



---

## Exemple de rapport d'audit SEO par RM Tech

---

Ceci est un exemple de rapport d'audit de référencement, généré par l'outil **Audit RM Tech** de la plateforme My Ranking Metrics.

Il s'agit d'un rapport d'audit réel qui a été anonymisé (certains liens ne sont pas actifs pour cette version). Il est très complet avec + de 50 pages !

- Vous êtes **prestataire SEO** (agence, freelance, consultant indépendant) ? Gagnez du temps, obtenez directement un rapport de ce type et gagnez en rentabilité.
- Vous êtes **SEO in-house** (chez l'annonceur ou pour vous-même) ? Gagnez aussi du temps, priorisez les problèmes, obtenez des recommandations détaillées.

Quel que soit votre profil, ce modèle de rapport d'audit de référencement vous sera très utile.

My Ranking Metrics est une plateforme SEO française mise au point par des experts reconnus (Olivier Duffez avec [WebRankInfo](#), Fabien Faceries avec [Ranking Metrics](#)), utilisée par + de 2100 clients...

Pour en savoir plus (présentation, prix...) : [myrankingmetrics.com](https://myrankingmetrics.com)

# 1 Introduction

## 1.1 Objectif de ce rapport

Ce document a pour objectif de vous présenter les résultats de l'audit SEO de votre site. Vous y trouverez une série de recommandations avec de nombreuses explications, ainsi que des annexes contenant la liste exhaustive de tous les éléments à améliorer sur votre site.

L'audit de votre site a porté sur **10 269 URL** ainsi que sur **25 560 images distinctes** (25 131 images sont hébergées sur example.com et 611 sur d'autres sites).

### Score RM Tech



Très forte marge de progression potentielle !

Meilleure note possible : 100

### Score Pages Actives



Très forte marge de progression potentielle !

Meilleure note possible : 100

Toutes les analyses qui en découlent doivent donc être considérées en fonction de l'état du site ce jour-là. Voici les conditions de l'analyse :

- Site analysé : example.com
- URL de départ : <https://www.example.com/>
- Protocoles : aucune restriction (HTTP et HTTPS)
- Nombre maximal d'URL à explorer : 25 000
- Motif d'inclusion : aucun
- Motif d'exclusion : .jpg OR .jpeg OR .png OR .gif
- User-Agent pour le site : Mozilla/5.0 (Windows NT 6.1; WOW64; rv:64.0) Gecko/20100101 Firefox/64.0
- User-Agent pour le robots.txt : Googlebot
- Vitesse de l'exploration : 1.0 requête par seconde

- Ignorer le robots.txt : non
- Prendre en compte les liens externes ayant un attribut rel="nofollow" : non
- Limiter l'exploration au sous-domaine www.example.com : non
- CMS déclarés : aucun
- Analyse des images : oui
  - Nb maximal d'images à explorer : 60 000
  - Seuil de poids pour considérer une image comme trop lourde : 100 Ko
  - Seuil de poids des images d'une page considérée comme trop lourde : 1 000 Ko
  - Nombre d'images maximal acceptable par page : 50
  - Explorer les images externes : oui
- Données Google Analytics : compte "example.com", propriété "https://www.example.com", vue "www.example.com", trafic "organic" uniquement, sur les 365 derniers jours
- Données Google Search Console : compte "contact@example.com", propriété https://www.example.com/, sur tous types d'appareil, sur les 365 derniers jours

Vous trouverez [à la fin de ce rapport](#) des explications techniques sur le fonctionnement de notre outil ainsi que des définitions des termes utilisés dans le jargon du référencement naturel (SEO).

Les données sont disponibles sous forme exhaustive dans des fichiers annexes au format Excel ainsi qu'au format TSV (Tab-separated values) que vous pouvez facilement ouvrir dans Excel ou tout autre tableur.

Vous pouvez télécharger l'ensemble de ces fichiers annexes dans un seul fichier au format ZIP en cliquant sur le bouton ci-dessous :

#### Annexes

Les annexes sont disponibles dans le fichier ZIP annexes-example\_com.zip.

