# Cocon.se : voir la structure de son site avec les yeux de Google

*Par Sylvain Deauré*

| | | |
|---|---|---|
| **Domaine :** | Recherche | **Référencement** |
| **Niveau :** | Pour tous | **Avancé** |

*La représentation graphique de la structure et de l'arborescence d'un site devient une étape indispensable lorsqu'on désire auditer l'optimisation d'une source d'information pour Google. Comprendre comment le contenu est articulé, comment le visite et le voit un moteur de recherche, permet de détecter rapidement et visuellement d'éventuelles erreurs et de les corriger dans la foulée. L'outil Cocon.se permet ainsi d'effectuer ce travail et de voir votre site « avec les yeux du moteur ». Explications par le concepteur de l'outil...*

En SEO, on utilise la plupart du temps deux grandes familles d'optimisations : le « off site » et le « on site ». Le « off site » s'oriente principalement autour du linking. Il s'agit sans doute du levier le plus souvent utilisé, mais également le plus facilement pénalisé par Google, notamment avec ses filtres Penguin.

Les leviers « on site » sont moins bien exploités et semblent parfois accessoires, moins importants. Pourtant, il n'en n'est rien. En levier « on site », on peut bien sûr optimiser le contenu, mais il ne faut pas oublier le **maillage interne**, la façon dont les pages de son site sont reliées entre elles. Google, dans ses guidelines (https://support.google.com/webmasters/answer/35769?hl=fr) , insiste d'ailleurs sur cet aspect. Voici quelques extraits choisis : *« Accéder à toutes les pages du site depuis un lien situé sur une autre page accessible » « Limitez le nombre de liens par page » « Empêcher l'exploration d'espaces infinis » « Évitez de faire appel à des ID de session ou à des paramètres d'URL » « Veillez à ce que tous les liens redirigent vers des pages Web en ligne »* et surtout, *« **Créez une arborescence de pages claire et conceptuelle.** »*

Il semble évident que pour que Google puisse correctement parcourir un site, l'indexer, et décider de quoi celui-ci parle, de la thématique à laquelle il appartient, du besoin auquel il répond, etc., la « cartographie » du site, les chemins qu'il propose sont un élément clé.

Dans cet article, nous allons vous présenter le site Cocon.Se (http://cocon.se/), un outil qui a pour vocation de vous montrer votre site comme le voit un robot, au travers de son arborescence et de ses liens. L'objectif de cet outil est d'obtenir une représentation graphique du site, sur laquelle se baser pour détecter des erreurs de structure et de conception.

## La genèse de l'outil

Tout est parti comme souvent d'un besoin personnel, et d'une intuition. Il y a deux ans de cela, on pouvait voir quelques visualisations de sites, mais cela ressemblait plus à un gadget à l'époque. On pouvait considérer ces visualisations comme expérimentales. Ou on les utilisait pour mettre en évidence un souci qu'on avait déjà identifié, ce qui limitait leur intérêt.

Nous avions l'intuition qu'on pouvait faire bien plus que cela, et qu'avec une bonne visualisation, une bonne spatialisation, on pouvait tirer des analyses concrètes de la représentation graphique d'un site.

L'idée a fait son chemin depuis, et d'autres personnes (en France comme à l'international) commencent à utiliser des visualisations graphiques pour montrer au client ses problèmes de structure, voir l'arborescence comme un cocon, pour diagnostiquer des soucis techniques ou conceptuels, plus facilement qu'avec des indicateurs purement numériques.

