

SCREAMING FROG : VISUALISATION ET ANALYSE APPROFONDIE

Posted on 16 décembre 2020

Après avoir exploré la configuration de l'outil Screaming Frog et initié l'interprétation des crawls effectués les mois précédents, faisons place ce mois-ci à deux problématiques importantes : la présentation des dernières fonctionnalités de la version 14 de l'outil, ainsi que ses différentes (et nombreuses) possibilités de représentation graphique de l'arborescence d'un site. Et, là aussi, les fonctionnalités sont nombreuses !



Troisième article de la série sur le crawler Screaming Frog, après avoir vu comment [bien configurer plusieurs paramètres son crawl](#), et [analyser une partie des résultats](#) tout en évitant des erreurs d'interprétation.

Nous verrons dans cet article les fonctionnalités de visualisation de l'outil, ainsi que l'option d'analyse d'exploration qui permet de calculer un certain nombre de métriques post-crawl.

Mais avant tout chose, nous allons rapidement passer en revue les nouveautés de la version 14.0 de l'outil, sortie le 23 novembre dernier.

Nouveautés de la version 14.0

The screenshot displays the Screaming Frog SEO Spider 14.0 interface. The main table lists 11 crawled URLs with their content types and status codes. The right-hand panel shows 'Crawl Data' with a table of internal links and a donut chart. The bottom panel shows detailed metadata for the selected URL.

Address	Content Type	Status Code	Status	Indexa
1 https://www.abondance.com/	text/html, charset=UTF-8	200	OK	Indexa
2 https://www.abondance.com/wp-content/uploads/2020/12/bandeau-moteur-engine-algorith...	image/jpeg	200	OK	Indexa
3 https://www.abondance.com/wp-content/uploads/2020/10/bandeau-traffic-route-voitures-m...	image/jpeg	200	OK	Indexa
4 https://www.abondance.com/wp-includes/images/rss.png	image/png	200	OK	Indexa
5 https://www.abondance.com/wp-content/uploads/2020/10/bandeau-justice-aveugle-statue-...	image/jpeg	200	OK	Indexa
6 https://www.abondance.com/wp-content/uploads/2020/11/bandeau-nessie-mythes-seo-ser...	image/jpeg	200	OK	Indexa
7 https://www.abondance.com/20201124-44193-search-console-le-bug-sur-les-actions-man...	text/html, charset=UTF-8	200	OK	Indexa
8 https://www.abondance.com/20201207-44259-goossips-displaynone-visibilityhidden-mem...	text/html, charset=UTF-8	200	OK	Indexa
9 https://www.abondance.com/presentation-abondance	text/html, charset=UTF-8	200	OK	Indexa
10 https://www.abondance.com/audit-seo	text/html, charset=UTF-8	200	OK	Indexa
11 https://www.abondance.com/20201126-44206-google-et-les-recherches-en-contexte.html	text/html, charset=UTF-8	200	OK	Indexa

Overview	Site Structure	Response Times	API	S
Total External URLs	13	18.57%		
Crawl Data				
Internal				
All	57	100%		
HTML	38	66.67%		
JavaScript	0	0%		
CSS	1	1.75%		
Images	18	31.58%		
PDF	0	0%		
Flash	0	0%		
Other	0	0%		
Unknown	0	0%		

Name	Value
Address	https://www.abondance.com/20201126-44206-google-et-les-recherches-en-contexte.html
URL Encoded Address	https://www.abondance.com/20201126-44206-google-et-les-recherches-en-contexte.html
Content Type	text/html, charset=UTF-8
Status Code	200
Status	OK
HTTP Version	HTTP/1.1
Indexability	Indexable
Size	62.6 KB
Title 1	Google et les recherches en contexte - Actualités SEO et moteurs - Abondance
Title 1 Length	76
Meta Description 1	26 novembre 2020 - Google nous explique dans un article de blog les progrès du moteur de recherche en termes de recherche contextuelle.

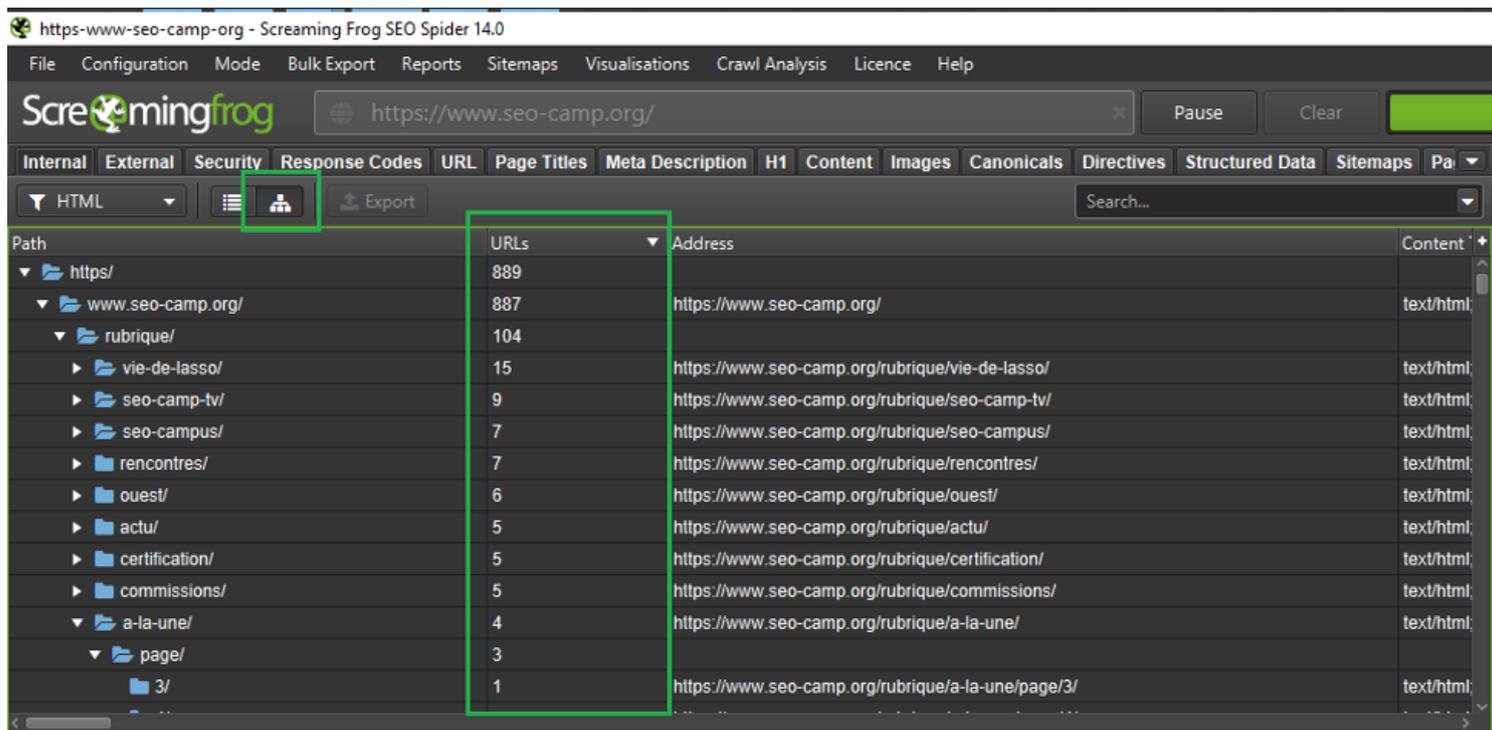
Modification de la couleur de l'interface (Menu Configuration > User interface > Theme)

Hormis l'arrivée de l'interface « Dark Mode » qui peut être plus agréable à l'œil, voici des nouveautés qui vous seront plus utiles d'un point de vue opérationnel:

- Possibilité de générer des exports via Google Sheets en ayant préalablement connecté votre compte (option également disponible pour les crawls préprogrammés et leurs exports) ;
- Extraction de l'ensemble des en-têtes http pour chaque requête : cela peut s'avérer très pratique pour identifier les directives de cache sur les URL spécifiques ;
- Possibilité de collecter les informations contenues dans les cookies.