## 1.2 Comment exploiter cet audit

Bien entendu, cet audit couvre tous les points essentiels d'un audit essentiellement technique : erreurs dans les liens (404...), URL canonique et balise meta robots noindex, temps de téléchargement, profondeur et maillage, balises title, meta description, H1-H6 ainsi que la taille du contenu dans la zone principale de la page. Le rapport explique tout et les annexes contiennent tout ce qu'il faut pour corriger.

Mais l'audit RM Tech permet d'être bien plus efficace grâce à des algorithmes exclusifs développés pour tirer le maximum du croisement entre les données de crawl et celles de Google Analytics et Search Console :

- analyse des pages actives et inactives sur 1 an
- calcul de l'indice zombie de chaque page
- identification des pages les plus rentables à travailler
- identification des pages créant le plus d'insatisfaction chez les utilisateurs
- identification des pages pouvant passer de page 2 à page 1 de Google

Voici ce que vous devriez faire, dans l'ordre :

1. s'il y a trop d'erreurs de crawl (3xx, 4xx, 5xx), corrigez-les
2. allez directement voir la répartition de vos pages indexables selon leur performance SEO (nb de sessions générées sur 1 an)
3. si vous avez des pages à fort indice zombie, traitez-les en priorité en suivant les explications
4. exploitez la liste des pages les plus rentables à travailler en 1er et celles à fort potentiel pour passer en page 1 de Google

Pour d'autres conseils sur les phases à enchaîner, consultez [cet article](#).

## 2 Synthèse des résultats

### 2.1 Scores RM Tech et Pages actives

Le score RM Tech pour cet audit est de **65,7** sur 100. Il représente le niveau d'optimisation du référencement naturel en ce qui concerne les éléments analysés sur le site (techniques, contenus, liens internes, qualité...).



Le score Pages Actives pour cet audit est de **21,4** sur 100. Il indique la performance globale des pages indexables trouvées par RM Tech (en fonction du nombre de sessions générées par chaque page selon Google Analytics, sur les 365 jours étudiés). Aucun échantillonnage n'a été effectué par Google Analytics.



Pour ces 2 scores, nous recommandons d'atteindre 95 (score en vert) ; entre 70 et 95 le score est en jaune et en-dessous de 70 il est en rouge.

### 2.2 Pages zombies

D'après notre algorithme Pages Zombies d'identification des pages irrécupérables, l'indice zombie moyen est de **40,6**. [Voyez les détails ici](#).

### 2.3 Pages à travailler en premier

Sur la période étudiée et avec les paramètres de crawl du site rappelés dans l'introduction, **1 599 785** sessions ont été générées par **5 331 pages** (pages indexables autres que l'URL de départ) via le référencement naturel.

Grâce à notre algorithme exclusif combinant des dizaines de critères SEO (détaillés dans ce rapport) et des données disponibles dans Google Analytics, vous trouverez ci-après la [liste des 100 pages les plus rentables à travailler en premier](#) (pour l'amélioration de votre référencement naturel), avec pour chacune des indications précises sur le travail à accomplir.

### 2.4 QualityRisk

D'après notre algorithme QualityRisk d'évaluation de la qualité, le risque moyen est de **38,3** sur l'ensemble des pages indexables (le pire possible étant 100).









En outre, **84,74 % des pages indexables sont de trop faible qualité** (elles ont un QualityRisk  $\geq$  20).

[Voyez les détails ici](#).