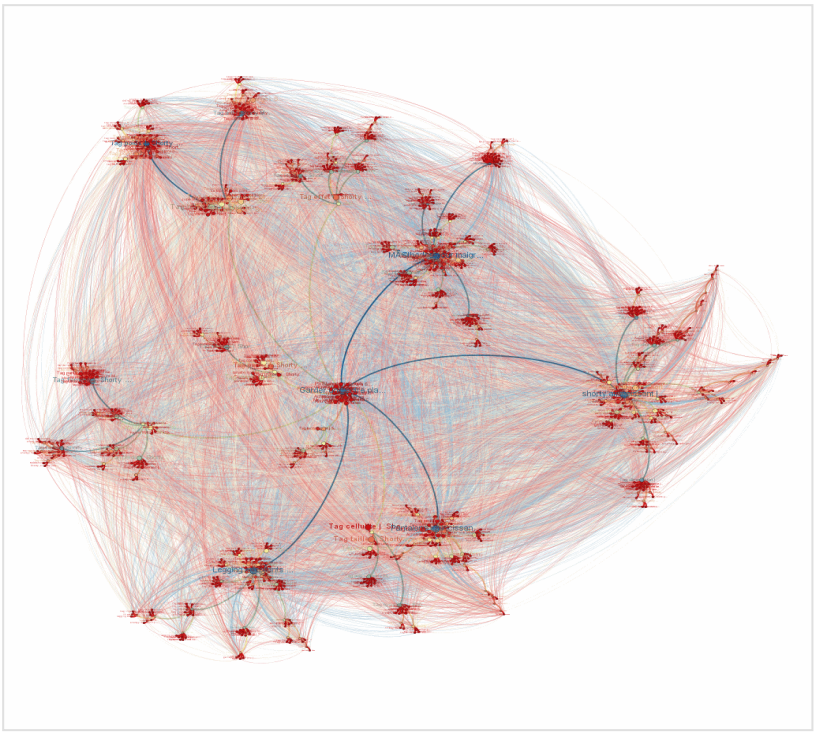

*Fig. 1 : Une des premières vues expérimentales, avec Gephi (voir article à son sujet par ailleurs dans la lettre).*

Au début, pour prototyper, expérimenter, nous utilisions Gephi. Il permet assez facilement d'obtenir des vues globales, macroscopiques. Mais on est vite confronté à des problèmes de lenteur, de ressources, de manipulations longues si on veut aller plus loin avec cet outil. C'était trop laborieux et une fois le principe validé, nous avons basculé sur du développement personnalisé.

## Comment ça marche, en quelques mots...

Le fonctionnement de l'outil Cocon.se est le suivant :

• On crawle le site à la façon de Google, en suivant les liens depuis une page de départ (le plus souvent la page d'accueil) ;
• Au passage, on détecte d'éventuelles erreurs de base ;
• On calcule des indicateurs de structure locaux et globaux, dont le PageRank interne ;
• On calcule des organisations spatiales des pages du site, qui rendent compte de sa hiérarchie telle que perçue par un moteur ;
• On affiche pages et liens sous une forme structurée qui met en évidence l'arborescence interne du site.

## Le crawl

Le Crawl est le premier contact d'un moteur avec votre site. C'est lors du crawl que moteur se fait une idée de votre site, des pages importantes. Avant même d'avoir tout crawlé, Google a déjà décidé de classer en premier les pages « tag » ou « catégorie » de votre site, par exemple, selon le maillage interne partiel qu'il a rencontré. C'est un élément vraiment critique, essentiel, et qu'il est pourtant facile (et courant) de faire de travers.