Structure de site agrégée

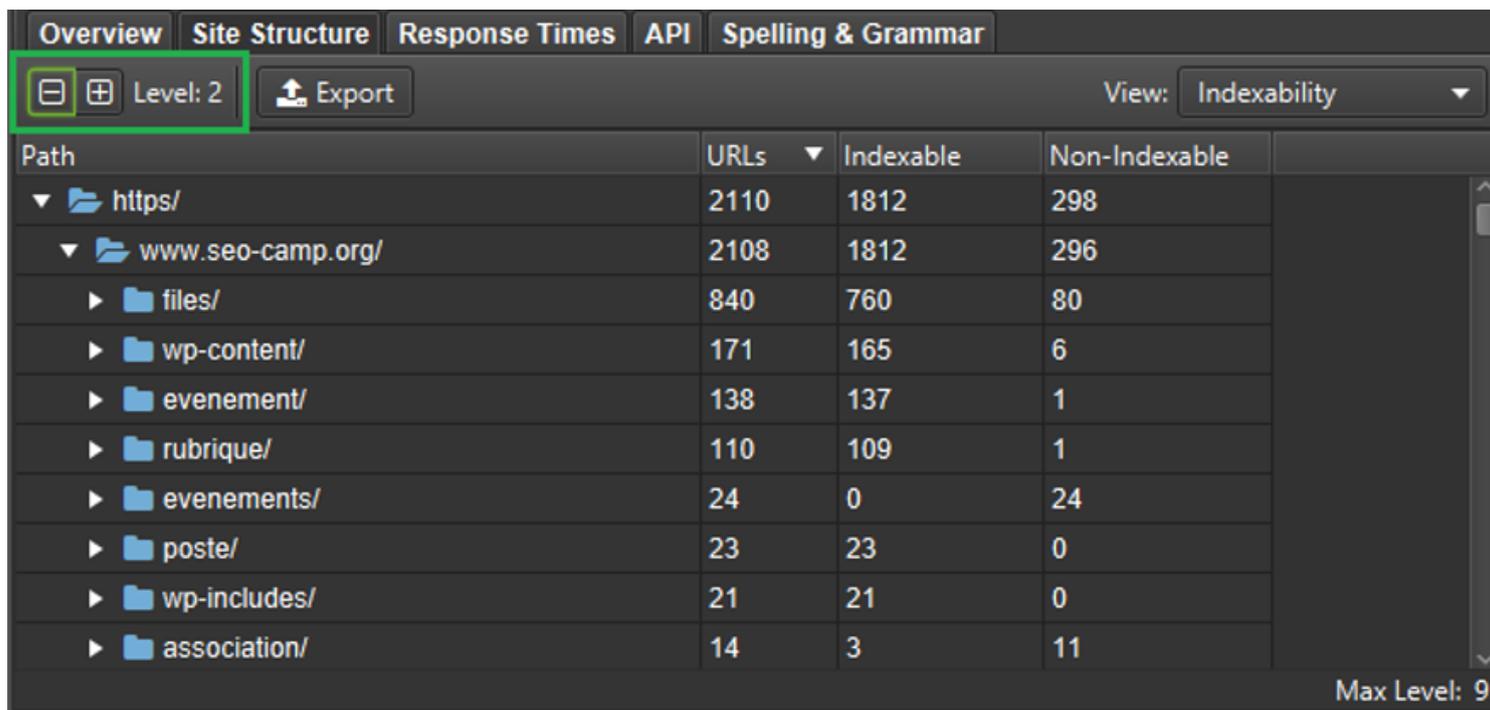
L'apparition d'une nouvelle colonne dans l'onglet « Internal » peut s'avérer pertinente pour certains sites : il s'agit de la possibilité de connaître le nombre d'URL en fonction de leur structure par répertoire (ou chemin, parfois virtuel *via* la réécriture d'URL) :



Structure agrégée du site avec le nombre d'URL par répertoire

On visualise ainsi rapidement le nombre de pages pour chaque catégorie du site, ce qui permet de se faire une bonne idée de ses contenus.

En complément, la partie graphique « Site structure » située dans la colonne de droite de l'outil, permet d'avoir cette vision avec la possibilité de naviguer niveau par niveau (à condition que les niveaux des répertoires représentent l'arborescence réelle des contenus pour que cela ait du sens) :



Path	URLs	Indexable	Non-Indexable
https/	2110	1812	298
www.seo-camp.org/	2108	1812	296
files/	840	760	80
wp-content/	171	165	6
evenement/	138	137	1
rubrique/	110	109	1
evenements/	24	0	24
poste/	23	23	0
wp-includes/	21	21	0
association/	14	3	11

Vision « Site structure > Indexability » : regroupement des pages en niveau 2 des répertoires (en fonction des URL)

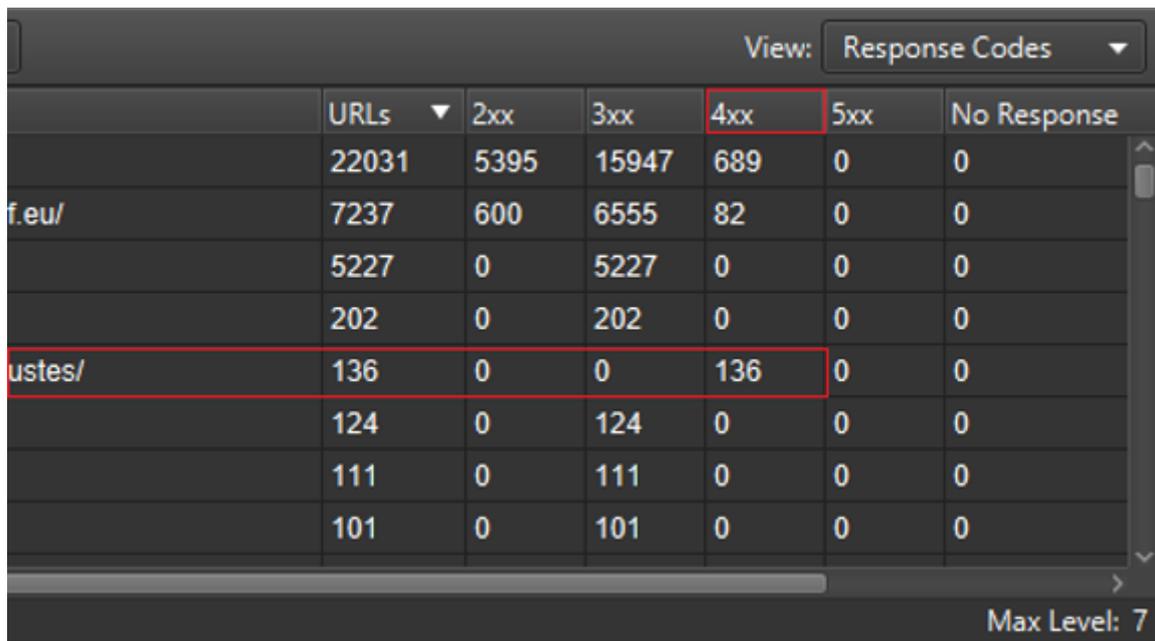
La navigation entre les niveaux de la structure des répertoires (à ne pas confondre avec la profondeur réelle des contenus, visible en utilisant « Crawl depth » dans le menu déroulant de ce volet) peut se faire grâce aux boutons « + » et « - ».

On peut ainsi voir pour chaque regroupement de contenu (en fonction des répertoires dans l'URL), différentes données dont :

- la répartition des codes réponses ;
- la profondeur des contenus ;
- l'indexabilité.

Ces données relatives à l'indexabilité aident à mieux visualiser les types d'éléments qui ne peuvent pas être indexés, comme les éléments contenus dans le répertoire « /evenements/ » dans l'exemple ci-dessus.

