

Utiliser les données de Google Webmaster Tools pour son référencement

[Retour au sommaire de la lettre](#)

Le site Google Webmaster Tools, créé initialement autour du standard Sitemaps, propose aujourd'hui de nombreuses informations qui peuvent être très utiles pour le référencement d'un site web, d'autant plus que ces données émanent du moteur de recherche leader lui-même... Voici comment faire pour les utiliser au mieux !

Sorti en juin 2005 afin de permettre aux webmasters de soumettre leurs fichiers Sitemaps, Google Webmaster Tools s'est étoffé au fil du temps de différents rapports et de différentes statistiques sur la façon dont le moteur de recherche « voit » un site. Chacun de ces rapports peut être une source potentielle d'information sur les optimisations à opérer afin de positionner vos pages au mieux. Il suffit, la plupart du temps, de savoir comme interpréter correctement les données disponibles. Voici quelques pistes de travail...

Google Webmaster Tools est en effet aujourd'hui un outil unique dans le monde du référencement. C'est actuellement la seule plateforme permettant d'obtenir des informations sur l'indexation d'un site Web directement à la source, d'après les données internes d'un moteur de recherche. En attendant une mise à jour plus complète de Yahoo! Site Explorer avec des fonctionnalités similaires et le développement d'un portail Webmaster chez MSN/Live.com, cette vue de l'intérieur permet de comprendre au mieux les réactions du moteur face à un site Web et à ses mises à jour.

Le site est-il crawlé régulièrement ?

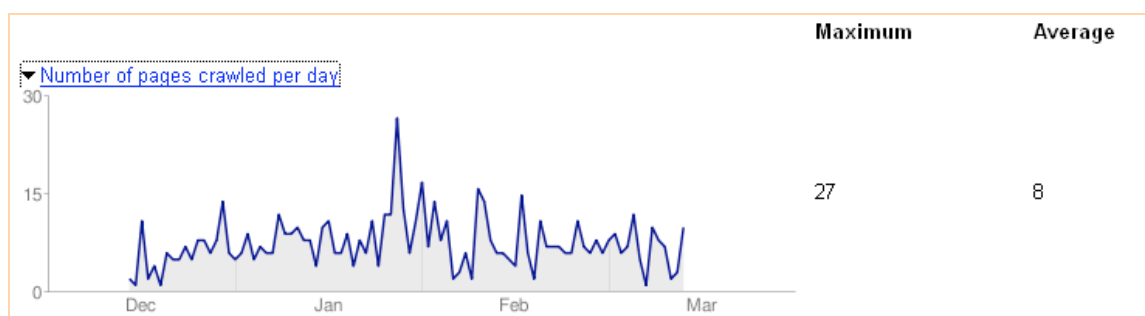
Si certaines données de Webmaster Tools sont trouvable avec d'autres outils d'analyse, celles concernant la fréquence de visites du moteur sur un site ne peuvent s'obtenir sans peine que par l'intermédiaire de Webmaster Tools. L'onglet "Diagnostic" contient la majorité de ces données. La page de résumé de cet onglet permet de savoir de quand date la dernière visite de Google Bot sur la page d'accueil de votre site :

Indexing Query stats »	
Home page crawl:	✓ Googlebot last successfully accessed your home page on Apr 2, 2007 .
Index status:	✓ Pages from your site are included in Google's index. See Index stats . [?]

Si cette date correspond bien à la dernière visite du robot d'indexation de Google, elle ne veut pas dire que la page d'accueil du site audité a été mise à jour dans l'index à cette même date. Les visites de Google Bot et les mises à jour de l'index de Google sont deux choses distinctes. On sait toutefois que plus les visites de Google Bot sont fréquentes, plus les chances de voir le contenu de vos pages mis à jour dans le moteur de recherche sont grandes.

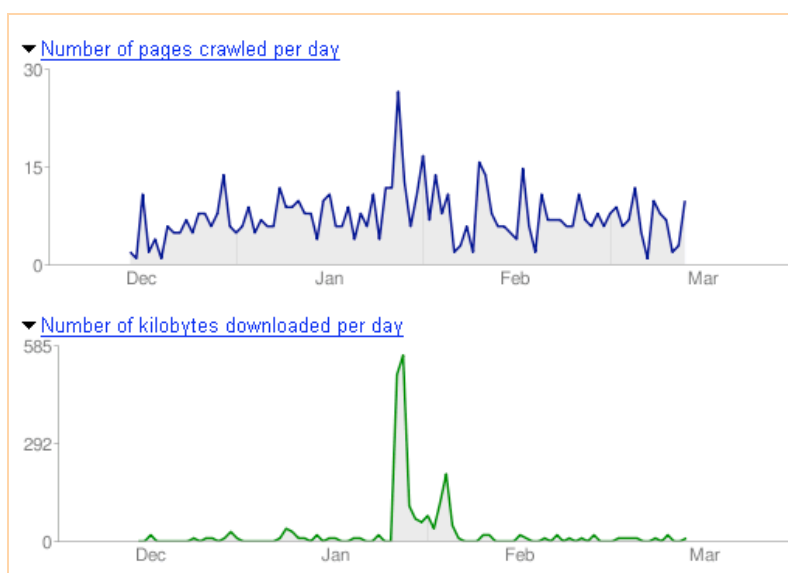
Pour favoriser les visites fréquentes de Google Bot, mettez à jour le contenu de votre page d'accueil de manière régulière. Google apprécie les nouveaux contenus, et la présence de nouveaux textes et liens l'incitera à revenir régulièrement voir si de nouveaux éléments à indexer ne sont pas présents sur votre page d'accueil. C'est bien pour cela que Google indexe aussi rapidement le contenu des grands sites d'actualité et des blogs.

