

Requête « site: » : quelle utilité en SEO ?



Par **Daniel Roch**

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Le moteur de recherche Google, comme ses concurrents, permet aux internautes d'utiliser des commandes pour effectuer des recherches plus précises. Parmi celles-ci, la commande `site:` est souvent utilisée pour affiner une recherche ciblée sur un site précis. Comment s'en servir efficacement lorsque l'on fait du référencement naturel, et quels sont les défauts de cette commande ? C'est à ces questions que nous allons répondre dans cet article, en plus d'un petit rappel des syntaxes d'interrogations avancées du moteur...

Qu'est-ce que la commande `site:` ?

La commande `site:` a été initialement créée pour restreindre les résultats à un domaine précis. Par exemple, la requête suivante permet de rechercher tous les contenus qui parlent de référencement sur le site [abondance](http://www.abondance.com) :

[référencement site:www.abondance.com \[https://www.google.fr/search?q=site:abondance.com+r%C3%A9f%C3%A9rencement\]](https://www.google.fr/search?q=site:abondance.com+r%C3%A9f%C3%A9rencement)

En théorie, elle va donc permettre d'affiner une recherche, bien qu'elle soit souvent utilisée à tort sans mot clé ou expression pour mesurer l'indexation ou non d'un site. Nous reparlerons de ce point plus tard dans cet article.

La commande fonctionne ainsi :

- [\[site:monsite.fr\]](https://www.google.fr/search?q=site:monsite.fr) : liste les contenus indexés par Google pour le domaine monsite.fr et pour tous les sous-domaines qui vont avec (www.monsite.fr, blog.monsite.fr, etc.) ainsi que monsite.fr sans sous-domaine.
- [\[site:www.monsite.fr\]](https://www.google.fr/search?q=site:www.monsite.fr) : liste les contenus indexés par Google pour le domaine

www.monsite.fr, sans les autres sous-domaines.

- [\[site:sousdomaine.monsite.fr\]](https://www.google.fr/search?q=site:sousdomaine.monsite.fr) : liste les contenus indexés sur le sous-domaine sousdomaine.monsite.fr.

Par défaut, la commande exclue les contenus dupliqués détectés par Google, ne montrant ainsi que les contenus uniques aux yeux du moteur de recherche (source : <http://googlewebmastercentral.blogspot.fr/2007/03/using-site-command.html>).

A noter également : elle fonctionne également sur les moteurs de Recherche Yahoo et Bing !

Vérifier l'indexation avec la commande `site:`

Une commande peu fiable

Malheureusement, le problème de cette syntaxe d'interrogation est qu'elle ne montre pas la totalité des pages indexées d'un site. Elle n'affiche en règle générale que des résultats partiels.

Sur des sites de petite taille, cette commande va afficher presque tous les contenus indexés. En revanche, plus le site sera

grand, moins ce sera le cas et plus les résultats seront tronqués.

D'ailleurs, le nombre de résultats renvoyés par la commande site peut parfois varier entre la page 1 de Google, et les pages suivantes de la recherche. En faisant une commande site:monsite.com, la page 1 peut en effet afficher 5000 résultats, et la page suivante 5500. Cela est causé par le fait qu'en page 1 Google cherche à vous donner un résultat rapide des pages indexées d'un site, et qu'au fur et à mesure de l'avancée de votre recherche, ce résultat est affiné. Il est donc fortement déconseillé d'utiliser de manière brute le nombre de résultats d'une requête qui sera très souvent inexact. Ce n'est d'ailleurs pas pour rien que le terme « Environ » est ajouté juste avant le nombre de résultats pour la recherche en cours (source : <https://www.youtube.com/watch?v=6g8MJKFSWJU>).

Les autres solutions

Si vous cherchez à connaître l'indexation d'un site Internet qui vous appartient, deux fonctionnalités du centre pour webmasters de Google vous donneront de biens meil-

leurs résultats. Sur des sites qui ne vous appartiennent pas, en revanche, il vous sera quasi impossible de connaître la réelle indexation de l'ensemble des contenus.

Le fichier sitemap

La première fonctionnalité à utiliser est la soumission d'un fichier sitemap dans le Centre pour webmasters de Google (<http://www.google.fr/webmasters/>).