Pour ce qui est des codes réponses, on pourra rapidement isoler l'ensemble d'un répertoire pour lesquels les éléments qu'il contient répondent en 404 :



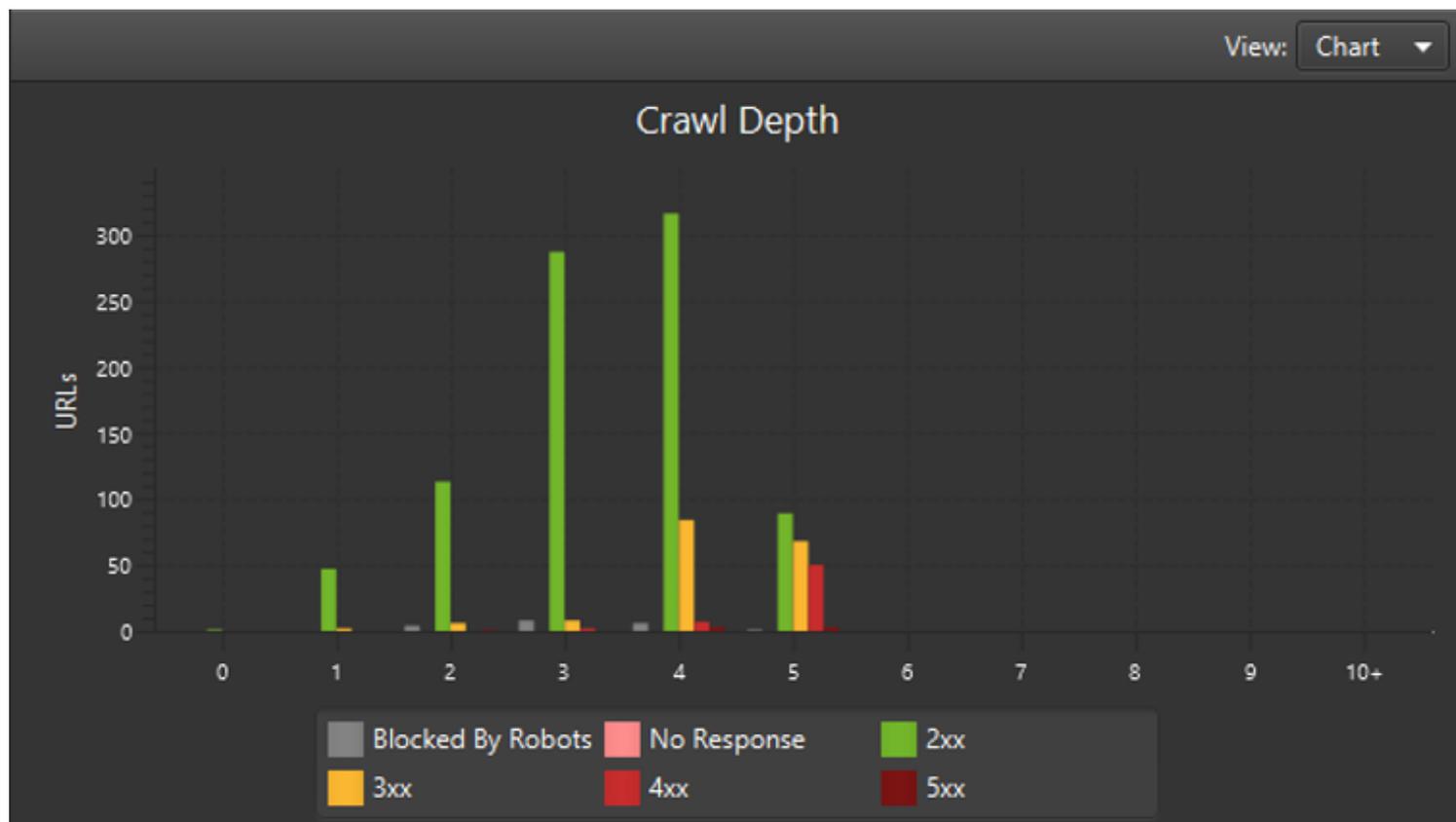
	URLs ▼	2xx	3xx	4xx	5xx	No Response
	22031	5395	15947	689	0	0
f.eu/	7237	600	6555	82	0	0
	5227	0	5227	0	0	0
	202	0	202	0	0	0
ustes/	136	0	0	136	0	0
	124	0	124	0	0	0
	111	0	111	0	0	0
	101	0	101	0	0	0

View: Response Codes ▼

Max Level: 7

Vision "Site structure" par codes réponses

L'une des dernières nouveautés de la version 14.0 se trouve dans le graphique de cet onglet « Site structure », et permet de visualiser les codes réponses (et la crawlabilité restreinte via le robots.txt) des contenus en fonction de leur profondeur :



Profondeur des URL en fonction des codes réponses

Nous constatons ici un nombre particulièrement élevé d'erreurs 4xx (403, 404, etc.) au cinquième niveau de profondeur, ainsi que de nombreuses redirections en niveau 4 et 5.

Visualisation

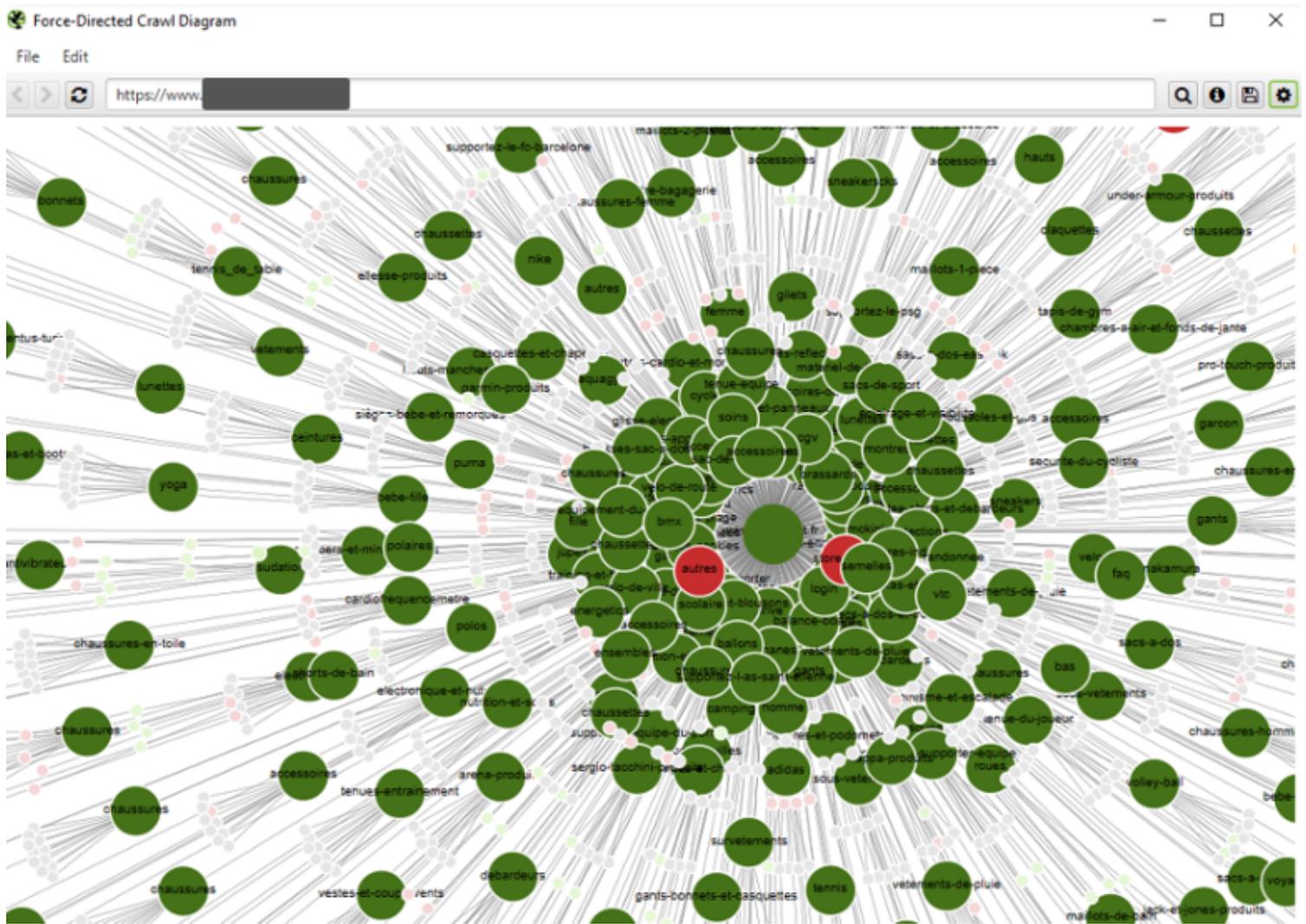
Depuis 2 ans, Screaming Frog dispose de différents outils de visualisations. Ils permettent de mettre en avant certaines faiblesses SEO de sites, mais encore faut-il bien comprendre la façon dont fonctionnent ces outils