Mais les visites à la page d'accueil d'un site ne font pas tout, il est important que la totalité des pages de votre site, ou du moins une grande partie d'entre-elles, soit visitée régulièrement. Pour s'en assurer, une visite dans la section "Crawl rate" de l'onglet Diagnostic s'impose. Cette page propose une série de graphiques détaillant les visites de Google Bot, la bande passante utilisée par celui-ci et son temps d'accès aux pages au cours de trois dernier mois.



Ces données sont à mettre en corrélation avec l'activité même de votre site. La règle énoncée plus haut, plus une page est mise à jour, plus elle a de chance d'être visitée par Google Bot, s'applique à la totalité des pages d'un site. Les pics de visites dans le graphique *Number of pages crawled per day* sont donc à rapprocher de l'activité réelle de votre site Web. Ceux-ci doivent correspondre, à quelques jours près, aux importantes mises à jour de votre site et donc aux moments où Google a identifié de nouveaux contenus sur celui-ci. De plus, ce graphique est l'une des rares données chiffrées fournie par Google. Profitez-en pour comparer le nombre de pages parcourues au maximum et en moyenne à l'inventaire réel de votre site, et au nombre de page réellement indexées par Google à l'aide de la commande "site:". Si ces données ne concordent pas, c'est peut-être que certaines sections de votre site Web ne sont pas accessibles au moteur de recherche.

Les deux autres graphiques sont également très utiles pour comprendre l'activité de Google sur un site. En comparant par exemple les données des visites de Google Bot et la bande passant utilisée par celui-ci, on peut avoir une idée des mises en cache massive du site, typiquement quand les pic de bande passante ne correspondent pas réellement aux pics de pages parcourues :



Google Webmaster Tools a d'abord été conçu pour recueillir les données du protocole SiteMaps, initié par Google. Dans cette partie aussi, les statistiques peuvent s'avérer très utiles. Les deux données critiques disponibles sous l'onglet "Sitemaps" de l'interface sont les suivantes :

View: All Sitemaps OK Sitemaps 1 to 1 of 1

Sitemap	Type	Submitted	Last Downloaded	Sitemap Status	URLs submitted
<input type="checkbox"/> sitemap.xml SITEMAP	Web	Mar 6, 2007	19 hours ago	OK	43

La date de dernier téléchargement du fichier Sitemaps (il y a 19 heures dans l'exemple ci-dessus) permet de savoir si celui-ci est parcouru de façon régulière. La fréquence de téléchargement dépend largement de la façon dont Google parcourt déjà le site, et de la date à laquelle un dernier *Ping* (une requête de mise à jour) a été envoyée à Google. Ici encore, plus on habitude Google à des mises à jour fréquentes (et réelles) d'un site, plus il visite régulièrement celui-ci afin de garder à jour le maximum d'informations. Il est donc important de veiller à ce que les délais de téléchargement d'un fichier Sitemaps correspondent à la réalité de ses mises à jour.

Deuxième donnée, le nombre d'URL soumises à Google par l'intermédiaire du fichier Sitemaps (43 dans cet exemple). Si le fichier XML est construit à la main, pas de soucis, le nombre inscrit dans cette section doit correspondre exactement au nombre de pages réellement disponibles sur le site. Si le fichier Sitemaps est en revanche créé de manière automatique, il peut être bon de s'assurer qu'il liste bien la majorité des pages disponibles, et est donc bien optimisé pour le référencement.

Repérer plus facilement les erreurs

Les rapports et chiffres détaillés ci-dessus permettent de s'assurer que Google Bot est capable de voir, de façon optimale, l'ensemble des pages qui composent votre site Web. Si jamais des problèmes dans la façon dont Google Bot parcourt votre site existaient, Google Webmaster Tools vous tiendrait également au courant de ceux-ci. La seconde partie de l'onglet "Diagnostic" de Google Webmaster Tools présente en effet l'ensemble des erreurs rencontrées par Google Bot lors de ses derniers accès au site (en général sur le dernier mois de visites).

