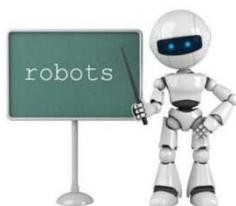


FICHER ROBOTS.TXT : MA DÉFINITION, SON USAGE ET SA CONCEPTION.

🕒 Vous n'avez pas 6 min ...

Tout le monde a une raison d'aimer le web !



Découvrir ce que les autres ont envie de dire. Délivrer un message ou faire grandir ses connaissances.

Chaque utilisateur a des besoins différents comblés par internet. Mais, comment serait internet sans les moteurs de recherche ? Ne vous sentiriez-vous pas comme celui qui se retrouve au milieu du désert sans connaître le chemin ?

À l'origine, les moteurs de recherche pouvaient se comparer à de simples panneaux indicateurs. Ils vous indiquaient les différents sites qui reprennent les mots de vos requêtes.

De nos jours, nous pourrions plutôt les comparer à des nomades qui se trouvent sur votre chemin au milieu du désert. Certes, ils recueillent vos questions. Mais surtout, ils tendent à se comporter comme des êtres humains en essayant de comprendre votre besoin pour vous indiquer les meilleures choses à faire pour répondre à votre besoin.

Cette évolution oblige les moteurs de recherche à devenir « omniscient ».

Le web devenant de plus en plus important, les spiders ont de plus en plus de travail pour crawler les sites web. Les



moteurs de recherche comme Google ou Bing se sont très vite rendus compte qu'explorer un site mal construit était très long à réaliser.

Ils nous ont imposé leur volonté de structurer les sites qu'ils allaient explorer.

Ils ont aussi compris qu'il n'était pas indispensable d'explorer toutes les pages d'un site. L'exploration prend beaucoup de temps. Alors, pourquoi perdre du temps sur des pages qui n'auraient pas d'importance pour l'utilisateur final ?

C'est à ce moment précis qu'ils ont demandé de l'aide, à nous les concepteurs de site.

1. Le fichier ROBOTS.txt, une liaison avec les moteurs de recherche



En mettant en place un simple fichier de type texte à la racine de notre site internet, ils nous offrent la possibilité de leur signaler les contenus ou répertoires qu'ils n'ont pas besoin d'explorer.

Ce fichier devient donc un recueil de contenus qui ne leur seront d'aucune utilité pour répondre aux besoins des internautes.

Il est évident que même s'ils nous demandent de l'aide, ils ne sont pas obligés de nous croire. Il est tout à fait possible qu'ils passent outre nos recommandations. Google ou un autre peut tout à fait crawler ces fichiers.

Mais il le fera seulement si cela lui confère un certain intérêt.

2. Le fichier Robots.txt pourrait-il servir à désindexer ?

Le fichier ROBOTS.txt ne permet pas de désindexer une page. Dites-vous bien que l'on n'oblige pas Google à réaliser quelque chose.

Par contre, il est possible que Google n'indexe pas votre nouvelle page et cela provient du simple fait qu'il ne l'a encore jamais crawlé ; Et que vous lui dites grâce à votre fichier Robots.txt qu'elle ne vaut pas la peine.

Vous avez vu la subtilité ? Voilà d'où est né le mythe de la désindexation par Robots.Txt.

3. Concevoir votre fichier Robots.txt



Le fichier Robots.txt est un fichier qui regroupe des instructions compréhensibles par les moteurs de recherches.

Voici la liste des instructions que vous allez pouvoir utiliser pour le créer :

User-agent : vous indiquez aux robots qu'ils doivent prendre en compte les règles à suivre. Il est important de bien connaître leur nom pour que la demande soit prise en compte. Pour exemple, chez Google ils se nomment GoogleBot, GoogleBot-News, GoogleBot-Image. Tandis que pour Bing, il se nomme BingBot. Il existe une multitude de robots. Prenez le temps de vous renseigner sur ceux qui pourront potentiellement vous intéresser. Cependant vous devez savoir que pour créer une demande qui s'adresse à tous les robots, il suffit de placer le paramètre * comme ceci : user-agent : *

Disallow : vous suggérez d'éviter de crawler la page ou le répertoire. Il n'apportera rien à l'utilisateur final.

Allow : Vous indiquez aux spiders que votre page ou votre répertoire est susceptible d'apporter une réponse à un utilisateur final.

Sitemap : Avec cette instruction, vous indiquez aux spiders que vous avez un fichier sitemap.xml. Les moteurs de recherche comme Google, Bing, Quant, Ask ont opté pour cette instruction. Avec cette instruction, il n'est plus nécessaire de passer par des extensions pour indiquer votre sitemap.

4. Comment le fichier Robots.txt peut-il alléger votre serveur ?



En effet, il se peut que le crawl simultané de votre site par divers spiders puisse engendrer des pertes de vitesse de chargement pour vos

utilisateurs finaux.

La solution de ralentir les crawls de votre site aura pour effet d'alléger la charge de votre serveur.

Il existe différents moyens de réguler la fréquence maximale de crawl par moteurs de recherche.

La mise en place de crawl-delay dans le fichier Robots.txt

Crawl-delay : Vous pouvez indiquer à certains moteurs de recherche de respecter un délai entre deux crawls de vos répertoires ou fichiers.

Cette instruction n'est prise en compte que par Bing, MSN et Ask.

Google ne prend pas en compte l'instruction crawl-delay.

Cependant, si vous avez besoin d'alléger votre serveur, Google vous offre la possibilité d'agir sur la fréquence de crawl des spiders.

Pour cela, vous devez vous rendre à l'adresse :

<https://www.google.com/webmasters/tools/settings>

Vous sélectionnez ensuite la propriété du site à impacter.

Il ne vous reste plus qu'à paramétrer la vitesse d'exploration maximale. OÙ laisser GoogleBot se charger de cela.

5. Le fichier Robots.txt, un exemple créé par WordPress

Voici l'**exemple du fichier Robots.txt** de qui est généré par WordPress au départ :

User-agent: *

Disallow: /wp-admin/

Allow: /wp-admin/admin-ajax.php

6. Conclusion

Le fichier robots.txt vous sera très utile pour indiquer aux spiders vos pages à référencer.

Ne vous en servez pas comme fichier de désindexation, vous seriez déçus !