Une fois le fichier (le plus exhaustif possible, bien sûr) soumis, le centre indiquera le nombre de pages réellement dans l'index parmi celles soumises.

Attention, le chiffre donné, « Dans l'index » (voir illustration ci-dessous), n'est pas le nombre total de pages indexées, mais les pages indexées parmi celles du fichier sitemap.

Le menu indexation

Dans le centre pour webmasters de Google, il existe également un menu appelé « Indexation > Etat de l'indexation », qui affiche le nombre de pages actuellement dans l'index.

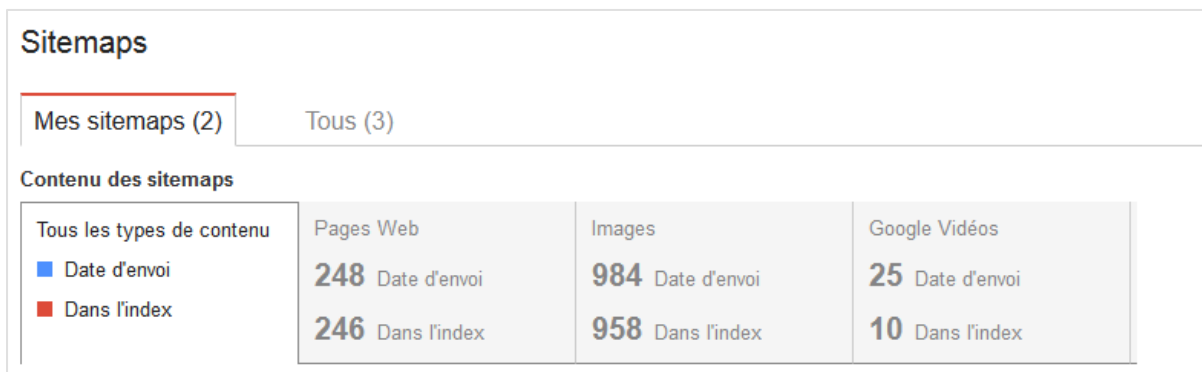


Fig.1. Zone d'analyse des Sitemap dans les Gogle Webmaster Tools

En cliquant sur le bouton « Avancé », il est même possible de connaître le nombre de pages bloquées par le fichier robots.txt ainsi que le nombre de pages supprimées de l'index par Google à votre initiative via ce centre pour webmasters. Malheureusement, impossible de télécharger une liste (au format texte ou csv) de ces pages indexées.

Comment la commande site: classe-t-elle les résultats ?

Il n'existe pas d'explication tangible sur l'ordre d'apparition des résultats sur les requêtes avec la commande site:, surtout quand cette dernière n'inclut aucun mot clé.

Selon certains référenceurs, les pages seraient classées par pertinence et popularité à la fois. Selon d'autres, le moteur de recherche afficherait de manière brute les pages d'un site. Autre hypothèse : le tri se ferait par PageRank descendant et taille des URL (de la plus courte à la plus longue : la page d'accueil, puis les répertoires, puis les pages finales dans l'arborescence). En tout état de cause, Google n'a jamais

officiellement communiqué sur le sujet, à notre connaissance.

D'après un test de Younès Merzouki datant de Décembre 2014 (<http://www.anthony-degrange.fr/test-seo-analyse-classement-commande-google-site/>), les pages ne seraient pas classées par popularité, mais selon d'autres critères, notamment le fait d'avoir une balise méta description correctement remplie ou encore des URL courtes, comme indiqué précédemment. Attention cependant, les résultats de ce test ne peuvent être pris comme règle de fonctionnement systématique.

Utiliser les autres commandes en complément

La principale utilité de la commande site: n'est pas de chercher à classer vos contenus ni de connaître l'indexation de votre site Internet, mais bel et bien d'affiner votre recherche.

