

Comment trouver les pages orphelines avec Screaming Frog ?

Quand il y a beaucoup de modifications sur un site web, il est fréquent d'oublier de faire une redirection, de changer une url ou de continuer à pointer sur une page qui retourne une 404.

Screaming Frog peut nous aider à les retrouver en ajoutant d'autres sources de données que le crawl lui-même.

Voici la marche à suivre en résumé

1. url du site dans le champ principal
2. Exclure / Inclure des portions du site
3. Si crawl multidomaines, se servir de l'option cdn (voir l'article associé pour le paramétrage)
4. Utiliser Google Analytics (configuration > api access > Google Analytics). Setter le date range et cocher la case "create new urls discovered in Google Analytics".
Dans user account, sélectionner le segment "all users" pour utiliser l'ensemble des données GA et pas seulement les données relatives aux sources organiques
5. Utiliser Google Search Console (configuration > api access > Google Search Console).
Setter le date range et cocher la case "create new urls discovered in Google Search Console"
6. Crawler le sitemap XML (configuration > spider > crawl > XML sitemap. Sélectionner "Crawl linked XML sitemap" et renseigner la /les méthodes pour trouver les sitemaps
7. Configurer l'analyse de crawl (crawl analysis > configure > check sitemaps, Analytics et Search Console + auto analyse at the end of the crawl)

et le tutoriel complet donné par Screaming Frog

<https://www.screamingfrog.co.uk/find-orphan-pages/>

Révision #2

Créé 12 October 2020 12:18:28

Mis à jour 27 May 2022 13:20:42 par Jean-Paul Diaz-Caneja