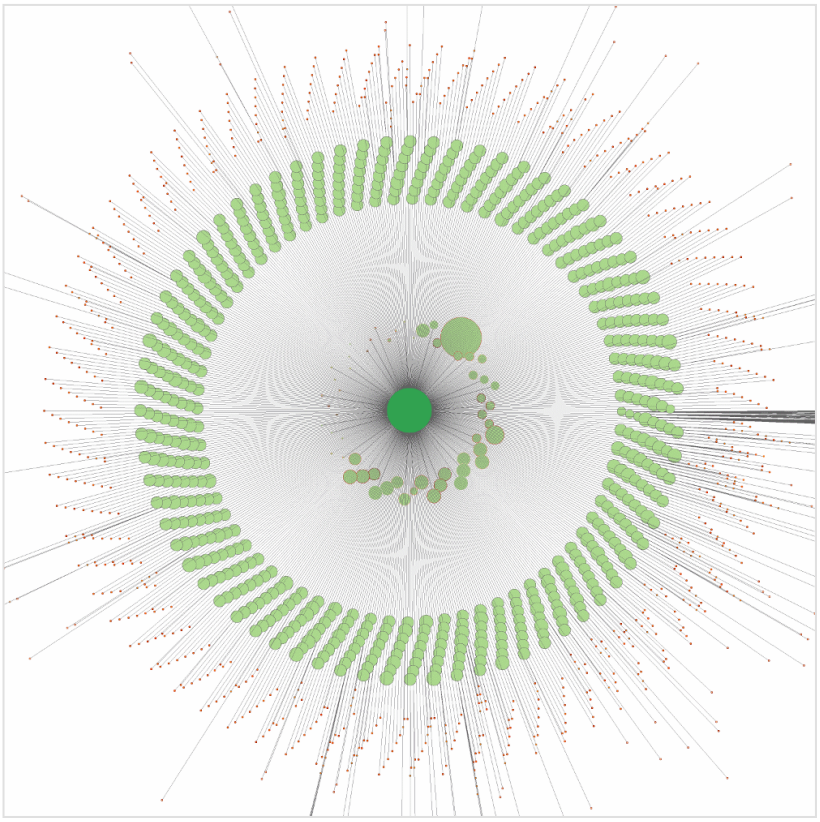

*Fig. 1 : Un exemple de site surmaillé, avec un énorme mega-menu (ronds verts dans le cercle autour de la page d'accueil)*

Le fait d'avoir développé notre propre crawler est un atout important : cela nous a permis de nous rendre compte du nombre de points essentiels qui peuvent être mal codés, ou qui posent des problèmes à un robot. Corriger les problèmes de crawl permet de se recentrer sur l'essentiel, de donner tout de suite toutes les billes à Google, toutes les infos au bon endroit, dans le bon ordre, et sans ambiguïté.

Même si Google utilise des tonnes de rustines pour corriger des problèmes de crawl et de structure (*canonical, redirections, no follow, robots.txt, rel prev/next*...) cela ne veut pas dire qu'il faut les utiliser à tout bout de champ.

Un site bien conçu dès le départ, qui n'aura pas besoin de toutes ces rustines, aura toujours une longueur d'avance sur un site rafistolé à coup de « verrues ».

*In fine*, notre crawler a pour objectif d'explorer votre site de la même manière que Googlebot, afin de voir les mêmes informations locales que Google, et fournir donc une représentation similaire à ce qu'en voit le moteur de recherche.

## Les algorithmes et calculs

Nous n'allons pas nous étendre ici sur les aspects purement techniques de notre infrastructure. Disons juste qu'elle a été pensée dès le début pour être performante et « scalable ».

Nous nous basons sur des algorithmes connus (parcours de graphe, arbre recouvrant, surfeur aléatoire / PageRank...) et avons également toutes une série d'algorithmes *maison* pour extraire les données importantes lors du crawl, optimiser les calculs, les traiter de manière répartie, et surtout calculer une organisation efficace des pages du site pour la visualisation.

Coté langages, nous utilisons, selon les composants, les technologies PHP, Python, Javascript, C#, Node.Js, etc.

## La visualisation du site

Une visualisation du maillage interne permet de lever le doute, d'avoir une vraie réponse à la question « mon site est-il vu par Google comme je l'ai pensé/conçu ? ». On y trouve souvent une bonne marge de manœuvre, qui donne un levier concurrentiel énorme, sans le risque associé à des actions de linking plus ou moins factice.

On pourrait penser que la restitution des datas est une brique « simple », et que plus on a de données, plus on a d'information. En fait, c'est le contraire. Plus on a de datas, plus on est perdu ! La difficulté n'est pas d'avoir beaucoup de données, mais bien de trouver celles qui sont significatives, qui vont structurer le reste et le rendre intelligible et actionnable.

Avec notre crawler, on a l'avantage de pouvoir collecter, filtrer, la data comme on veut, pour la présenter, on l'espère, sous une forme utilisable.