Elle peut donc permettre entre autre :

- De voir si certains contenus d'une thématique sont indexés ;
- De trouver tous les contenus traitant ou

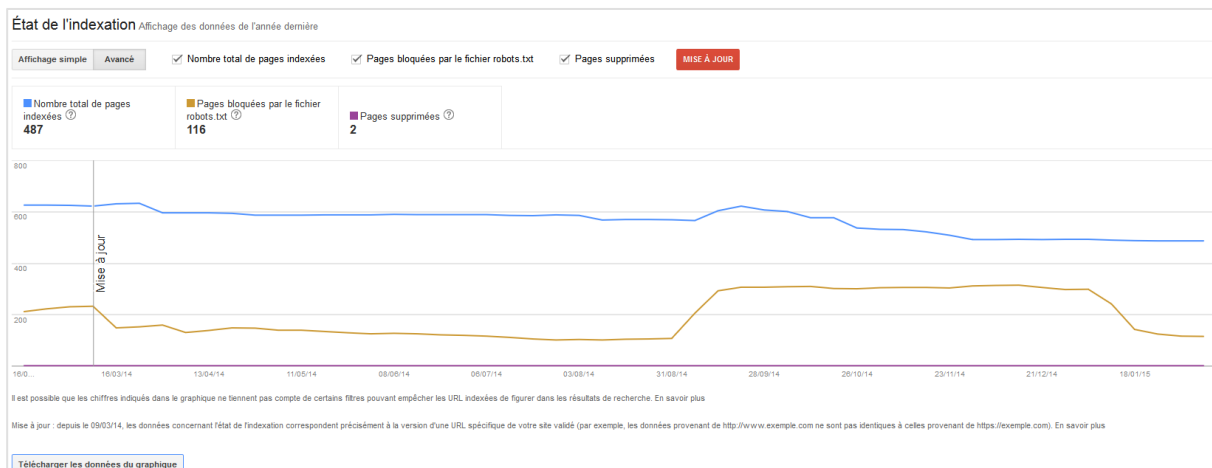


Fig.2. Nombre de pages indexées sur les 12 derniers mois, dans les Gogle Webmaster Tools

ne traitant pas d'une thématique (pour les modifier, pour y rajouter un lien, pour les supprimer...);

- De faire une veille concurrentielle ou sur un sujet bien précis.

Les autres commandes

On peut en effet utiliser d'autres commandes avec l'opérateur site:. Parmi celles qui peuvent vraiment vous être utile de manière générale, on peut utiliser les commandes suivantes :

- **OR** : pour une recherche sur un terme ou un autre (l'opérateur doit obligatoirement être en majuscules, pour ne pas le confondre par exemple avec le métal précieux);
- ***** : pour trouver un terme « joker ». Par exemple, [chocolat *](#) contiendrait tous les résultats avec chocolat et n'importe quel autre terme juste après ;
- **Guillemets** : "**X Y**" pour une recherche sur l'expression X Y ;
- **Deux points consécutifs (..)** : pour une recherche sur une plage de chiffres. Par exemple, [5..10](#) fait un recherche sur tous les contenus qui contiennent des chiffres compris entre 5 et 10 ;
- **-inurl:toto** : pour supprimer tous les résultats qui contiennent "toto" dans l'URL ;
- **inurl:toto** : pour cibler tous les résultats qui contiennent "toto" dans l'URL ;
- **-intitle:toto** : pour supprimer tous les résultats qui contiennent "toto" dans la balise Title ;
- **intitle:toto** : pour cibler tous les résultats qui contiennent "toto" dans la balise Title ;
- **filetype:pdf** : pour trouver les résultats ayant pour extension de fichier "pdf" ;
- **safesearch:toto** : pour trouver les résultats d'un terme sans les résultats adultes ;
- **-intext:toto** : pour supprimer tous les résultats qui contiennent le texte "toto" ;

- **intext:X** : pour cibler tous les résultats qui contiennent le texte "toto" ;
- **-inanchor:toto** : pour supprimer tous les résultats qui contiennent "toto" dans les ancres des liens vers cette page (et non pas dans cette page);
- **inanchor:toto** : pour cibler tous les résultats qui contiennent "toto" dans les ancres des liens vers cette page.

Notons cependant que cette liste n'est que théorique. A l'usage, on s'aperçoit que, parfois, Google n'en fait qu'à sa tête et renvoie des résultats qui ne correspondent pas forcément à 100% avec la demande effectuée.

L'utilisation avec site:

Comme expliqué auparavant, l'intérêt est donc de combiner ces commandes avec l'opérateur site: pour affiner vos recherches sur les résultats de votre site et ceux de vos concurrents, par exemple pour étudier des backlinks ou des thématiques précises sur un site.

