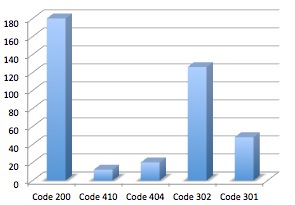
Audit technique :

**Crawler :**Outils : Xenu - <http://www.screamingfrog.co.uk/seo-spider/>

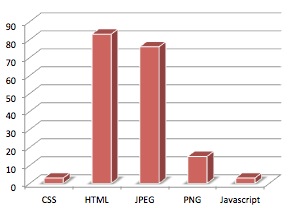
**1 – 3 crawls différents :**

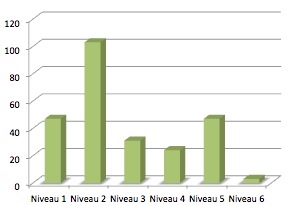
- **premier crawl sans paramétrage particulier  
-** un **second crawl en mettant en user-agent** celui du moteur de recherche qui vous intéresse (ok, on se fait passer pour Googlebot)  
- un **troisième crawl en désactivant le javascript, voire le CSS**

**2 – Les statuts des pages :**

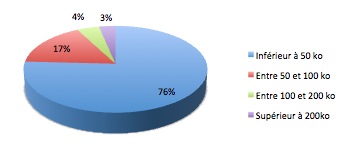


**3 – Le type de fichiers :**

****

**4 - Le niveau de profondeur des pages :  
**

**5 - La construction des URL**Sont-elles SEO friendly ? Contiennent-elles des paramètres ? Des ID de sessions ? Ont-elles une logique dans la construction ?

**6 - Le poids des fichiers  
**

**7 - Les directives Robots :**Globalement, ça se passe au niveau des balises **meta robots** (dans le code) et/ou des **X-robots-tag** (dans les en-têtes HTTP). On cherche ici à identifier des facteurs réellement bloquants comme une demande de non-indexation ou un non suivi des liens (typiquement les directives noindex et nofollow).  
  
Mais les informations ne s'arrêtent pas là, on trouve par exemple des directives concernant la mise en cache des fichiers dans les en-têtes HTTP. C'est une vraie mine d'or, apprenez à les interpréter ;-)

1. **- Les canonical**Les pages dupliquées ont-elles une URL canonique ? Cette URL existe-t-elle ? Est-elle bien accessible ?
2. **- L'arborescence du site**Les pages importantes se retrouvent-elles bien en **haut de l'arborescence** ? Sont-elles liées depuis d'autres pages importantes ? Les **pages qui convertissent sont-elles facilement accessibles** ? etc.  
   **Maillage interne** : Mise en place des liaisons verticalement, mais également horizontalement ?   
   **Pagination :** Est-elle correctement mise en place ? Est-elle optimale ? Favorise-t-elle une réduction du niveau de profondeur de certaines pages ? Les balises rel="prev" et rel="next" sont-elles bien en place ?
3. **– Le robots.txt**

Existe-t-il ? Comporte-t-il des restrictions sur des dossiers ou des pages importants Au besoin, vous pouvez relire [mon article sur le robots.txt](http://noviseo.fr/2013/01/larticle-ultime-sur-le-robots-txt/).  
Google Webmaster Tools vous permettra de tester le robots.txt

1. **- Le Sitemap.xml**Existe-t-il ? Y a-t-il des pages orphelines ? Est-il mis à jour ? Existe-t-il une page listée dans le sitemap et qui est en même temps bloquée au crawl par le robots.txt ?
2. **- Le fichier .htaccess :**

Celui-ci vous informera sur les directives données aux robots (comme vu plus haut, mais tout est sous les yeux), les informations sur la mise en cache (idem), sur la réécriture d'URL, etc. Pour en savoir un peu plus sur le fichier htaccess, je vous recommande de lire [cet article](http://openweb.eu.org/articles/le-fichier-htaccess).