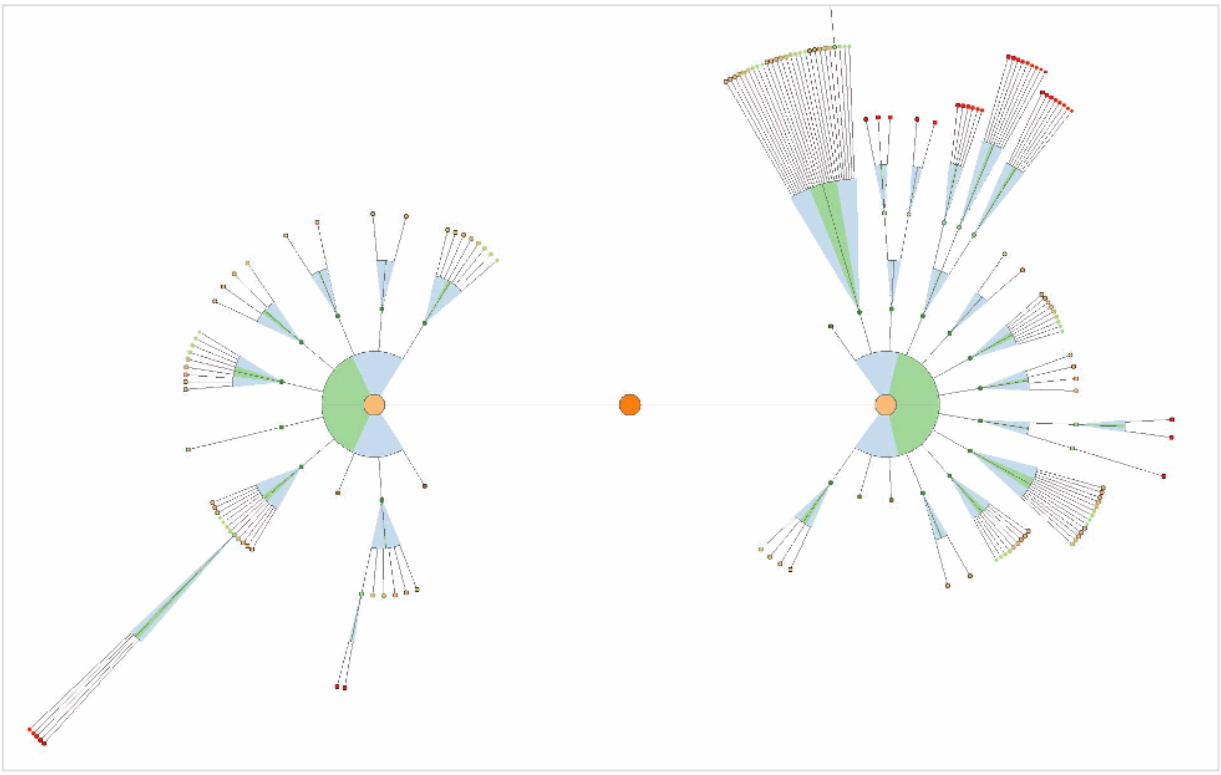

*Fig. 3 : Exemple de visualisation de 2 catégories (silos) d'un site.*

Lorsqu'on gère un site web, on pense parfois que le CMS que l'on utilise est « optimisé seo » ou que le plugin « super opti seo » que l'on a installé est fait pour, que si il nous affiche une coche verte dans son administration, tout se passera parfaitement bien. Ce n'est malheureusement jamais aussi simple. Un blog, avec ses catégories, ses articles chronologiques, son nuage de tags en vrac, n'est pas optimisé pour le Seo. Un CMS E-Commerce, de type Prestashop, par défaut, n'est pas optimisé pour le SEO.

Quand on fait une visualisation de ce type de site, on se rend tout de suite compte de la différence énorme entre l'organisation « pensée » du site, et celle que voit le moteur, qui est bien bien plus brouillonne et complexe !

## *En pratique, comment voir mon site avec Cocon.Se ?*

Voici la procédure à suivre pour visionner votre site avec notre outil :

• Inscrivez-vous sur https://self.cocon.se ;
• Vous aurez accès à quelques sites déjà cartographiés si vous voulez prendre l'interface en main ;
• Renseignez votre profil (important pour la gestion de la TVA et des factures) et choisissez un type de compte. Le compte « *pay as you go* » est gratuit et permet d'acheter des crawls à la demande. Les abonnements sont sans engagement de durée et plus économiques ;
• Cliquez sur « *Nouveau crawl* », copiez/collez l'adresse de votre site depuis la barre d'adresse de votre navigateur vers la zone « *start url* » ;
• Vérifiez et lancez le crawl ;

• Quelques minutes plus tard (selon la taille et vitesse de votre site), vous avez une première photo de votre site disponible dans votre espace d'administration ;
• Vous pouvez ensuite créer d'autres vues, selon d'autres facettes ou depuis d'autres pages de départ.

