

Structures optimisées pour sites (extra) larges



Par Cédric Cherchi

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Le principal défi des sites à forte volumétrie est de proposer une visibilité équitable à l'ensemble de leurs contenus qui sont clairement inutiles tant qu'ils ne sont pas vus des utilisateurs... et de Google. La plupart des plus grands sites ont trouvé une astuce imparable pour offrir une exposition optimale à leurs contenus : une structure soigneusement optimisée au travers d'un plan de site pour les internautes bien conçu.

Catégorisation des contenus vs structure de site

La structure d'un site reflète généralement la façon dont on catégorise ses pages. Cette catégorisation peut avoir des conséquences dramatiques sur sa/son (in)capacité à promouvoir les contenus auprès des utilisateurs. Cette notion est d'autant plus délicate à appréhender que nous avons tendance à confondre catégorisation et structure. Or, si la première influence la seconde, cette dernière représente notamment la traduction des relations (liens hypertextes) entre les différentes pages du site.

Quelle profondeur pour votre site ?

De la structure du site peut-être déterminé un élément essentiel en termes d'expérience utilisateur : le nombre de clics nécessaires pour accéder à une page (depuis la page d'accueil). Nous parlons ici de la profondeur de la page. Ce facteur fait directement référence à la règle bien connue des 3 clics [1, voir références en fin d'article] :

Users of a site will become frustrated and often leave if they cannot find the information within the three clicks.

Bien évidemment, se borner à maintenir toutes les pages d'un site à 3 clics de la page d'accueil n'a pas de sens. D'ailleurs, cette règle est régulièrement discréditée par des essais utilisateurs qui, par ailleurs, sont de plus en plus souvent encouragés à utiliser les formulaires de recherches internes aux sites. Toutefois, elle rappelle que le nombre de clics aura toujours, à un moment donné, un impact sur l'efficacité d'une page. Et puisque les crawlers, comme GoogleBot, ne peuvent s'appuyer sur les moteurs de recherches internes pour explorer les sites de façon exhaustive, le nombre de clics

demeure un facteur-clé pour le SEO. Sachant cela, combien de clics (ou pages intermédiaires) ne doit-on pas dépasser pour garantir une visibilité optimale aux contenus ? Au cours d'un Google Hangout organisé le 24 octobre 2014, John Mueller a donné de très timides éléments de réponse [2] :

Home pages is where we forward the pagerank within your website. Depending on how your website is structured, if content is closer to the Home page, then we'll probably crawl it a lot faster, because we think it's more relevant. But it's not something where I'd say you artificially need to move everything three clicks from your Home page.

A défaut d'une recommandation précise, John confirme ici que le nombre de clics (ou la profondeur) a un impact sur le crawl de GoogleBot et donc sur sa propension à indexer et positionner une page dans les SERP de Google.

In fine, le référencement d'une page peut être dégradé par la catégorisation du site (rubrique/sous-rubrique/tags etc.) mais également par la pagination potentiellement utilisée au sein des rubriques et SERP internes. Si les pages les plus profondes d'un site d'actualité sont *a priori* les moins critiques (puisque plus anciennes et moins concurrentielles) il en va autrement d'un site de e-commerce ou d'un site de *social networking* dont toutes les pages ont, *a priori*, une importance égale.

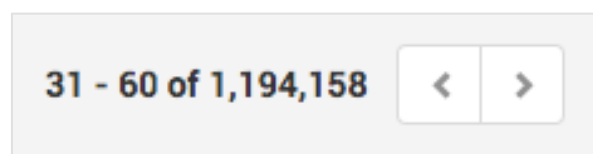


Fig.1. Les boutons "précédent/suivant", le spider-trap qui peut augmenter considérablement la profondeur d'un site

Jusqu'où irait Google pour une page ?

Vérifions cette hypothèse. Une analyse des logs HTTP de plusieurs sites de e-commerce à forte volumétrie nous donne une idée plus précise de la relation entre profondeur et fréquence de crawl. Googlebot peine à crawler les pages situées à plus de 3 clics de la page d'accueil. Au delà de 5 clics, les pages ne sont crawlées qu'une fois sur la période (quand toutefois elles sont crawlées – voir fig.2).