1. **Les performances du site :**Si le poids des fichiers et le .htaccess nous ont déjà renseignés sur ce point là, il peut être intéressant d'aller plus loin. Pour ça, plusieurs outils sont à notre disposition, et certains nous donnent même des conseils pour améliorer les performances :

* Vitesse du site dans Google Analytics
* [Pingdom Tools](http://tools.pingdom.com/) : <http://tools.pingdom.com/fpt/>
* [Page Speed Insight](https://developers.google.com/speed/pagespeed/insights) : <http://developers.google.com/speed/pagespeed/insights/>
* [Gtmetrix](http://gtmetrix.com/)
* L'extension Yslow pour Firebug

1. **Le reste :**

* l'âge du domaine (avec n'importe quel service de WhoIs)
* le voisinage du domaine (avec [YouGetSignal](http://www.yougetsignal.com/" \t "_blank)), sa localisation géographique
* le type d'hébergement (dédié ou mutualisé)
* le nombre de liens par pages (pas trop de liens internes ou externes, tout de même quelques liens externes)
* la présence de liens en nofollow (à éviter s'ils pointent vers des pages importantes, à méditer s'ils pointent vers des pages peu importantes)
* l'éventuelle duplication du domaine (le www est-il bien redirigé vers la version sans www, ou vice-versa ?)
* la présence d'éléments cachés en CSS (à éviter bien évidemment !)

Audit de contenu

Ça y est, vous avez les bases en ce qui concerne la partie technique d'un audit de référencement, maintenant il va falloir s'attaquer à la partie contenu. Vous avez trouvé le travail sur l'aspect technique fastidieux ? Réjouissez-vous, ici ça va être encore pire !

Pire parce qu'il est quasi impossible d'automatiser une analyse de contenu car souvent les ressources vous feront défaut ; pire parce que cette même analyse est purement subjective, contrairement à tous les autres aspects de l'audit SEO ; pire parce qu'il va falloir presque tout faire à la main ; pire parce que plus il y a de pages, plus vous aller y passer du temps. Vous êtes quand même prêt à continuer ? C'est parti !

1. **Repérage du contenu dupliqué**

Idéalement, votre crawler a repéré la zone où se trouve le contenu effectif des pages (ie le contenu sans l'en-tête, les sidebars et le footer), il ne vous restera donc plus qu'à filtrer les données récupérées et vérifier la présence de doublons.

S'il n'est pas aussi puissant, notre crawler a tout de même récupéré les balises Title, Meta description, H1 et H2. Screaming Frog le fait, d'ailleurs ;-) On va pouvoir commencer par filtrer ces données pour repérer les doublons, au moins pour la balise Title. Si on trouve des Title doubles, c'est mal, et si après vérification on voit que les pages ont le même contenu, c'est encore pire !

Google Webmaster Tools nous indique les balises Title et Description dupliquées dans la catégrie Optimisations ==> Améliorations HTML.

Une vérification "à la scout" sera toujours intéressante : vous sélectionnez quelques pages, vous copiez une ligne du contenu au hasard et vous allez la coller dans Google sous la forme suivante :

"ligne que vous venez de copier" site:site-audité.fr

Vous saurez ainsi si Google connait plusieurs pages de votre site ayant le même contenu. Évidemment vous n'allez pas le faire sur toutes les pages du site mais simplement sur une sélection de différentes typologies de pages.

Concernant le contenu dupliqué sur d'autre sites, c'est un peu la même façon de faire, la partie "site:" en moins. Plus simplement, vous pouvez également demander au client comment sont gérés ses contenus : les fiches produits sont-elles des copies des fiches du fabricant ? Il saura vous répondre, et devra le faire. Hey ! Vous bossez **avec** lui, pas uniquement **pour** lui !