Maillage

Dans les différents menus, le menu « Visualisations » permet d'avoir une représentation visuelle du site de deux façons différentes :

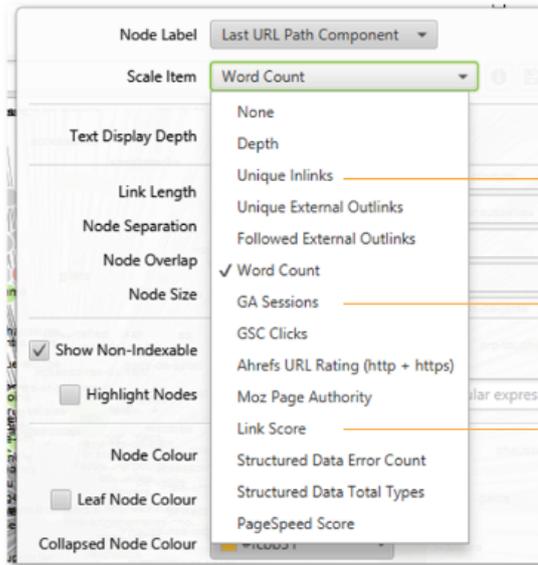
- « *Force-Directed Directory Tree Diagram* » permet d'effectuer une visualisation de la structure du site en fonction de l'arborescence des répertoires ;
- « *Force-Directed crawl Diagram* » permet d'effectuer une visualisation du maillage du site, en fonction du niveau de profondeur des URL.

Attention cependant, « *Force-Directed crawl Diagram* » ne représente **pas l'intégralité du maillage interne, mais uniquement une partie de la navigation descendante**, où ne sont affichés que le chemin le plus rapide via les liens pour arriver à chaque URL. Par exemple, si une page Catégorie est accessible depuis la page d'accueil, mais également via une page de sous-catégorie, c'est le chemin le plus rapide (via la *homepage*) qui sera dessiné.



Visualisation - navigation descendante

Cette fonction, largement paramétrable, permet de visualiser différentes problématiques en fonction de la configuration effectuée :

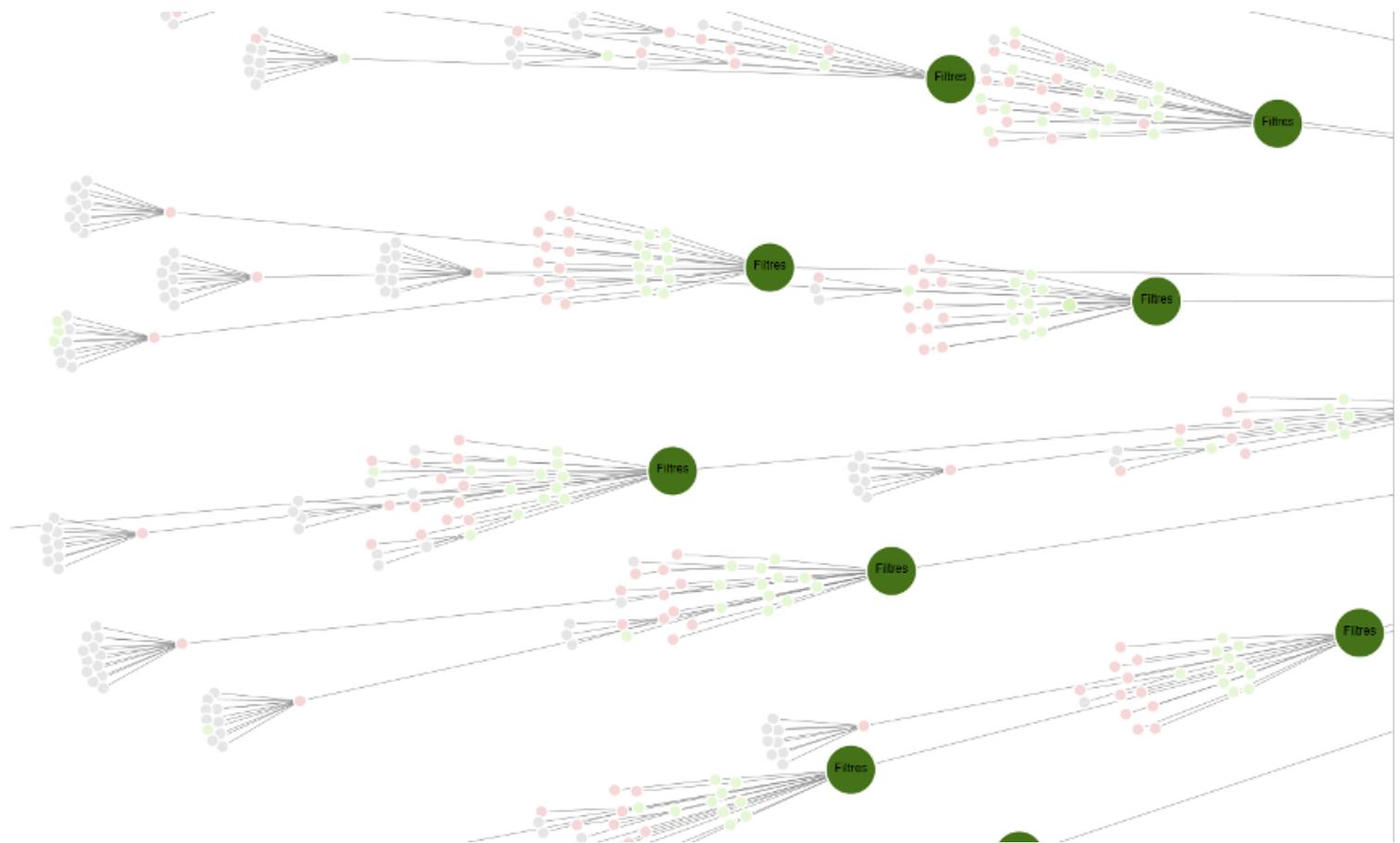


- Voir si des URL fortement maillées ne sont pas indexables (Canonical, redirection, noindex) → perte de PageRank interne
- Voir où se situent dans le maillage les URL qui génèrent le plus de trafic
- [Link score](#) : page rank interne de screaming Frog