## *Comment interpréter ce que je vois ?*

Les codes suivants sont utilisés par l'outil pour visualiser graphiquement votre arborescence :

• Les pages du site sont représentées par des disques colorés ;
• La taille du disque rend compte de l'importance relative de la page (Pagerank interne) ;
• La couleur du disque représente le nombre de « clic » depuis la page d'accueil ;
• Les liens entre les pages ne sont, volontairement, pas tous représentés. Vous voyez les liens qui constituent le plus court chemin vers chaque page, depuis l'accueil (la vue interactive, elle, permet d'afficher tous les liens). Ceci permet de mieux mettre en évidence la hiérarchie de vos pages ;
• Les pages sont numérotées, vous avez la correspondance avec l'URL via le menu 'top pages', ou directement au survol de la souris, en passant en vue interactive.

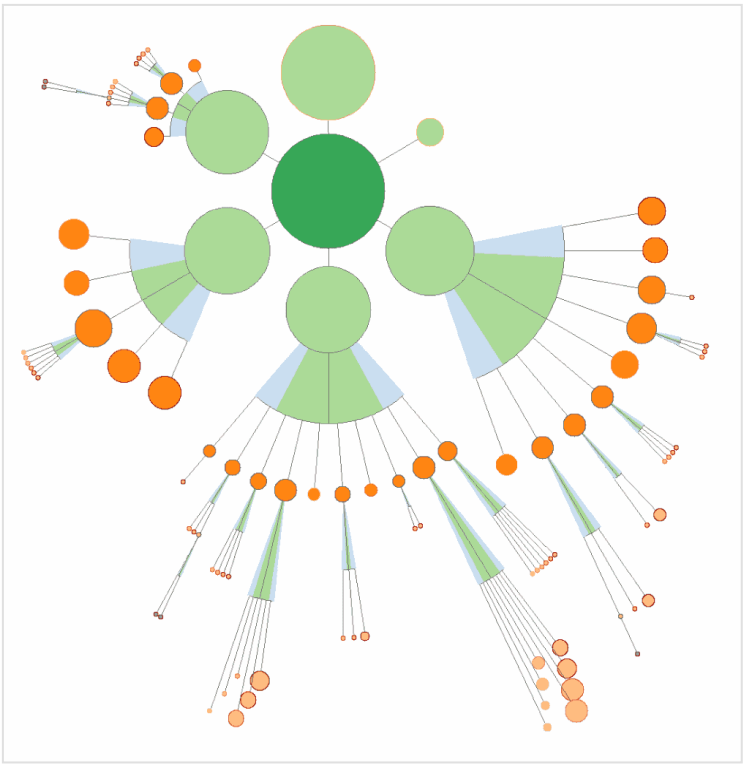

*Fig. 4 : Un exemple typique de visualisation d'une arborescence en silos.*

Si votre site est de taille conséquente, vous ne verrez clairement que les premiers niveaux. Vous pouvez alors créer d'autres vues centrées sur des catégories de votre site, en indiquant le N° de la page de départ, pour entrer dans le détail, section par section.