1. **La qualité du contenu**

Sujet brûlant tant il dépend de la stratégie et donc des objectifs, la qualité est tout de même importante. Même si globalement ça devrait aller assez vite sur ce point, on commence par ce qui va plaire à Mme Duchez (votre prof de français de troisième), ses dissertations et son Bescherelle :

* vérification de l'orthographe
* vérification de la grammaire
* vérification de la valeur ajoutée du texte

Une fois ces vérifications effectuées, il va falloir comprendre si le champ lexical est optimal : vise-t-il les meilleures expressions ? Les expressions visées sont-elles réalistes (ie : le positionnement sur ces expressions est-il réalisable, en prenant en compte le background du site) ? Sont-elles génératrices de trafic ? Et surtout, le plus important, sont-elles adaptées à la cible ?

1. **Keyword stuffing et "keyword missing"**

Le keyword stuffing est la répétition d'un mot clé, mais de façon bien trop importante. C'est une pratique que les moteurs repèrent et sanctionnent à plus ou moins forte échelle, c'est donc une pratique à éviter. Il est donc important de se poser la question si le site est sujet ou non au keyword stuffing :

* plusieurs occurrences de l'expression dans le Title
* utilisation de l'expression dans l'URL
* utilisation de l'expression dans le H1 et chacun des H2
* répétition trop importante de l'expression dans le texte
* mise en place de l'expression dans l'attribut alternatif des images
* etc.

À l'inverse on retrouve ce qu'on pourrait appeler le "keyword missing". Pour faire simple, c'est le fait de vouloir se placer sur une expression sans l'utiliser dans la page, ou en l'utilisant trop peu. Au final, on reprend à peu près les mêmes points que précédemment mais on recherche l'effet inverse :

* l'expression n'est pas utilisée dans le Title
* l'expression (ou une de ses variantes) n'est pas utilisée dans le H1
* l'expression n'est jamais citée dans les textes
* etc.

Mais comment repérer tout ça ? Quel est le taux idéal d'occurrences d'un mot clé dans une page ? Honnêtement, je n'en ai aucune idée, je m'en contrefous et je ne suis même pas sûr qu'il y ait un "taux idéal" quelque part dans l'algorithme de Google. La façon la plus simple de repérer un de ces deux extrêmes est de lire le texte. **Si ce dernier utilise une expression si souvent qu'il en devient indigeste, alors c'est du keyword stuffing**. Si vous n'arrivez pas à résumer le sujet d'une page en deux ou trois mots, ou que l'expression qui semblerait la plus pertinente n'apparait pas, alors c'est du keyword missing.

1. **La taille des textes**

Là encore on est vraiment dans le "très subjectif" et ici aussi, je serais tout simplement incapable de définir un volume de texte optimal. Lorsque vous auditez les textes d'un site, le plus pertinent est de se mettre dans la peau d'un internaute : se sent-il lésé concernant l'information partagée ou au contraire se sent-il étouffé par une quantité de textes si énorme qu'elle en devient imbuvable ? Dans ce deuxième cas, le texte ne mérite-t-il pas d'être divisé en plusieurs pages ?

Google veut proposer quelque chose qui plait à l'internaute. Si à la première vision un texte ne vous plait pas, il en sera autant pour tous les autres internautes.

1. **Les autres types de contenu**

Si le texte est le seul type de contenu que Google sait lire, c'est loin d'être le seul contenu disponible sur internet. Les autres contenus sont pourtant bien susceptibles d'être présents sur le site audité. Quid des images ? L'attribut alternatif est-il renseigné ? Le nom du fichier est-il descriptif ? Un texte explicatif les accompagne-t-il ? Même chose pour les vidéos, les sons, les présentations, les formulaires, etc. : sont-ils accompagnés d'un environnement susceptible de les définir ?



Nous venons de terminer un bon tour du propriétaire après avoir réalisé la partie technique et la partie contenu de l'audit de référencement. Vous pensez en avoir fini ? Non, bien sûr que non ! Ça tombe alors très bien puisque vous pourrez revenir demain pour vous attaquer à la partie statistique de l'audit ;-)