Paramétrage de la visualisation

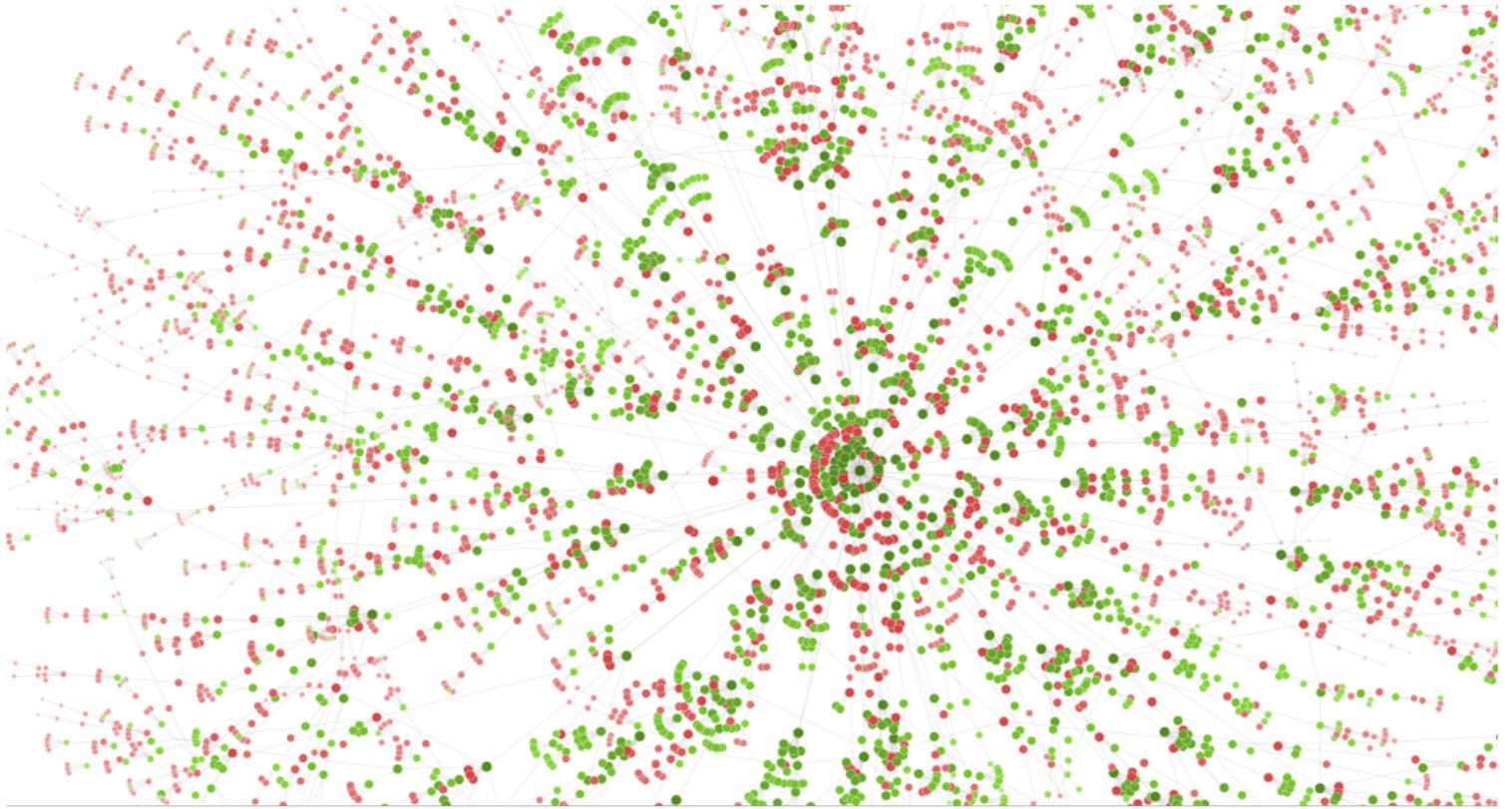
La largeur de chaque cercle (représentant une URL précise) peut représenter différents KPI. Par défaut, ce diamètre est associé à la profondeur de l'URL (plus le rond est petit, plus l'URL est profonde).

Le fait de pouvoir afficher en rouge (couleur modifiable) les pages identifiées comme « Non-indexable » (URL non canoniques, pages en « noindex », pages redirigées », etc.) peut mettre en avant des fuites de popularité, avec du pagerank interne diffusé de à des pages non indexées :



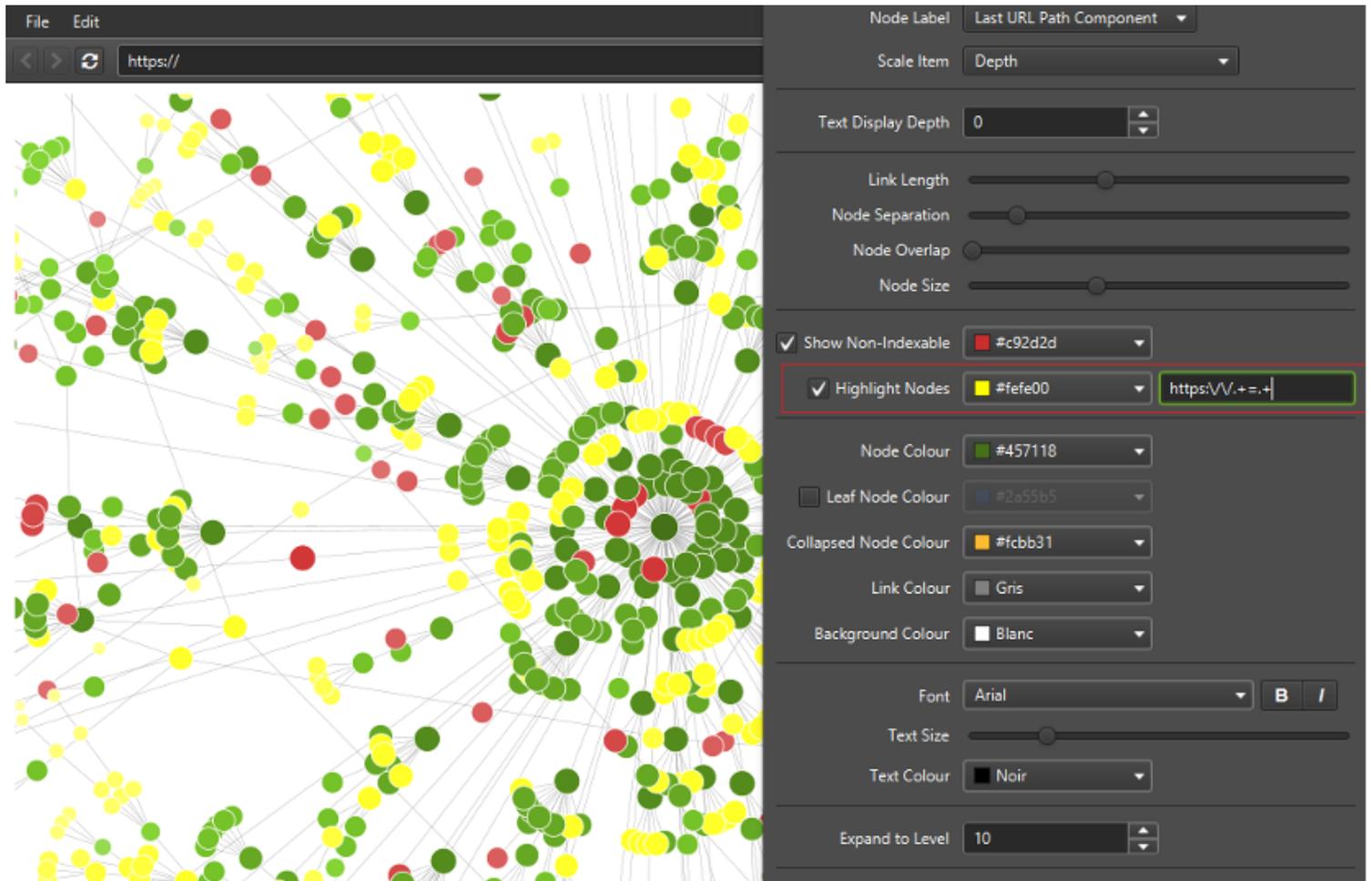
URL maillées mais non indexables

Dans l'exemple ci-dessous, nous constatons la présence de nombreuses pages avec une URL canonique (après avoir survolé une partie des points rouges pour avoir le détail par URL), ce qui signifie la présence de liens vers ces pages qui déséquilibrent la structure du site, et augmente probablement le taux de crawl sur des URL pouvant être considérées comme dupliquées :



Présence d'un nombre important de pages avec une balise canonique en rouge

Pour isoler visuellement certaines URL en fonction de leur structure, il est possible d'utiliser des expressions régulières, *via* le champ suivant dans les paramètres de la visualisation. Ici, nous allons mettre en avant les URL qui contiennent des paramètres (ex : `/path?cle=valeur`) :



Utilisation d'expressions régulières dans la visualisation

Nous utilisons l'écriture suivante pour indiquer que nous cherchons les URL qui contiennent « = » dans l'URL : `https://.+=.`

L'utilisation d'un mode de calcul du PageRank interne est également possible. Il faudra au préalable lancer un « Crawl analysis » dans le menu correspondant. Avant cela, en modifiant le « Scale item » sur « Unique Inlinks », on pourra visualiser les pages peu maillées (ou à l'inverse les pages fortement maillées) :



Les cercles les plus larges correspondent aux pages recevant de nombreux liens uniques

De la même manière, une fois connecté à la Search Console, on pourra visualiser de façon rapide où se situent les pages qui génèrent le plus de trafic (« Scale item » > « GSC clicks »), et mettre en avant les pages relativement hautes dans l'arborescence qui en génèrent peu...

