

Le déréférencement de Google

La visibilité d'un site Web est largement assurée par sa présence dans l'index des principaux moteurs de recherche. Certaines sociétés ont pu voir leur site retiré ou « déréférencé » sans que toutes les raisons ne soient réellement connues. Google est le moteur de recherche le plus utilisé avec 65 % de parts de marchés selon l'institut Nielsen. Dans la course à la taille des index les trois géants (Google, Yahoo! et MSN) ont arrêté de compter officiellement le nombre de pages référencés depuis 2005. Un bon référencement est devenu un impératif pour toute société qui veut être présente sur le Web. La visibilité sur Internet d'une société dépend en bonne partie de son bon référencement. Le déréférencement volontaire ou non d'un site web peut être lourd de conséquences pour une société dont le site web est un outil commercial.

Le référencement

Avant d'être déréférencé, il est nécessaire d'être référencé par l'index de Google. Les Googlebot (les robots de référencement), parcourent en permanence le Web pour l'indexer et augmenter le nombre de pages connus. Il existe une autre méthode pour être référencé, « l'ancienne méthode », entrer soi même l'adresse du site internet à intégrer dans la base de Google. Cette méthode si elle a le mérite d'être efficace n'est pas la plus rapide puisque qu'il peut se passer plusieurs semaines avant que le référencement soit effectif. De plus un référencement dans l'index ne vous assure pas forcément de figurer dans les premières pages. Une fois dans l'index c'est là que les manipulations peuvent arriver et qu'un site peut se retrouver volontairement ou non déréférencé.

Le déréférencement

Google n'a pas pour habitude de donner de nombreuses explications sur les raisons d'un déréférencement. Pour preuve cette déclaration de Matt Cutts, « software engineer » et porte-parole technique de Google qui date de la fin de l'année 2004 qui illustre bien le point de vue du moteur de recherche. « Traditionnellement, Google n'informe jamais les sites qui sont identifiés comme spammeurs ». En effet certains sites ont eu la surprise de se voir déréférencé du jour au lendemain.

Les sites internet des sociétés BMW (www.bmw.de) et Ricoh (www.ricoh.de) ont fait l'objet de mesure de suppression temporaire de l'index par Google en août 2006 pour utilisation trop importante des pages satellites (pages dont le contenu visible pour l'internaute est identique). En effet les sites qui ne respectent les règles établies par Google risque de se voir non référencé par le moteur de recherche. Google n'aime pas qu'on l'attaque et n'hésite pas à punir les sociétés l'ayant attaqué. Dans l'affaire contre les journaux belges francophones Google a supprimé les journaux en question de l'index de Google actualités comme imposé par la justice belge, mais a également supprimé les sites webs des journaux comme Le Soir, La Libre Belgique ou la Dernière Heure de l'index général. On voit dans cette décision que le pouvoir de nuisance de Google peut être important car au delà du coup médiatique, cet événement pose le problème du libre accès au site Web.

Il n'y a pas que Google lui-même qui a le pouvoir de supprimer des pages Web, les internautes bien ou mal intentionnés peuvent faire disparaître des pages web de l'index de Google. Un tiers avec de bonnes connaissances en informatique peut tout à fait réussir à déréférencer une page ou un site.

La suppression d'un site de l'index de Google nécessite des informations personnelles comme un compte google (une adresse gmail ou un compte google personnalisé). Toutefois en utilisant l'outil de suppression des liens morts il faut simplement posséder une adresse mail pour pouvoir accéder à cet écran. La suppression proprement dite est difficile à réaliser car elle demande de pouvoir accéder au contenu même du site pour y insérer un fichier « robots.txt » bien calibré qui indique au googlebot de ne pas parcourir le site en question et de cette manière le faire sortir de l'index.

Marc Ajalbert
Veilleur
IFIE

Termes associés :

Cloaking : procédé technique qui consiste à berner le Googlebot (le robot qui indexe continuellement les pages) en donnant des informations qui sont différentes de celle que voient les internautes.

Google Bombing : technique de référencement visant à influencer le classement d'une page dans les résultats de Google. Cette technique exploite une caractéristique du moteur, le PageRank, qui accorde un poids plus important au texte avec un lien hypertexte vers une autre page.

Spamdexing : Mot composé à partir de spam et de indexing, que l'on peut traduire par référencement abusif. Il s'agit de l'ensemble de méthode visant à inclure des données susceptible de tromper les algorithmes et les robots des moteurs de recherches pour améliorer son référencement.