Élément analysé	Optimisation	Erreurs constatées sur les pages	
		en pourcentage	en nombre de pages
QualityRisk (évaluation de la qualité)		84,74 % <sup>(3)</sup>	6 538

### 2.5 Taux d'erreur

Ce tableau regroupe en 8 familles les dizaines d'éléments audités sur votre site :

Élément analysé	Optimisation	Erreurs constatées sur les pages	
		en pourcentage	en nombre de pages
URL répondant en erreur		1,65 % <sup>(1)</sup>	169
Indexabilité des pages HTML		23,61 % <sup>(2)</sup>	2 385
Temps de téléchargement		66,60 % <sup>(2)</sup>	6 727
Profondeur et maillage interne		92,22 % <sup>(3)</sup>	7 115
Balise title		0,43 % <sup>(3)</sup>	33
Balise meta description		1,17 % <sup>(3)</sup>	90
Contenu des pages		65,12 % <sup>(3)</sup>	5 024
Liens sortants externes		48,63 % <sup>(4)</sup>	2 996

- (1) pourcentage calculé par rapport au nombre d'URL explorées, soit ici 10 269 URL
- (2) pourcentage calculé par rapport au nombre d'URL de pages HTML sans erreur (code 200), soit ici 10 100 URL
- (3) pourcentage calculé par rapport au nombre d'URL de pages HTML indexables, soit ici 7 715 URL
- (4) pourcentage calculé par rapport au nombre d'URL externes distinctes trouvées dans les liens sortants follow, soit ici 6 161 URL

Explications sur le code couleur :

- vert = succès : moins de 5% des pages sont en erreur
- jaune = avertissement : entre 5% et 30% des pages sont en erreur
- rouge = erreur : plus de 30% des pages sont en erreur

Ce tableau de synthèse est basé uniquement sur ce que nous considérons être des **erreurs**, mais vous verrez que par endroits dans ce rapport nous listons également des **avertissements** : il s'agit de points à vérifier ou d'optimisations moins importantes à mettre en place. Les taux d'erreurs sont en gras (mais pas ceux concernant les avertissements).

Certaines cases de nos tableaux contiennent le libellé "n/a" ce qui signifie "non applicable" pour indiquer qu'il est normal que la case ne contienne pas de valeur.