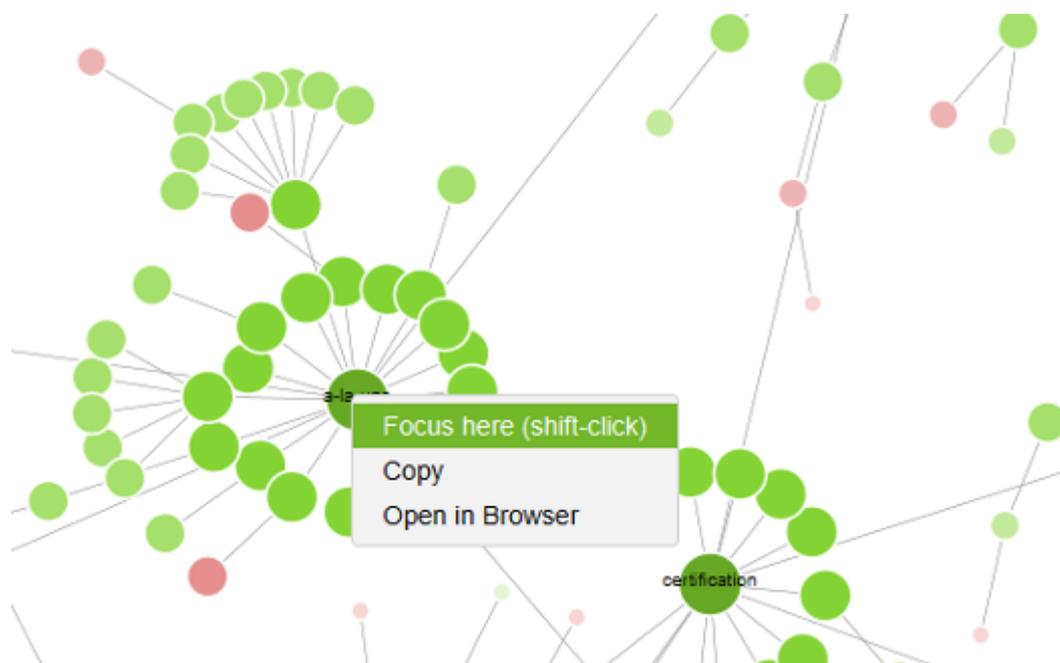
Cette visualisation est disponible au global, mais également à partir d'une URL précise ce qui permettra une vision plus claire sur un groupe de pages :

200	https://www.yapasdequoi.com/category/seo	4	text/html; charset=UTF-8	301
201	https://www.yapasdequoi.com/category/seo-people	1	text/html; charset=UTF-8	200
202	https://www.yapasdequoi.com/category/seo-people/		text/html; charset=UTF-8	301
203	https://www.yapasdequoi.com/category/seo-people/page/2		text/html; charset=UTF-8	200
204	https://www.yapasdequoi.com/category/webperf		text/html; charset=UTF-8	200
205	https://www.yapasdequoi.com/cms/1135-drupal-et-les-balises-title-et		text/html; charset=UTF-8	200
206	https://www.yapasdequoi.com/cms/1576-allez-pschiit-les-commentai		text/html; charset=UTF-8	200
207	https://www.yapasdequoi.com/cms/1692-ajout-du-tag-more-automat		text/html; charset=UTF-8	200
208	https://www.yapasdequoi.com/cms/224-afficher-un-bloc-drupal-sur-u		text/html; charset=UTF-8	200
209	https://www.yapasdequoi.com/cms/2332-wordpress-et-les-url-de-cat		text/html; charset=UTF-8	200
210	https://www.yapasdequoi.com/cms/2902-wordpress-entre-guillemets		text/html; charset=UTF-8	200

name	Value
adresse	https://www.yapasdequoi.com

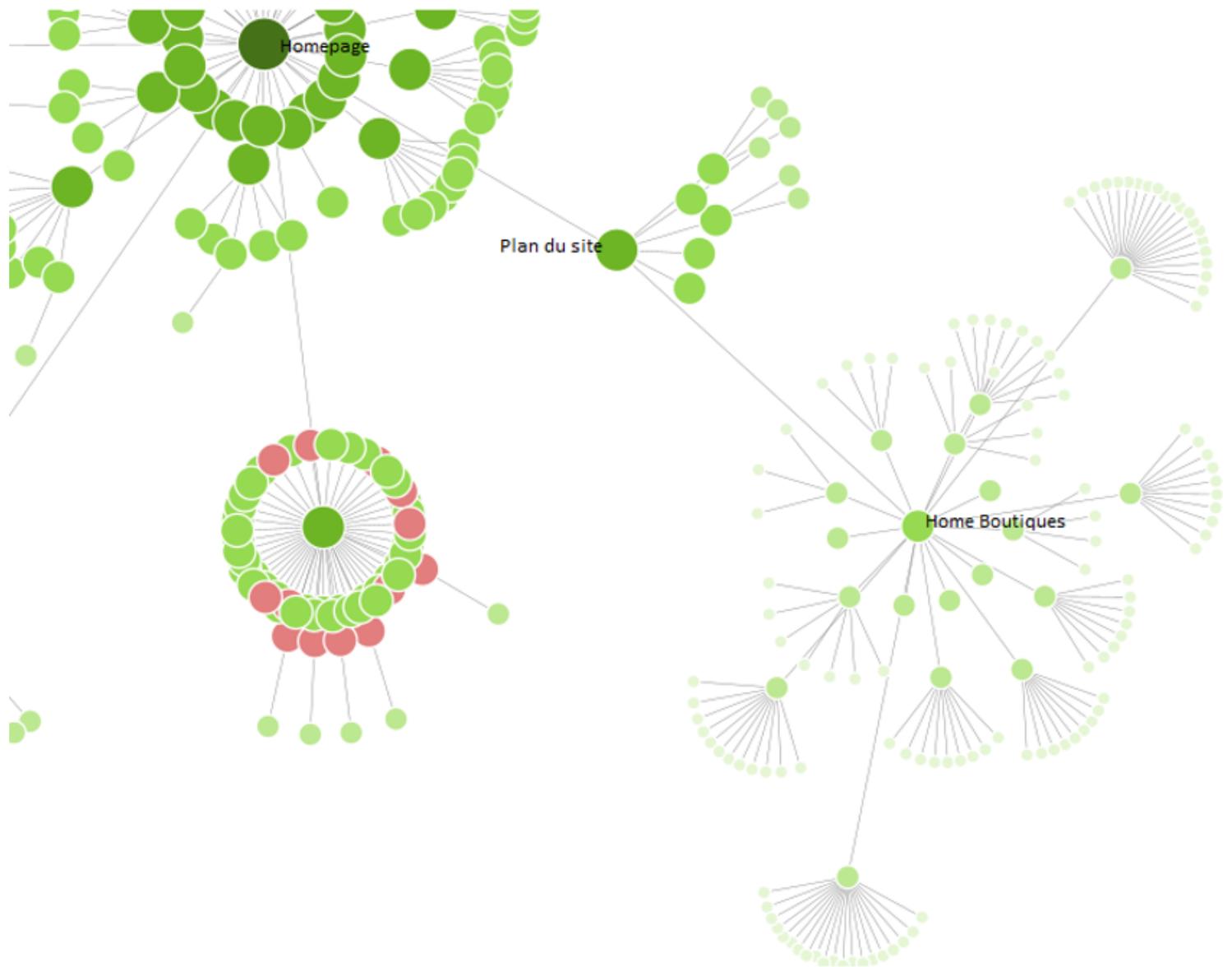
Visualisation sur une URL spécifique

La même visualisation pour des URL spécifiques est également possible via un clic-droit directement sur le diagramme, puis l'option « Focus here » :



Focus à partir d'une URL spécifique

Dans l'exemple ci-dessous, nous pouvons facilement identifier un défaut de maillage sur la version mobile du site crawlée :



Visualisation d'un défaut de maillage sur une version mobile

Ce site vitrine avec des boutiques ayant pignon sur rue possède bien un lien depuis sa page d'accueil vers le *store locator*, mais uniquement dans sa version pour Ordinateur. La version mobile du site possède un menu différent, et ne propose pas de liens vers le *store locator* depuis la page d'accueil. Il est nécessaire de passer par la page « Plan du site » pour accéder aux boutiques, ce qui n'est pas idéal d'un point de vue utilisateur, et ne facilite pas la transmission de popularité vers les pages boutiques (et département) *via* la navigation descendante.