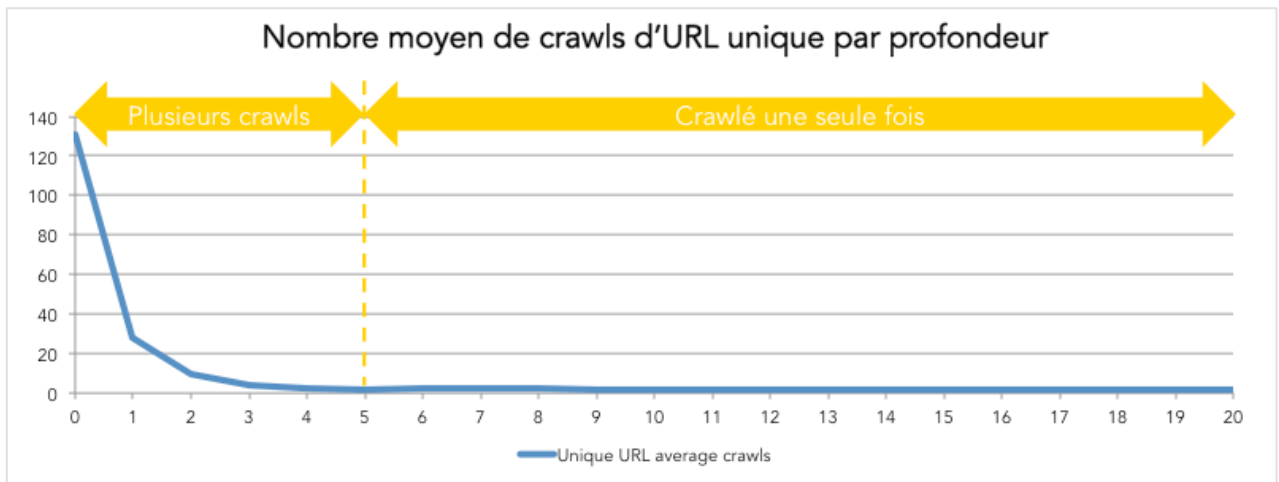


Fig.2. Nombre moyen de crawls d'URL unique par profondeur

Une analyse de l'efficacité des pages (pages crawlées générant au moins 1 visite SEO) par niveau de profondeur confirme qu'une page plus profonde est moins encline à générer des visites (fig.3.).

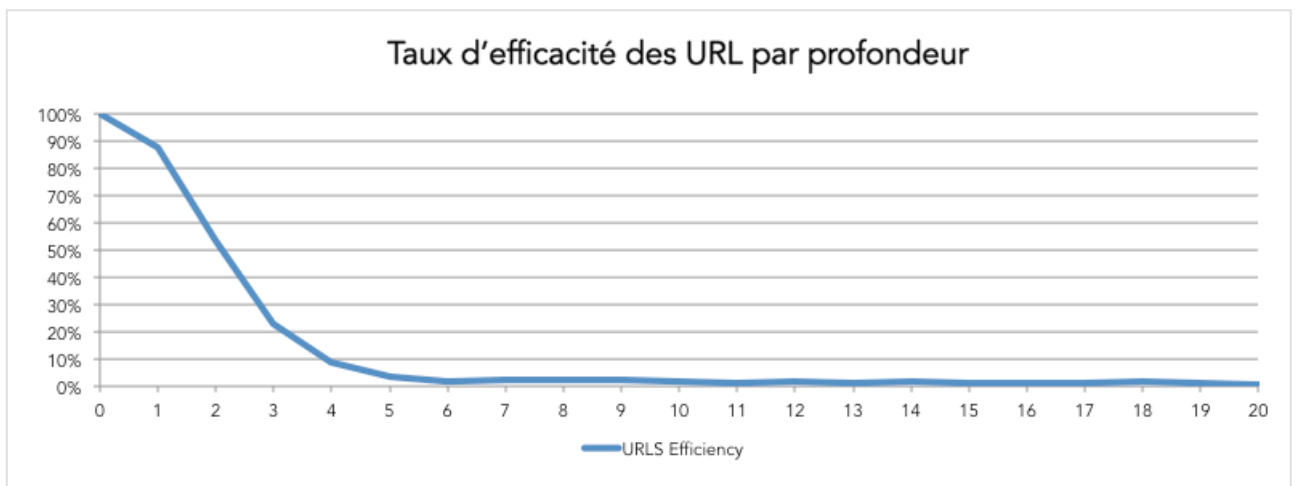


Fig.3. Taux d'efficacité des URL par profondeur.

Evidemment, ces résultats peuvent varier d'un site à l'autre. D'ailleurs ils seront probablement différents sur un site d'actualité par rapport à un site e-commerce !

Un site map pour les relier tous!

Pour atténuer ce phénomène, les responsables SEO de la plupart des grands sites appliquent intelligemment une recommandation pourtant simpliste du Google Search Console Help [3] :

Offer a (HTML) site map to your users with links that point to the important parts of your site. If the site map has an extremely large number of links, you may want to break the site map into multiple pages. Keep the links on a given page to a reasonable number.

On parle bien ici de plan du site pour les internautes et non pas de fichiers Sitemap XML... Même si le nombre de lien n'est *a priori* plus un problème depuis 2008 [4], tout l'exercice repose dans la conjugaison de deux contraintes antagonistes pour générer une navigation alternative :

- Limiter au maximum le nombre de clics entre la page d'accueil pour diminuer la profondeur moyenne des pages ;
- Limiter au maximum le nombre de liens par page pour préserver au maximum le jus (*linkjuice*) diffusé sur l'ensemble des pages.

Prenons l'exemple d'un site contenant 10 millions de pages et imaginons que nous nous donnions comme objectif de placer 100 liens maximum sur chacune des pages de notre sitemap HTML.

La méthode la plus simple pour générer les pages de ce site map est de réunir les pages du site (indépendamment de leur catégorisation) par groupe de 100 et de réunir chaque groupe par groupe de 100, etc.