Voici quelques exemples concrets des combinaisons possibles (sachant que, pour chacune de ces requêtes, le résultat ne sera pas obligatoirement exhaustif, comme indiqué auparavant) :

- **site:www.monsite.fr** : résultats indexés sur [www.monsite.fr](#) ;
- **site:www.monsite.fr wordpress** : résultats indexés sur [www.monsite.fr](#) et contenant le terme "wordpress" ;
- **site:www.monsite.fr WordPress OR Joomla** : résultats indexés sur [www.monsite.fr](#) et contenant les termes WordPress ou Joomla ;
- **site:www.monsite.fr wordpress *** : résultats indexés sur [www.monsite.fr](#) et contenant le terme "wordpress" ainsi que n'importe quel autre mot clé ;

- [site:www.monsite.fr](http://www.monsite.fr) "WordPress versus Joomla" : résultats indexés sur www.monsite.fr et contenant l'expression "WordPress versus Joomla" ;
- [site:www.monsite.fr](http://www.monsite.fr) 1..4 wordpress : résultats indexés sur www.monsite.fr et contenant le terme "wordpress" ainsi que des chiffres compris entre 1 et 4 ;
- [site:www.monsite.fr](http://www.monsite.fr) -inurl:wordpress : résultats indexés sur www.monsite.fr et ne contenant pas le mot "wordpress" dans l'URL ;
- [site:www.monsite.fr](http://www.monsite.fr) inurl:wordpress : résultats indexés sur www.monsite.fr et contenant le mot "wordpress" dans l'URL ;
- [site:www.monsite.fr](http://www.monsite.fr) -intitle:wordpress : résultats indexés sur www.monsite.fr n'ayant pas dans sa balise Title le terme "wordpress" ;
- [site:www.monsite.fr](http://www.monsite.fr) intitle:wordpress : résultats indexés sur www.monsite.fr ayant dans sa balise Title le terme "wordpress" ;
- [site:www.monsite.fr](http://www.monsite.fr) -intext:wordpress : résultats indexés sur www.monsite.fr n'ayant pas dans leur contenu le terme "wordpress" ;
- [site:www.monsite.fr](http://www.monsite.fr) intext:wordpress : résultats indexés sur www.monsite.fr ayant dans leur contenu le terme "wordpress" ;
- [site:www.monsite.fr](http://www.monsite.fr) -inanchor:wordpress : résultats indexés sur www.monsite.fr n'ayant pas dans les ancres de liens vers cette page le terme "wordpress" ;
- [site:www.monsite.fr](http://www.monsite.fr) inanchor:wordpress : résultats indexés sur www.monsite.fr ayant dans les ancres de liens entrants le terme "wordpress" ;
- [site:www.monsite.fr](http://www.monsite.fr) filetype:pdf : fichiers PDF indexés sur www.monsite.fr ;
- [site:www.monsite.fr](http://www.monsite.fr) safe-search:wordpress : résultats indexés sur www.monsite.fr contenant le terme "word-

press", mais sans les résultats pour adultes ;

Tout dépend en réalité de vos besoins et de l'élément à analyser, mais voici les utilisations possibles et réellement utiles de ces différentes syntaxes :

- Pouvoir lister ou exclure tous les contenus d'une thématique, que ce soit au niveau du contenu, de l'URL ou de la balise Title ;
- Pouvoir tester si certaines ancres de liens sont faites à tort vers certains contenus (avec la syntaxe [inanchor](#)) ;
- Trouver tous les fichiers d'un certain type et/ou avec une thématique précise ;

Conclusion

Comme on peut le constater dans ce guide, la commande `site:` est malheureusement bridée pour l'indexation. Elle perd ainsi son principal intérêt pour le référencier, puisqu'il est impossible de connaître avec précision la correcte indexation de ses contenus, ni celles des concurrents.

En revanche, combinée avec certaines requêtes ou termes supplémentaires, la commande `site:` devient très pratique pour faire des recherches poussées sur un site web. Si vous aimez affiner la compréhension de votre site dans Google, trouver certains types de contenus ou thématiques, ou, encore mieux, cerner le travail de vos concurrents, cette commande devient idéale !



Daniel Roch, Consultant WordPress, Référencement et Webmarketing chez SeoMix (<http://www.seomix.fr/>).