Ancre et corps d'une page

D'autres outils de visualisation sont disponibles, dont un qui permet de mettre en avant les principales occurrences



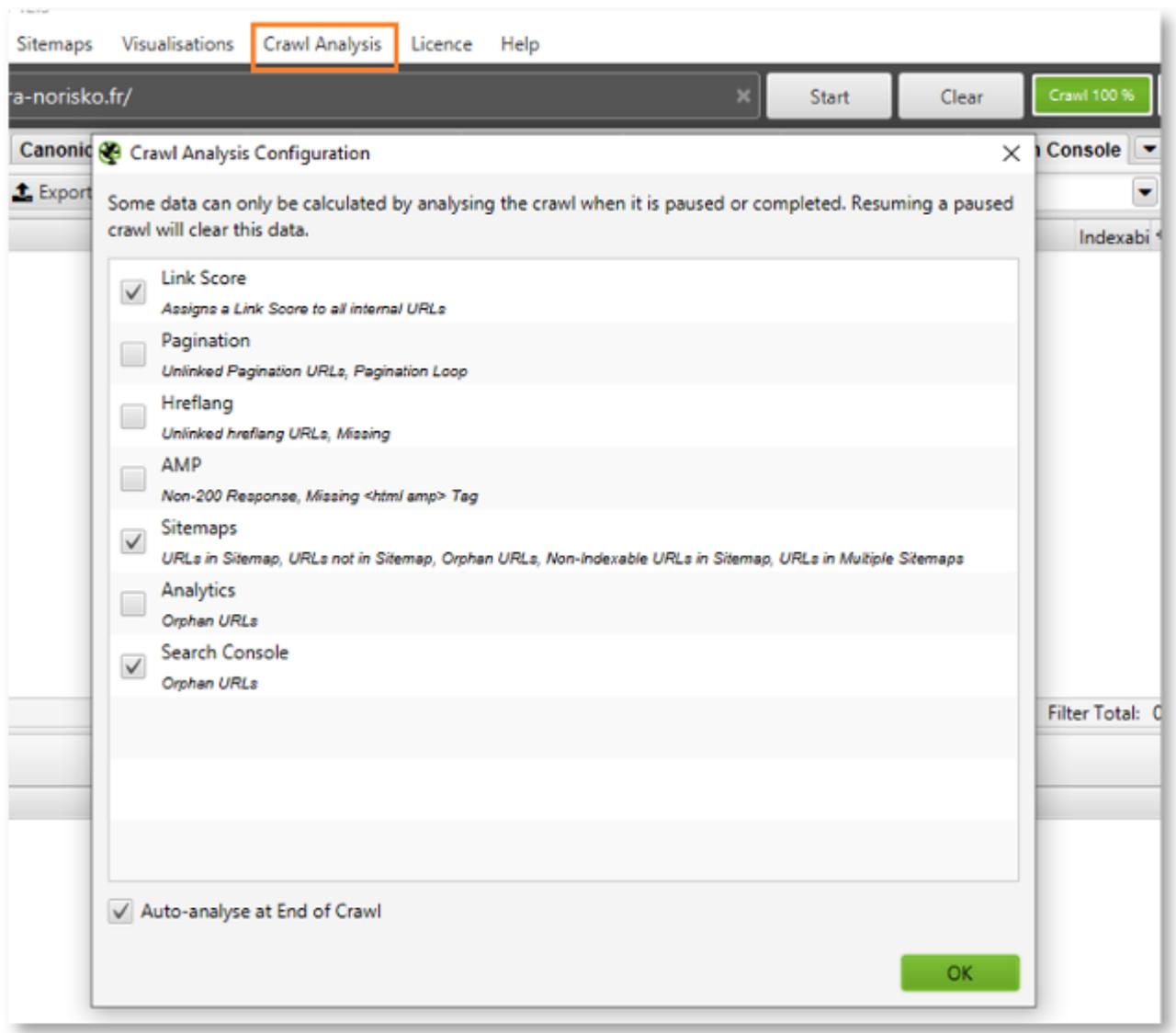
Nuage des ancres de liens vers une page

Cela ne renforce pas le poids de la page sur des requêtes plus ciblées (les mois de l'année étant peut pertinent).

Analyse approfondie avec « Crawl analysis »

Cette fonction, à exécuter une fois le crawl terminé, permet entre autres d'exécuter un certain nombre de comparaison entre les URL trouvées via différents moyens, et les URL crawlées.

On détectera ainsi les URL orphelines (URL trouvées mais ne recevant pas de liens), afin de mettre en évidence des problèmes de maillage (et/ou de configuration). Un calcul du pagerank interne (Link Score) pourra également être exécuté.