## *Quelques exemples de vues et d'analyses simples*

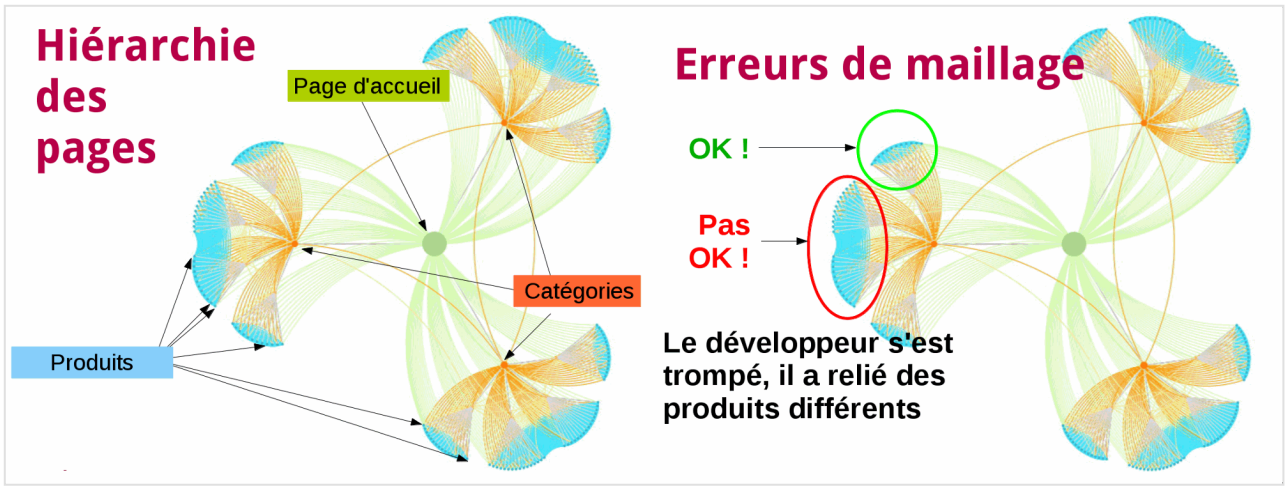


*Fig. 5 : Exemples d'analyse et détection d'erreur de maillage*

Sur la cartographie du site Fig. 5, on voit plusieurs choses : d'une part, l'arborescence des pages chère à Google est bien respectée. On distingue une hiérarchie claire avec une page centrale, en vert, suivie de pages « catégories » en orange (plus petites, donc secondaires en terme de Pagerank interne) et enfin des pages « produits » en bleu, encore plus petites, reliées en grappes, chacune dans son silo (sa catégorie).

Chacune dans son silo ? Pas tout à fait ! On voit des liens bleus qui passent d'une branche (catégorie) à une autre : ce sont des produits d'un silo qui font des liens vers le silo voisin. Ce n'était pas voulu, et cela aurait été fort difficile à détecter autrement que sous cette forme graphique.

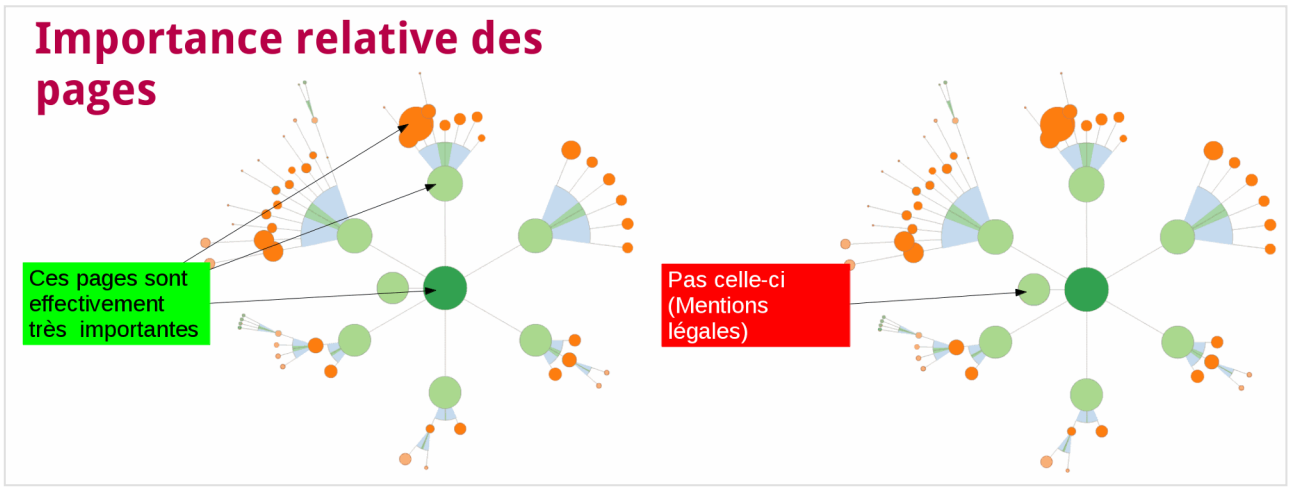


*Fig. 6 : Exemples d'analyse : Importance relative des pages*

Pour le site de la Fig. 6, on s'est concentré sur les pages importantes en termes de Pagerank interne, donc celles qui le mieux placées pour se positionner.

