	2	3	4
	Résultats différents	1	5	4	4
biodiversité	Résultats identiques	5	5	6	5
	Résultats classés différemment	2	1	0	1
	Résultats différents	3	4	4	4
gérard depardieu	Résultats identiques	6	2	3	3
	Résultats classés différemment	1	2	2	3
	Résultats différents	3	6	5	4

Même conclusion, peut-être encore plus marquée que pour les versions anglophones des outils de recherche :

- On trouve un "air de famille" aux différentes pages de résultat, indiquant clairement qu'ils sont tous servis par la technologie YST.

- Mais aucun outil ne propose les mêmes liens, dans le même ordre, pour un même mot clé. Parfois, les différences sont mêmes très importantes.

Il semblerait, selon les tests complémentaires que nous avons effectués, que le fait de mettre en place une recherche filtrée au niveau linguistique (ici la langue française) peut induire de fortes différences d'un outil à l'autre. Cela nous semblera logique en fin d'article, lorsque nous étudierons les possibilités dont dispose chaque moteur pour "customiser" ses résultats...

Enfin, comme précédemment, nous avons fait un comparatif des résultats selon la date de fraîcheur des liens obtenus sur cinq sites web francophones d'actualité :

Fraîcheur des résultats :

					
lemonde.fr	9 mai 2004	9 mai 2004	9 mai 2004	9 mai 2004	8 mai 2004
dna.fr	6 mai 2004	6 mai 2004	6 mai 2004	6 mai 2004	6 mai 2004
abondance.com	7 mai 2004	7 mai 2004	7 mai 2004	7 mai 2004	7 mai 2004
journaldunet.com	9 mai 2004	9 mai 2004	9 mai 2004	9 mai 2004	9 mai 2004
www.sporever.fr	1er mai 2004	1er mai 2004	1er mai 2004	1er mai 2004	1er mai 2004

Ici aussi, on s'aperçoit que tous les index semblent être "en ligne" au niveau de la fraîcheur des liens... L'hypothèse d'un même index peut donc être envisagée sérieusement...

D'où viennent les différences d'un moteur à l'autre ?

D'où peuvent bien venir les différences d'un moteur à l'autre ? Bien sûr, il est tout à fait possible que Yahoo!/Inktomi utilise plusieurs index pour mieux répartir son trafic. Chaque index étant mis à jour en parallèle mais pas obligatoirement en simultané, un moteur "attaquant" un index peut fournir des résultats proches mais différents d'un autre outil utilisant un autre index. C'est une première explication, tout à fait plausible (et vérifiée, voir ci-dessous).

Mais ce n'est pas la seule. En effet, chaque outil utilisant la technologie YST a à sa disposition plusieurs paramètres qu'il peut modifier à sa guise lorsqu'il effectue sa demande de résultats à Yahoo! (ces paramètres sont décrits dans un document fourni par Yahoo! et intitulé "Inktomi Data Protocol"). En d'autres termes, l'outil de recherche (par exemple AltaVista) reçoit la requête de l'internaute et effectue une demande à Yahoo! de résultats-liens correspondant à cette demande. Dans la configuration de cette demande, AltaVista peut "jouer" sur plusieurs paramètres :

- La langue : il est possible de ne demander des résultats que dans une langue donnée (anglais, français, allemand...) ou de demander un "mix" entre plusieurs langues. Par exemple, 40% de résultats en Français, 40% en Anglais et 20% en Allemand...
- La région : YST gère un argument régional. Il est possible de demander des résultats français ou allemands, etc. La région est déterminée par une analyse du document, de l'adresse IP du serveur et de l'interconnexion des pages du web. Il est également possible de "mixer" certaines régions.
- Index : YST gère également plusieurs bases de données (appelées "databases") définies selon la langue et la pays. Par exemple, la base de données www-fr correspond à la France, tandis que www-fr-ca correspond à la base de données en Français pour le Canada, etc.
- La syntaxe booléenne : le portail-client peut choisir de demander un ET ou un OU entre chaque mot demandé. En règle générale, c'est cependant un ET, pratique couramment admise aujourd'hui, qui est demandé. L'équivalent d'une phrase (comme si la requête avait été faite avec des guillemets) est également possible. Plusieurs possibilités de syntaxes assez complètes (et complexes) sont proposées par YST.
- Stemming : Le moteur peut demander ou non à YST qu'une approximation soit faite sur l'orthographe du mot demandé. Par exemple, que pour le mot clé *run* soient également trouvés les occurrences *running*, *ran* et *runs*. Cette option n'est pas proposée par défaut.
- Filtre familial : Il est possible de choisir ou non l'utilisation d'un filtre familial qui s'applique aux résultats.

On le voit, YST propose de nombreuses façons de configurer les résultats fournis aux moteurs de recherche à qui il délivre des résultats. Il y a fort à parier qu'AltaVista, MSN et les autres utilisent, à la base, les mêmes données que Yahoo! (le même index de départ), mais que les réglages de recherche sont différents dans chaque cas. D'autre part, chaque moteur peut, évidemment, et

comme le montrent les comparatifs proposés précédemment, afficher plus ou moins d'informations dans ses pages de résultats.

On peut d'ailleurs supposer que Yahoo! teste, sur ses moteurs AltaVista et AllTheWeb, d'autres réglages que ceux adoptés sur les sites Yahoo!, afin d'améliorer au fur et à mesure la pertinence de ses résultats. Bref, utiliser AltaVista et AllTheWeb comme des "laboratoires de recherches"...