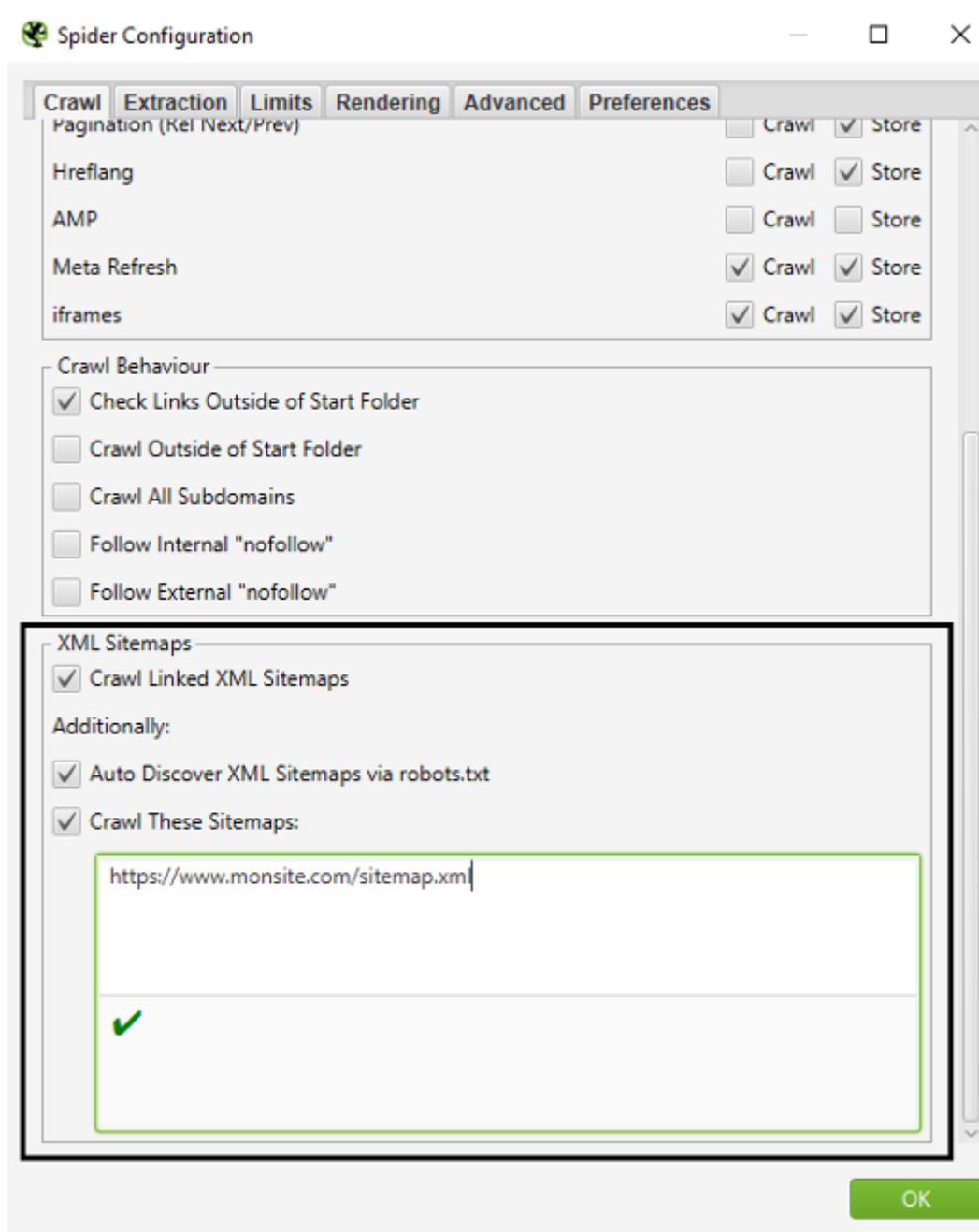
Crawl analysis

Voici les différentes comparaisons qui peuvent être effectuées entre les URL crawlées *via* des liens, et celles découvertes par d'autres biais que des liens :

- URL de pagination (balises prev/next) ne recevant pas de liens ;
- Balises hreflang manquantes ou non maillées ;
- URL AMP qui ne répondent pas correctement ou absence de bi-directionnalité (canonical vs AMP URL) ;
- URL maillées et absentes des sitemaps, URL présentes dans les sitemaps et non maillées, URL non-indexables dans les sitemaps, etc. ;
- URL détectées dans Google Search Console ou Google Analytics (ayant potentiellement généré du trafic ou des

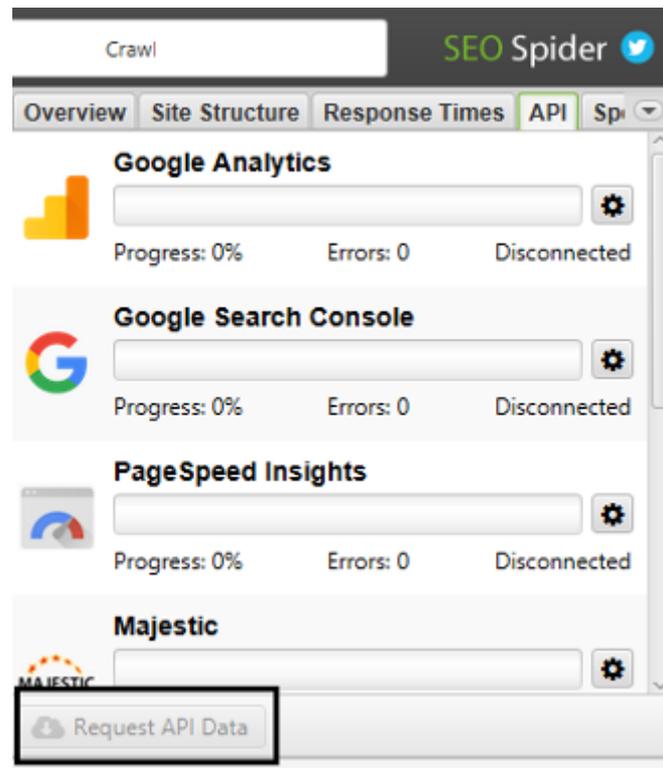
impressions), mais non découvertes *via* le système de liens.

Concernant les sitemaps, il aura fallu paramétrer en amont du crawl les URL de ces derniers :



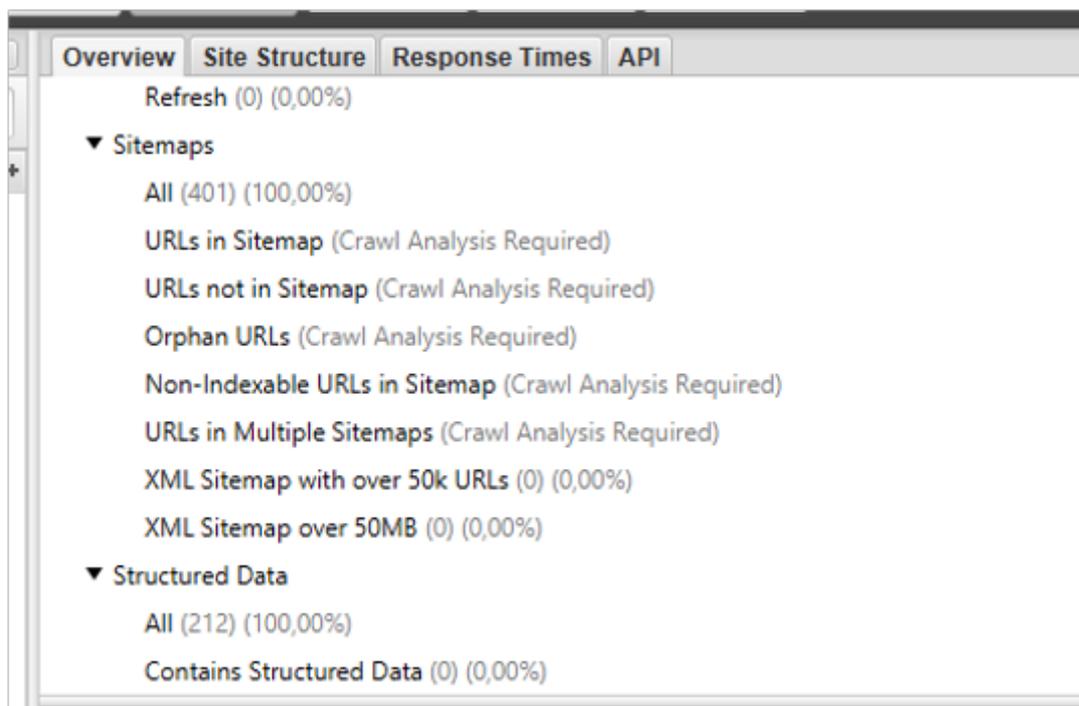
Utilisation des Sitemaps dans le cadre du crawl

Pour ce qui est des outils de Google, il est possible de les connecter *a posteriori* (*Configuration > API access*), puis de lancer des requêtes aux API avant de lancer la fonction « Crawl analysis » :



Requêtes vers les API afin de découvrir des URL orphelines

Pour chacun des éléments observés, de nouveaux filtres seront disponibles dans les onglets concernés. Voici un exemple pour la partie « Sitemaps » :



URL des Sitemap vs crawl : nécessité de lancer une analyse

Un fois l'analyse de crawl lancée, on pourra identifier la présence d'URL détectées dans les Sitemaps, mais qui ne reçoivent pas de liens *via* le maillage interne, afin de mener à bien les optimisations adéquates.

Un export de l'ensemble des URL orphelines est disponible dans le menu « Report » (*Orphan pages*), afin de découvrir l'ensemble des URL trouvées et non maillées, ainsi que leur provenance (Sitemap, Search Console, etc.). Cette fonction offre de réelles opportunités, afin de permettre une amélioration du maillage interne, les URL orphelines n'étant pas rares sur de nombreux sites web.

Une meilleure distribution de la popularité, tout en respectant les affinités sémantiques entre vos pages vous permettra donc d'améliorer votre visibilité.

Saviez-vous que les outils de crawl comme Screaming Frog recelaient tant de fonctionnalités ?



Aymeric Bouillat, Consultant SEO senior chez Novalem (<https://www.novalem.fr/>)