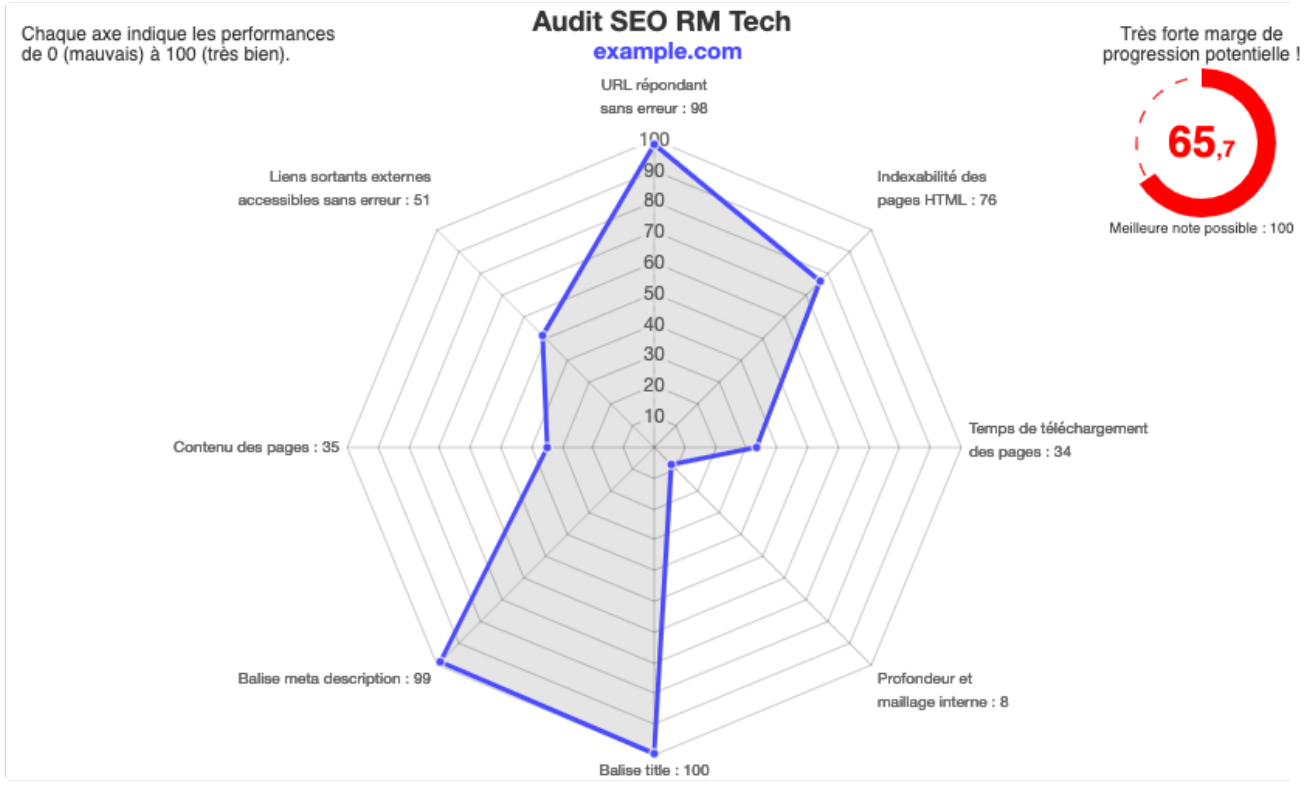
## 2.6 Performance

Les résultats présentés ci-dessus (et dans le reste du rapport) sont exprimés en termes de taux d'erreur. Nous vous proposons une autre façon d'analyser les résultats, de façon très visuelle, en calculant des scores de performances (en gros, c'est l'inverse du taux d'erreur).

Pour chaque famille de critères, le score de performances d'un site est égal au nombre de pages (ou de liens) SANS erreur divisé par le nombre de pages (ou de liens) étudié(e)s, exprimé en pourcentage. Par exemple, concernant le temps de téléchargement des pages :

- un taux d'erreur de 12 % signifie que 12 % des pages étudiées sont trop lentes à télécharger
- dans ce cas, le score de performances est de 88 % (= 100 - 12)

Le graphique en radar ci-dessous présente une synthèse des performances du site pour chaque famille de critères.



## 3 Crawl et indexabilité

Voici la méthode suivie par RM Tech pour l'analyse du crawl et de l'indexabilité :

1. d'abord, nous faisons le point sur les (éventuelles) URL qui n'ont pas été crawlées en raison d'une interdiction figurant dans un fichier robots.txt
2. ensuite, nous listons l'ensemble des URL crawlées (qu'elles soient accessibles ou pas) ; leur nombre correspond au nombre de crédits consommés par cet audit
3. enfin, parmi les pages HTML accessibles, nous étudions celles qui sont indexables (nous ne vérifions pas si Google les a indexées, mais aucun critère technique n'empêche qu'elles le soient)

Vous trouverez [dans la conclusion](#) une annexe qui liste toutes les URL crawlées avec pour chacune la raison de ce crawl (la plupart du temps il s'agit d'un lien mais ça peut aussi être une URL canonique ou une redirection par exemple).

### 3.1 Bilan des URL bloquées par le(s) fichier(s) robots.txt

RM Tech respecte les directives indiquées dans les fichiers robots.txt (il doit y en avoir un pour chaque sous-domaine), concernant Googlebot (le crawler de Google).

Nous vous conseillons de mettre en ligne un fichier robots.txt pour chaque sous-domaine, même s'il est vide. Ceci évite de générer des erreurs 404 à chaque fois qu'un robot cherche à y accéder. C'est la raison pour laquelle nous indiquons le code HTTP du fichier robots.txt de chaque sous-domaine crawlé (ce code doit donc être 200).

Voici la répartition des URL bloquées par un fichier robots.txt et qui reçoivent pourtant au moins un lien interne follow (ou qui sont indiquées dans une URL canonique) :

Fichier robots.txt	Code HTTP	Nombre d'URL bloquées
<a href="https://www.example.com/robots.txt">https://www.example.com/robots.txt</a>	200	37
<a href="http://www.example.com/robots.txt">http://www.example.com/robots.txt</a>	200	0
<b>Total</b>		37

Bien entendu, aucune URL stratégique pour votre business ne doit être bloquée par un fichier robots.txt ! Notre analyse et surtout l'annexe ci-dessous vous permettent de le vérifier.