Web crawl errors	
HTTP errors	✓ 0
Not found	⚠ 2
URLs not followed	✓ 0
URLs restricted by robots.txt	✓ 0
URLs timed out	✓ 0
Unreachable URLs	✓ 0
Total:	2

Ce diagnostic présente les erreurs rencontrées dans 6 domaines principaux :

HTTP errors : Ce rapport liste l'ensemble des pages renvoyant des erreurs HTTP à Google Bot. Il peut s'agir par exemple de problème d'identification (code d'erreur 401, lorsqu'une section de site est protégé par un mot de passe), d'accès totalement interdit (code 403) ou d'autre problème..

Not found : Les pages qui sont bien liées sur votre site mais renvoient un code d'erreur 404 lorsque Google Bot essaie d'y accéder. Ce rapport peut s'avérer des plus pratiques pour repérer les éventuels liens morts d'un site, même si un passage régulier d'un logiciel de vérification des liens (Xenu Link Sleuth par exemple) peut vous garantir une meilleure visibilité sur le maillage de votre site Web.

URLs not followed : L'ensemble des pages que Google Bot n'a pas été capable de suivre dans son parcours de votre site Web. Plusieurs raisons à cela : une redirection serveur de type 301 ou 302 mal configurée, une syntaxe d'URL étrangère à Google Bot (ce qui peut arriver avec des technologies de création de site et d'hébergement pointues)...

URLs restricted by robots.txt : Les pages auxquelles le robot n'a pu accéder en raison de la présence d'un fichier robots.txt à la racine du site Web. Si parmi les pages listées ici se trouvent certaines pages qui devraient être indexées et remonter dans les résultats des moteurs de recherche, il faut considérer la suppression ou la réécriture de votre fichier robots.txt afin d'éviter que ces pages ne soient pénalisées.

URLs timed out : Les pages qui ont renvoyé un code Time out lors de la dernière requête envoyée par Google Bot. Un time out signifie généralement que le serveur hébergeant votre site n'a pas été

capable de renvoyer une réponse assez rapidement au serveur de Google. Cela peut provenir d'une surcharge du serveur en question ou d'un encombrement passager du réseau Internet menant au site.

Unreachable URLs : L'ensemble des pages que Google Bot n'a pu atteindre sur le site, soit en raison d'un délai de réponse trop long soit en raison de l'indisponibilité totale du serveur d'hébergement du site Web cible.

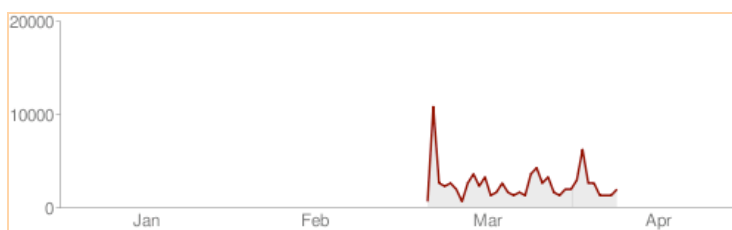
Pour l'ensemble de ces rapports d'erreur, Google liste non seulement les URL des pages concernées, mais également le code de l'erreur rencontrée et la date à laquelle cette erreur a été constatée, données accessibles depuis le liens Détails de la page d'accueil.

Errors 1 to 2 of about 2		
URL	Detail	Last Calculated
http://www.search-engine-feng-shui.com/del.icio.us	404 (Not found) [?]	Apr 1, 2007
http://www.search-engine-feng-shui.com/page/2/	404 (Not found) [?]	Mar 27, 2007
Download this table Errors 1 to 2 of about 2		

Vous pouvez ainsi facilement savoir si l'erreur en question est récurrente sur plusieurs pages, ou ponctuelle et due à une panne occasionnelle de votre service d'hébergement. Car les rapports fournis par Google ne servent pas uniquement au référencement de votre site Web, les visites régulières de Google Bot peuvent vous aider à diagnostiquer les éventuels problèmes liés à l'hébergement de votre site Web ou à la structure même de celui-ci. Cela peut notamment être le cas si votre site utilise un nombre important de sessions pour tracker les visiteurs tout le long de leurs visites.

Suivant la façon dont ces sessions sont définies, suivies et gérées par le moteur technique d'un site Web, Google Bot peut rencontrer un certain nombre de problèmes lors du crawling du site. Gardez à l'esprit que Google Bot ne peut exécuter certains scripts, qu'ils soient Javascript ou parfois serveur, qu'il ne peut stocker des valeurs en cookie. Il ne possède ainsi pas toutes les caractéristiques d'un visiteur classique. Si la disponibilité même d'une page de votre site dépend de la présence d'un cookie sur l'ordinateur de vos visiteurs, il y a de très fortes chances que Google Bot ne puisse y accéder et que vous retrouviez régulièrement cette page dans les rapports d'erreur détaillés plus haut. Google sera alors incapable d'indexer ces pages en question et votre référencement y perdra d'autant plus en qualité. Une analyse attentive des rapports d'erreur de Google Bot se révèle vite indispensable.