La page d'accueil est effectivement importante, ainsi que les grosses catégories du site (disques vert clair autour). Une page orange ressort également : C'est une page qui a un objectif et de positionnement et de conversion, c'est donc normal.

La page « Mentions légales » est elle aussi indiquée comme importante. Pourtant on ne cherche pas à me positionner sur ce mot clé, et il est dommage de « gaspiller » du Pagerank sur cette page.

En ajustant le maillage du site, il sera possible de redistribuer le PR plus efficacement.

Comme la visualisation se fait dans les mêmes conditions, on peut facilement faire un comparatif avant/après : que ce soit lors d'une refonte complète ou des mises à jour incrémentales, on voit nettement la différence.

Dans la rubrique « Historique » de l'outil, on voit les effets d'actions de maillage, d'ajout de lien sitewide, d'ajout de contenu, de suppression de pages parasites. On peut ainsi voir ce qui a changé, ce qu'il est difficile de faire de manière fiable avec Gephi.

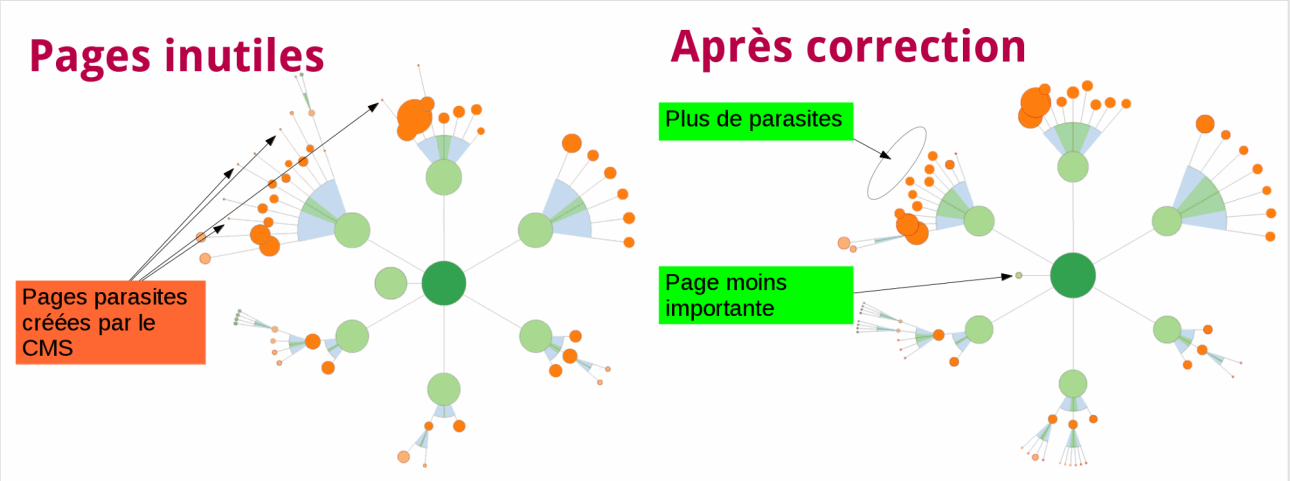


*Fig. 7 : Avant – Après : Correction du maillage et des pages inutiles*

Sur la Fig. 7, on voit le même site (il s'agit de cocon.se) avant et après correction. On voit que la page « mentions légales » a une importance moindre après correction. De même, les pages parasites crées à tort par le CMS ont pu être supprimées. On remarque également de nouvelles pages, en orange clair, qui sont apparues en bas de la cartographie, sans pour autant chambouler toute l'organisation. On peut comparer avant/après.

C'est cette stabilité et cette facilité dans les visualisations qui permet d'agir sur la structure, d'observer les changements et d'améliorer la situation par touches successives pour aller vers un site « parfait ».

## Typologie des utilisateurs de Cocon.se

Au départ, nous pensions que les clients allaient surtout être des agences web. En effet, ils connaissent les sites de leur clients, leurs objectifs, et ont les pré-requis techniques pour analyser et corriger les sites. En fait, à l'usage nous avons également comme utilisateurs de l'outil beaucoup de webmasters et de clients finaux qui travaillent eux-même et en direct sur la structure de leurs sites.