#### Annexe

Le fichier annexe [example\\_com-1\\_url-bloquees-robots-txt.xlsx](#) contient la liste des URL bloquées au crawl.

#### Annexe

Le fichier annexe [example\\_com-robotstxt.zip](#) contient l'ensemble des fichiers robots.txt récupérés par RM Tech.

### 3.2 Nombre total d'URL crawlées

A partir d'ici et pour tout le reste de ce rapport, nous ne tenons compte que des URL autorisées au crawl (non bloquées dans un fichier robots.txt).

Dans votre cas, nous avons pu analyser 100% des URL de votre site (trouvables par des liens depuis la page de départ). Le nombre d'URL crawlées a été de **10 269**.

Voici la répartition selon les protocoles HTTP et HTTPS :

Protocole	Nb d'URL	% des URL crawlées
HTTP	88	0,86 %
HTTPS	10 181	99,14 %

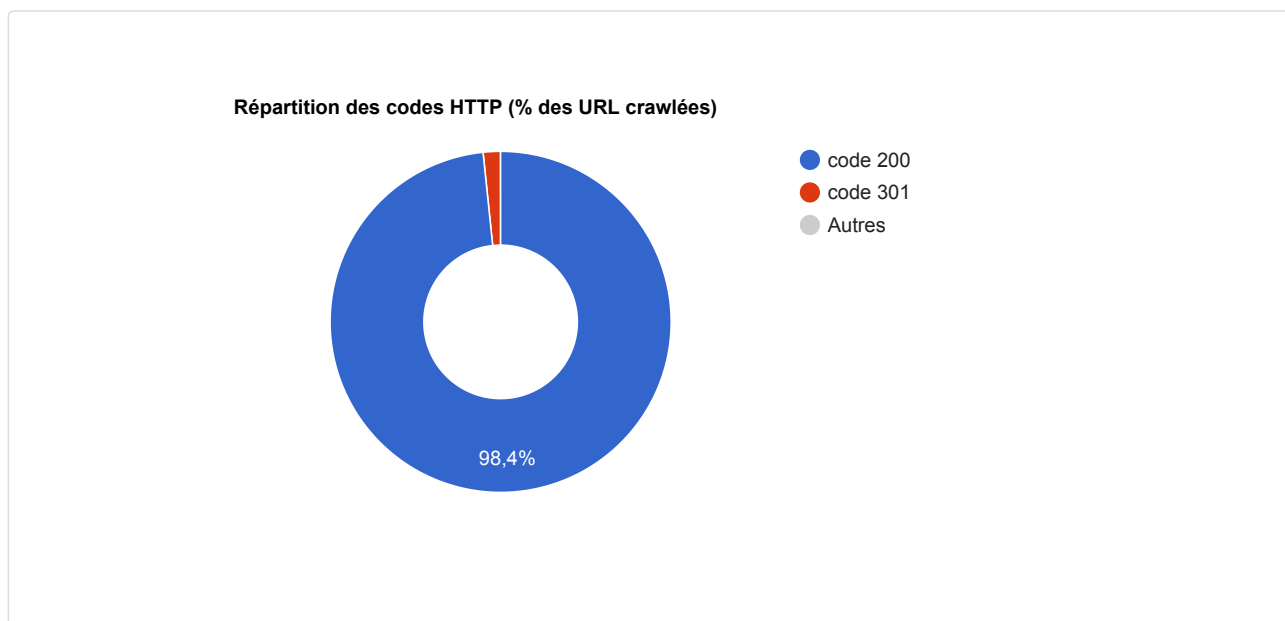
### 3.3 État des URL crawlées

Voici le bilan des codes HTTP rencontrés sur votre site (lisez leur signification [en fin de rapport](#)) :

Code HTTP	Nb d'URL en HTTP	Nb d'URL en HTTPS	Nb d'URL (HTTP et HTTPS)	% des URL crawlées
200	0	10 100	10 100	98,35 %
301	88	80	168	1,64 %
404	0	1	1	0,01 %
<b>Total erreurs 3xx</b>	88	80	168	1,64 %
<b>Total erreurs 4xx</b>	0	1	1	0,01 %
<b>Total erreurs</b>	88	81	169	<b>1,65 %</b>

Remarque : même si les redirections ne sont pas à proprement parler des erreurs techniques, c'est pénalisant pour le référencement. C'est pourquoi nous les avons incluses dans le calcul des erreurs. Nous vous conseillons de faire des liens directs (directement vers la page finale) au lieu de passer par une ou des redirection(s) qui sont des étapes intermédiaires.

