

Voici le message que nous avons reçus fin janvier 2013.

La question: le volume de données indexées par Google chute

Nous nous sommes rencontrés lors de la formation Référencement naturel en décembre 2012. Nous étions une dizaine de participants, je suis traffic manager chez nnnnnn.com

J'espère que vous allez bien depuis cette session de formation.

Je me permets de vous contacter d'une part pour maintenir le contact, et d'autre part pour vous faire part d'une problématique référencement naturel que je rencontre actuellement et pour laquelle je suis dans l'impasse.

Depuis le 12 janvier, il semble que Googlebot n'arrive à lire que l'entête http des pages : je le vois quand je lance l'outil "Explorer comme Google" dans Webmaster Tools. En effet, quand je lance ce rapport d'exploration dans webmaster tools, le rapport ne me renvoie rien hormis les infos http 200 etc., alors que dans les rapports précédents tout mon html s'affichait. Par exemple, dans les rapports d'avant le 12 janvier février google renvoie content-length : 27998, alors qu'après le 12 janvier il me renvoie content-length : 20.

=> Vous trouverez ci-joint une capture d'écran qui illustre cela.



Par ailleurs, dans les statistiques d'exploration j'ai une chute brutale du nombre de ko téléchargés et donc du temps de chargement à partir du 12 janvier

=> Vous trouverez également en pièce jointe une capture d'écran qui l'illustre.

Problème d'exploration Google: chute des données indexées

Écrit par Raphaël Richard



Statistiques sur l'exploration Googlebot durant les 90 derniers jours

Symptômes:

- depuis le 12 janvier, le volume de données chargées par Google (dans Webmaster tools) chute
- les tests lancés depuis la page 'Explorer comme google qui renvoient l'entête HTTP de la page et le code HTML de la page) ne renvoient plus que l'entête HTTP de la page
- le robot.txt n'a pas été modifié, pas plus que le reste du site
- le site ne semble pas avoir été hacké (la présence de code malicieux pouvant stopper l'indexation par Google)

Notre réponse

J'aurais eu tendance à dire que cela aurait pu être lié à une attaque, avec injection dans vos pages de code compressé permettant d'installer des chevaux de Troie, par exemple, qui aurait provoqué un blocage par "l'antivirus" de Googlebot, mais je viens d'appeler vos pages et Bitdefender ne détecte rien. Je pense qu'il me manque un élément pour trouver la réponse à votre problème.

D'où venait le blocage de Google bot ?

Nous avons eu la réponse le lendemain dans notre boîte mail. Involontairement, nous avons mis notre client sur la solution du problème. Il s'agissait effectivement d'un problème lié à la sécurité.

Un grand merci pour votre réponse !

Nous venons de trouver la cause de notre problème !

Notre système de sécurité côté serveur était trop dur, une IP Google a été bloquée par erreur. Une fois l'IP libérée, Webmaster Tools peut lire à nouveau le HTML de mes pages et je vois dans Google que le cache se met à jour. J'ai également cliqué sur le bouton "indexer" dans Webmaster Tools, pour ma homepage, par sécurité.

Commentaire sur la résolution du problème

Un élément manquait effectivement dans le diagnostic. Bien que le site lui même n'ait pas été modifié, avant que les problèmes de se produisent, le système anti hacking du serveur web sur lequel le site était hébergé avait été modifié. C'est ce qui a provoqué le blocage de Googlebot.

Ce client utilise probablement un système permettant de bloquer les machines de hackers dès qu'elles multiplient les connexions, ce qu'un internaute ordinaire ne fait pas.

Comme Google appelle un grand nombre de pages, il sollicite le service HTTP plus qu'un utilisateur normal et a du être assimilé a un "hacker" et, conséquence de quoi, a été bloqué.