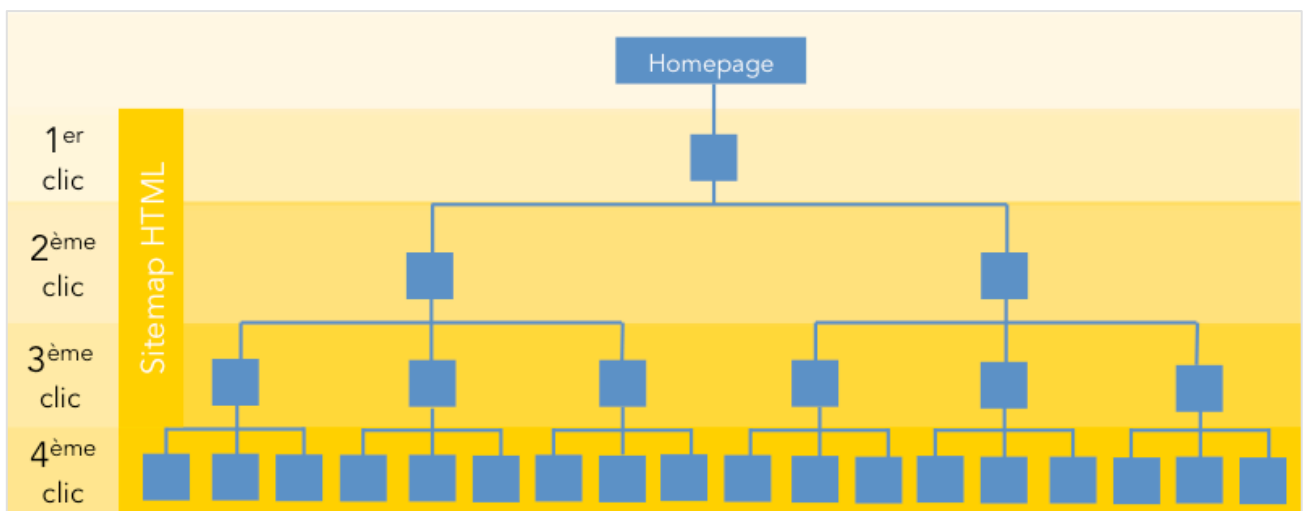


Fig.4. Schéma d'un sitemap HTML optimisé sur 4 niveaux

Cette méthode nous donne la répartition de liens suivante :

- Page d'accueil, niveau 0 (1 lien vers le sitemap)
- Sitemap, niveau 1 (10 liens/page) : 10 liens
- Sitemap, niveau 2 (100 liens/page) : 1,000 liens
- Sitemap, niveau 3 (100 liens/page) : 100,000 liens
- Sitemap, niveau 3 (100 liens/page) : 10,000,000 liens

Il est donc tout à fait possible de rendre accessible toute page d'un site de 10 millions de pages en 4 clics et en ne dépassant pas 100 liens par page. Mieux encore, passer le nombre de liens de niveau 1 de 10 à 100 permettrait de relier 100 millions de pages avec le même nombre de clics. Un site à plus faible volumétrie peut nécessiter moins de liens par pages et, surtout, moins de clics.

The screenshot shows the Facebook search interface. At the top, there's a navigation bar with the Facebook logo, an 'Inscription' button, and login fields for 'Adresse e-mail ou mobile' and 'Mot de passe'. Below this is a search bar with the text 'Rechercher des personnes' and a 'Rechercher' button. The main content area is titled 'Parcourir par nom' and includes a search filter for 'Personnes'. Below the filter, there are two rows of navigation tabs: 'Latin Character Names' (A-Z) and 'Non-Latin Character Names' (0-25). The search results are displayed in a grid format under the heading 'Résultats de type Personnes pour Cedric Chenot - Cedric Cherrier'. The results list various names starting with 'Cedric', such as 'Cedric Chenot', 'Cedric Cher', 'Cedric Cherel', etc. At the bottom, there are language selection options and a footer with 'Facebook © 2016'.

Fig.5. Une page du sitemap HTML du site facebook.com

Twitter (300 millions d'utilisateurs) [5], LinkedIn (400 millions d'utilisateurs) [6] et Facebook (1,5 milliards d'utilisateurs) [7] sont autant d'exemples d'utilisation de ce type de sitemap HTML. D'ailleurs, seulement 6 clics séparent la page d'accueil de LinkedIn de mon profil [8]. Et sur votre site ?

Références

- [1] https://en.wikipedia.org/wiki/Three-click_rule
- [2] <https://www.youtube.com/watch?v=s7kP7LWM87Y>
- [3] <https://support.google.com/webmasters/answer/35769?hl=en>
- [4] <https://www.seroundtable.com/google-100-links-17727.html>
- [5] <https://twitter.com/i/directory/profiles>
- [6] <https://www.linkedin.com/directory/people-a>
- [7] <https://www.facebook.com/directory/people/>
- [8] <https://www.linkedin.com/in/ccherchi>



Cédric CHERCHI, *Expert SEM, agence Onelead* (<http://www.onelead.fr>).