Voici ce que cela donne sous forme graphique :



#### Annexe

Le fichier annexe `exemple_com-3_codes-http.xlsx` contient les données associées.



**Annexe**

Afin de vous permettre de localiser les erreurs 4xx et 5xx sur le site, le fichier annexe `exemple_com-3_liens-vers-erreurs-4xx-5xx.xlsx` contient les URL des pages contenant les liens en erreur.

**Annexe**

Afin de vous permettre de localiser les erreurs 3xx sur le site, le fichier annexe `exemple_com-3_liens-vers-erreurs-3xx.xlsx` contient les URL des pages contenant les liens en erreur.

### 3.4 Types MIME

Voici la liste des types MIME (lisez leur signification [en fin de rapport](#)) rencontrés pour les pages répondant correctement (code HTTP 200) :

Type MIME	Nb d'URL	% des URL crawlées en code 200
text/html	10 100	100 %

**Annexe**

Le fichier annexe `exemple_com-4_types-mime.xlsx` contient les données associées.

### 3.5 Pages HTML crawlées sans erreur

Le référencement naturel se travaille essentiellement avec les pages HTML. Nous avons donc regroupé dans un fichier TSV les principales informations disponibles pour l'ensemble de vos pages HTML crawlées sans erreur (code HTTP 200) :

- URL
- Title
- Meta description
- Balise d'URL canonique
- URL canonique via en-tête HTTP
- Meta robots
- En-tête HTTP X-Robots-Tag

**Annexe**

Le fichier annexe `exemple_com-5_pages-sans-erreur.xlsx` contient les données associées.

### 3.6 Pages HTML indexables par les moteurs

Le référencement naturel se travaillant essentiellement pour les pages HTML, notre analyse porte ici exclusivement sur les pages HTML.

Nous souhaitons étudier l'indexabilité de vos pages, c'est-à-dire la possibilité qu'elles soient indexées. On parle donc de pages « indexables ».

Les pages HTML indexables sont toutes celles :

- qui renvoient un code 200
- et qui ne sont pas interdites d'indexation (via une balise meta robots noindex ou none, ou bien un en-tête HTTP X-Robots-Tag)
- et qui ne contiennent pas d'URL canonique différente de l'URL crawlée

Voici le décompte pour votre site :

	Nb d'URL	% des URL HTML code 200
<b>Pages HTML avec code HTTP 200 (A)</b>	10 100	
<b>Dont pages interdites d'indexation</b>	2 385	23,61 %
<b>Dont pages avec URL canonique différente de l'URL crawlée</b>	0	0 %
<b>Nb total de pages non-indexables (B)</b>	2 385	23,61 %
<b>Nombre de pages HTML indexables (A - B)</b>	7 715	76,39 %

Sauf cas particuliers dont vous connaissez bien l'existence sur votre propre site, le pourcentage de pages indexables devrait être élevé (proche de 100%).

En effet, il n'est pas efficace de laisser de nombreuses pages se faire crawler alors que vous exigez qu'elles ne soient pas indexées.

Dans votre cas, nous avons identifié 2 385 pages HTML interdites d'indexation représentant 23,61 % des pages HTML en code 200. Nous vous conseillons de vérifier si c'est normal d'interdire l'indexation de ces pages. Vous trouverez la liste des URL concernées dans les annexes.

#### Annexe

Le fichier annexe example\_com-6\_indexabilite.xlsx contient les données associées.