Veillez également au délai de réponse de votre site Web vis-à-vis de Google Bot. Pour cela, mettez en corrélation les données issues du rapport d'erreur "URLs Timed out" (si vous en trouvez) et le graphique "Time spent downloading a page (in milliseconds)" disponible dans la section "Crawl Rate" de Google Webmaster Tools.



Si le nombre d'erreurs est élevé et que le temps de téléchargement des pages de votre site est également assez élevé, vous pouvez demander à Google Bot de parcourir moins de pages de votre site à la fois à l'aide de l'option *Slower rate* disponible en bas du panneau "Crawl rate". Google Bot visitera moins de pages de votre site au cours de la même journée, mais celui-ci sera en contre partie plus accessible pour les internautes.

Le site est-il recherché et trouvé ?

Vous savez désormais que votre site Web est bien parcouru par Google Bot, de façon régulière, et vous avez pu identifier et corriger les éventuelles erreurs rencontrées par le robot d'indexation sur celui-ci. Reste à savoir si le site peut également être trouvé par ses visiteurs potentiels et s'il est bien visité. Pour l'ensemble de ces données, faire appel à un outil de statistique détaillé et précis (Google Analytics, Xiti...) pourra bien entendu vous révéler plus d'information. Google Webmaster Tools n'a jamais prétendu être une plateforme de *reporting* ou de statistiques précise et ses aides dans ce domaine ne peuvent être considérées que comme des indications de la visibilité de votre site dans le moteur.

Mais la lecture des informations fournies par la plateforme sur les requêtes effectuées peuvent s'avérer utiles dans la mesure où elles présentent à la fois les requêtes qui ont permis au visiteur d'accéder à votre site, et celles qui n'ont pas abouti à une visite, par ordre de popularité.

Rate this tool: 😊 😐 😞

Top search queries	Average top position
1. hotfrog	5
2. hotfrog google	3
3. feng "shui com"	9
4. hotfrog belgique	3
5. lsf network	4
6. commande link "live com"	3
7. vidéo de lsf pour le mots exposer	7
8. communiquera	10
9. msnlivesearch	12
10. arnaque casino770	18
11. seo concours casino770	19
12. concours casino770 mars 2007	42

[Download this table](#)

Ces deux données se trouvent dans la section "Statistics" de Google Webmaster Tools, dans la partie *Query stats*. Deux tableaux présentent ici les 20 premières requêtes ayant affiché votre site comme résultats (Top search queries) et les 20 premières requêtes ayant entraîné un clic vers votre site (Top search query clicks). Les chiffres disponibles en face de ces requêtes reflètent la position des pages de votre site dans les résultats de Google vis à vis de ce mot clé. N'oubliez pas, en consultant ce rapport, que vous pouvez filtrer l'ensemble de ces données à la fois par pays (France, Belgique...) et par type de moteur (Web, Image) afin d'obtenir une image encore plus précise de la popularité et de la visibilité de votre site sur Google.

Choose search type:

Choose search location:

Un enseignement à tirer de ces rapports ? Google Webmaster Tools est le seul outil à pouvoir actuellement fournir des informations sur les requêtes effectuées mais qui n'ont pas entraîné de visites. Ces informations sont très précieuses, puisqu'elles vous permettent de connaître les requêtes sur lesquels votre site est bien positionné et d'étudier le pourquoi de l'absence de visite : page de résultat non en adéquation avec la recherche, titre HTML peu incitatif, résumé automatique de la page peu incitatif également...

Autant de paramètre sur lesquels vous pouvez influencer afin d'acquérir plus de visiteurs. L'immense avantage c'est qu'ici vous n'avez plus à vous battre pour un positionnement, mais pour une meilleure visibilité de votre page. Si le travail n'est pas forcément plus facile, il repose sur des leviers plus facilement gérables et uniquement internes à votre site.

Un autre point sur lequel l'interface Google Webmaster Tools peut s'avérer très utile, c'est dans le cadre de la mise en place d'une stratégie de Net Linking intelligente. Google propose en effet

plusieurs rapports sur le nombre de liens renvoyant vers chacune des pages d'un site, sur les mots clés utilisés dans ces liens et sur le réseau interne des liens de votre site Web.

A l'heure où les liens constituent encore une grande partie des critères de référencement d'un site Web, ces rapports sont plus importants que jamais. **Nous étudierons ces données dans le prochain article de cette série.**

François Houste ([LSF Interactive](#))