### Avantages pour les agences :

• Une photo du site, construite toujours de la même manière ;
• Pas de paramétrage compliqué, pas besoin de grosse machine ni de temps de calcul, pas de tâtonnements pour obtenir une vue ;
• Les erreurs graves sont visibles immédiatement ;
• Identification de gros leviers d'amélioration ;
• Une base concrète pour expliquer au client ce qui ne fonctionne pas. Le client est convaincu beaucoup plus facilement de la nécessité de mise à jour, d'action, quand il a vu, quand on lui a expliqué image à l'appui ce qui ne va pas ;
• Le travail effectué lors d'une refonte est visible. Sur le site lui-même, ce n'est pas forcément évident. Sur la vue, c'est flagrant : il est plus facile de valoriser l'impact de la refonte... ;)
• Plus on a l'habitude des vues, et plus vite on détecte les problèmes, plus facilement on les lit et on en extrait des infos actionnables et donc de plus en plus rentables au fil du temps ;
• Effet « whaouh » pour le client qui peut enfin « voir » son site, qui a une représentation concrète de quelque chose qui sinon, reste assez abstrait pour lui.

### Avantages pour les webmasters :

Nous avons également de très bons retours d'expérience à ce niveau. L'outil permet de ne plus travailler en aveugle. Il permet des tests beaucoup plus rapides et efficaces. Au lieu d'avancer à tâtons, on peut en *pre-prod* simuler un changement de menu, de structure, ajouter un silo, et voir, réellement, l'impact que ce changement a sur le site, sur les pages qui sont boostées (ou pas), sur la topologie du site. On sait quelle image on présente à Google, et quelle est la différence avec la version précédente.

Pour analyser ensuite les changements de positionnement, cela n'a pas de prix. On en déduit plus facilement des corrélations. Bien sur, cela suppose qu'on ait réfléchi en amont à la structure de son site, c'est là qu'on voit de suite si « ça colle » ou si on découvre une grosse divergence entre ce qu'on imagine et ce que voit un robot.

Il est conseillé de commencer à utiliser l'outil sur petits sites, pour bien appréhender la signification des vues, le parallèle entre la structure du site et comment elle transparaît dans la vue.

## Donnez du sens à votre structure de site

On nous demande souvent : « Voilà la vue de mon site, est ce que c'est bien, comment je peux améliorer ? » C'est hélas impossible à dire comme ça... Il n'existe pas une

structure magique qui fonctionne pour tout ! Ce qui fonctionne, c'est une structure, une hiérarchie logique, qui a un sens pour l'utilisateur, qui a un sens d'un point de vue conversion, et qui se retrouve dans le site tel qu'exploré par le robot. Là on a tout bon.

Ce qu'on peut voir de immédiatement sur une visualisation Cocon.Se, ce sont les gros problèmes structurels, et il certains sont très fréquents :

- Site sur maillé ;
- Mega menus ;
- Liens site-wide vers des pages inutiles (sans conversion) ;
- Tunnels infinis ;
- Erreurs de canonical, de http/https, voire de redirections ;
- Etc.

Pour le reste, il n'y a pas de réponse à l'emporte pièce : il faut avoir eu une réflexion préalable sur l'objectif du site, son organisation, et on cherchera ensuite à voir si cette organisation se retrouve effectivement dans la vue.

*- J'ai voulu isoler des sections, faire des silos cohérents, ces silos sont-ils détectés ? Sont-ils bien isolés ou fuient-ils ?*
*- Je veux « pousser » cette page sur telle thématique, le PageRank interne circule-t-il bien comme je le pense, cette page est-elle effectivement bien mise en valeur ?*

**Bref, cocon.se est un outil qui met en évidence bon nombre de points, mais qui ne se substitue pas à une réflexion de fond sur la finalité et l'organisation de son site. C'est un outil qui a besoin d'un humain et d'un cerveau aux commandes !**

Sylvain Deauré *est ingénieur de formation, et développeur à son compte depuis 1997. Il édite plusieurs services grand public et s'intéresse principalement aux défis techniques ainsi qu'aux applications pratiques de la technologie. Pour en savoir plus : https://twitter.com/SylvainDeaure.*