#### Annexe

Afin de vous permettre de localiser les erreurs liées à une directive noindex (balise meta ou X-Robots-Tag) sur le site, le fichier annexe example\_com-6\_liens-vers-noindex.xlsx contient les URL des pages contenant les liens en erreur.

## 4 Temps de téléchargement

Plus vos pages mettent de temps à être téléchargées, plus c'est pénalisant :

- pour les internautes, car le temps de chargement complet de la page risque d'être long ;
- pour les robots comme Googlebot, car ils crawlent moins de pages par jour si elles sont lentes

Nous vous conseillons donc d'optimiser la conception de votre site, le contenu de vos pages ainsi que votre serveur d'hébergement afin d'avoir des pages qui se téléchargent rapidement. Dans cette partie, nous étudions le temps de téléchargement de chaque page HTML répondant sans erreur (code HTTP 200). Selon notre expérience, il faut que chaque page soit téléchargée en 600 ms maximum.

Voici les résultats de notre analyse :

	Nb d'URL	% des URL HTML code 200
<b>Pages trop lentes (&gt; 600 ms)</b>	6 722	66,60 %
<b>dont : pages vraiment trop lentes (&gt; 1000 ms)</b>	4 548	45,03 %
<b>Nb de pages HTML code 200</b>	10 100	
<b>Taux d'erreur</b>		<b>66,60 %</b>

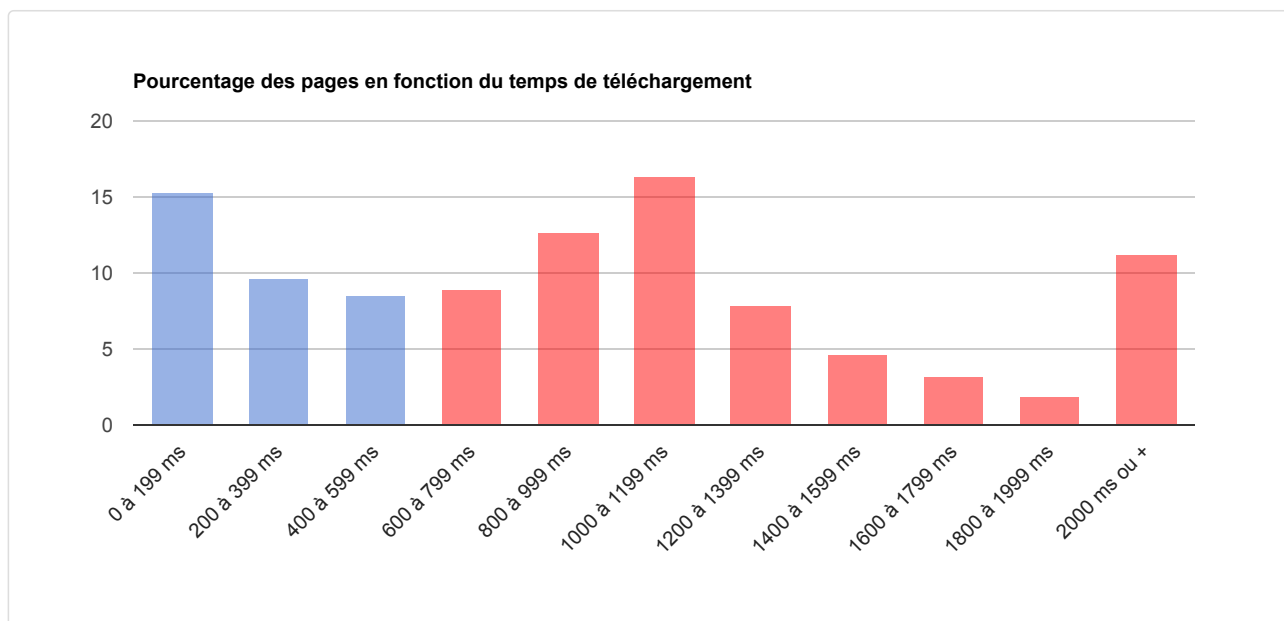
Remarque : **il ne s'agit pas ici du temps de chargement complet de la page**, puisque nous ignorons tous les fichiers externes pour ce calcul (Javascript, CSS, images, etc.). Il s'agit du temps mis par le serveur pour répondre à la requête, générer la page HTML et la télécharger. C'est donc le même principe que le temps indiqué dans Google Search Console (ex Webmaster Tools) rubrique Exploration > Statistiques sur l'exploration > Temps de téléchargement d'une page (en millisecondes). Cependant, notre calcul ne tient compte que des fichiers HTML alors que les données fournies par la Search Console incluent également les CSS, images, PDF et autres documents crawlés par Googlebot.

