

La syntaxe site: sur les moteurs de recherche majeurs

[Retour au sommaire de la lettre](#)

Sur Google, Yahoo! et MSN, la syntaxe site: permet de rechercher des informations sur un site particulier. Mais elle permet également des requêtes bien plus approfondies et / ou plus précises. D'autre part, est-elles prise en compte par Exalead, AskJeeves, Gigablast et Voilà ? Réponse...

Peut-être connaissez-vous déjà la syntaxe "site:" qui vous permet de restreindre vos recherches à un site donné sur les moteurs de recherche majeurs comme Google, Yahoo! et MSN. Mais en connaissez-vous toutes les subtilités ? C'est ce que nous allons essayer d'explorer dans cet article...

Back to basics...

Dans un premier temps, revenons sur l'utilisation "classique" de cet opérateur : il permet de restreindre une recherche à un site web. Par exemple :

google site:www.abondance.com

donnera comme résultat : "toutes les pages du site www.abondance.com qui contiennent le mot google". Cette syntaxe marche parfaitement sur Google, Yahoo! Search et MSN Search mais aussi sur Exalead, AskJeeves, Voila ou Gigablast. Bref, un vrai standard... Mais comment est-il possible de l'utiliser autrement ?

Combien de pages d'un site web sont indexées par les moteurs ?

Une autre utilisation possible est de taper uniquement la requête "**site:www.abondance.com**" qui vous donnera comme résultat "toutes les pages du site web indexées par le moteur de recherche" en question.

Là encore, cette syntaxe fonctionnera sur tous les moteurs... sauf Voila et AskJeeves, outils sur lesquels la syntaxe "site:" doit obligatoirement être accompagnée d'un mot clé. Dans ce cas, en tout cas, Voila semble donner parfois des résultats exacts, mais pas tout le temps, bref, il ne semble pas que cette syntaxe soit parfaitement gérée par le moteur de France Telecom... AskJeeves, lui, la refuse.

Recherche sur les sous-domaines d'un site

Il peut également être intéressant d'effectuer des recherches sur les sous-domaines d'un site ("quelquechose.votresite.com"). Par exemple, la requête "**google site:abondance.com**" aura pour vocation de rechercher "tous les documents appartenant à un sous-domaine du site abondance.com - par exemple www.abondance.com, actu.abondance.com ou outils.abondance.com - contenant le mot google".

Cette syntaxe fonctionne parfaitement sur Google, Yahoo!, MSN, Gigablast, AskJeeves, Exalead et Voila (avec quelques dysfonctionnements parfois observés sur Voila).

Recherche de toutes les pages de tous les sous-domaines

De la même façon que nous avons lancé la requête "site:www.abondance.com", la saisie de "**site:abondance.com**" devrait nous donner la liste de toutes les pages indexées par le moteur pour tous les sous-domaines du site en question. Et, en effet, tous les moteurs, hormis AskJeeves et Voila (avec un bémol sur ce dernier, voir la remarque ci-dessus) la prennent en compte.

Recherche sur les "TLD"

Mais la requête "site:" peut également vous aider à effectuer des recherches filtrées sur les "tld" (pour *Top Level Domain*), c'est-à-dire les domaines comme ".com", ".edu", ".fr" ou ".gov".

Par exemple, la requête "**chirac site:gov**" aura pour vocation de rechercher "toutes les pages des sites du Gouvernement américain - dont les sites se terminent en ".gov" - contenant le mot chirac".

Ce type de syntaxe marche parfaitement sur Google, Yahoo!, MSN, Exalead et AskJeeves mais ni sur Gigablast ni sur Voila. Ces requêtes sont notamment intéressantes pour des "tld" spécifiques comme .edu, .gov, .mil, .org, etc. Elles le sont moins pour des "tld" génériques et parfois, il faut bien le dire, un peu "poubelles" comme .com, .net et autres... Un site français pouvant être "immatriculé" en .com, une requête prenant en compte la syntaxe **site:fr** ne sera peut-être pas réellement efficace non plus... Bonne nouvelle cependant, des requêtes comme **ibm (site:ch OR site:fr)** fonctionnent bien sur certains moteurs sauf sur Google, AskJeeves, Gigablast et Voila...

De la même façon, une requête comme **google (site:www.abondance.com OR site:www.webrankinfo.com)** (je recherche "toutes les pages contenant le mot google et situées sur les site www.abondance.com OU sur le site www.webrankinfo.com") fonctionnera sur les mêmes moteurs et sera moins bien comprise (voire pas du tout) par les autres...

Recherche sur les terminaisons des noms de site

Autre type de recherche qui peut s'avérer intéressante : la recherche sur les terminaisons de noms de sites. Par exemple, imaginons que nous avons à faire une recherche inverse à celle ci-dessus : toutes les pages des sites gouvernementaux français qui parlent de George Bush. La requête sera alors "**bush site:gouv.fr**", les sites du Gouvernement français se terminant tous en ".gouv.fr". Là encore, tous les moteurs s'en sortent bien sauf Voila qui trouve, bizarrement, des documents issus des sites du Gouvernement... canadien !

Conclusion

Nos différents tests peuvent être synthétisés dans le tableau suivant, qui affiche la façon dont chacun des moteurs testés prend en compte les requêtes identifiées :

Syntaxe	Signification	Google	Yahoo!	MSN	Exalead	Voila	AskJeeves	Gigablast
google site:www.abondance.com	Toutes les pages du site www.abondance.com qui contiennent le mot google	OUI	OUI	OUI	OUI	OUI	OUI	OUI
site:www.abondance.com	Toutes les pages du site www.abondance.com indexées par le moteur	OUI	OUI	OUI	OUI	NON ?	NON	OUI
google site:abondance.com	Toutes les pages des sous-domaines du site abondance.com qui contiennent le mot Google	OUI	OUI	OUI	OUI	OUI ?	OUI	OUI
site:abondance.com	Toutes les pages des sous-domaines du site abondance.com indexées par le moteur	OUI	OUI	OUI	OUI	NON ?	OUI	OUI
chirac site:gov	Toutes les pages des sites ayant pour tld ".gov" contenant le mot chirac	OUI	OUI	OUI	OUI	NON	OUI	NON
bush site:gouv.fr	Toutes les pages des sites ayant	OUI	OUI	OUI	OUI	NON	OUI	OUI

	pour terminaison ".gouv.fr" et contenant le mot bush							
geneve (site:ch OR site:fr)	Toutes les pages contenant le mot geneve et sur des sites ayant pour tld ".fr" ou ".ch".	NON ?	OUI	OUI	OUI	NON	NON ?	NON

Conclusion : la plupart des moteurs "s'en sortent bien" pour ce qui est de la prise en compte de cet opérateur, sauf Voila dont les résultats sont le plus souvent difficiles à interpréter en raison d'"effets de bord" complexes à comprendre, voire déconcertants...

Plus précisément, les grands vainqueurs de ce comparatif sont Yahoo!, MSN et Exalead qui prennent en compte parfaitement toutes les variantes de la syntaxe "site:". Bravo.