En complément du temps de téléchargement de la page (code HTML), **le poids et les dimensions de vos images** ont été étudiés dans [l'analyse des images ci-après](#).

Voici la répartition par tranches de 200 ms (pour les pages HTML code 200) :

Temps de téléchargement (ms)	Nb de pages	Pourcentage des pages
de 0 ms à 199 ms	1 544	15,29 %
de 200 ms à 399 ms	974	9,64 %
de 400 ms à 599 ms	855	8,47 %
de 600 ms à 799 ms	901	8,92 %
de 800 ms à 999 ms	1 278	12,65 %
de 1000 ms à 1199 ms	1 644	16,28 %
de 1200 ms à 1399 ms	792	7,84 %
de 1400 ms à 1599 ms	465	4,60 %
de 1600 ms à 1799 ms	324	3,21 %
de 1800 ms à 1999 ms	188	1,86 %
2000 ms ou +	1135	11,24 %

Voici ce que cela donne sous forme graphique :



En plus du temps de téléchargement d'une page, il est intéressant d'étudier aussi le poids du code HTML brut (en Ko). Attention, ceci n'inclut aucune ressource de la page : ni les images, ni les javascript, CSS et autres.

Même si ce n'est pas toujours problématique, il est déconseillé d'avoir un poids trop important. Google crawle des documents très volumineux (plusieurs Mo) mais Bing semble s'arrêter autour de 700 Ko (ce qui est déjà très gros).

C'est également la limite utilisée dans cet audit, si bien que si certaines pages dépassent ce seuil, une partie seulement de leur contenu est prise en compte. Dans ce cas, il est possible que des liens présents dans la page (vers d'autres pages internes) ne soient pas pris en compte par cet audit.

L'annexe ci-dessous indique le poids du code HTML brut de chaque page, ainsi qu'un indicateur qui précise s'il dépasse 700 Ko.

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
<b>Nb de pages dont le poids du code HTML brut dépasse 700 Ko</b>	n/a	0	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>		0 %	

#### Annexe

Le fichier annexe exemple\_com\_04\_temps-telechargement.xlsx contient les données associées.

## 5 Profondeur des pages et maillage interne

### 5.1 Analyse de la profondeur des pages

Pour qu'une page soit efficace en référencement naturel, il faut éviter qu'elle soit trop profonde dans le site, la profondeur étant mesurée par le nombre de clics nécessaires (sur des balises <a href>) pour l'atteindre depuis la page d'accueil.

Dans votre cas, notre algorithme a évalué que **la profondeur maximale recommandée est de 4**. Concrètement, cela signifie que nous conseillons que la majorité des URL soit à une profondeur inférieure ou égale à 4. Plus vous aurez d'URL situées à des profondeurs supérieures, plus vous aurez des problèmes de crawl, d'indexation et même de positionnement.

Remarque importante : attention, nous ne conseillons pas de réduire artificiellement la profondeur des pages uniquement pour ne pas dépasser la profondeur recommandée. Nous conseillons de privilégier les liens utiles à l'internaute, c'est-à-dire situés à des endroits faciles à identifier, entre des pages de thématique proche. Par exemple, c'est une mauvaise solution que de mettre des liens vers toutes les pages du site dans une page plan du site, elle-même reliée depuis le pied de page. Il n'est pas toujours possible ou facile d'éviter que certaines pages soient trop profondes, par contre il faut s'assurer que les pages stratégiques sont très faciles d'accès (profondeur 1 si possible).

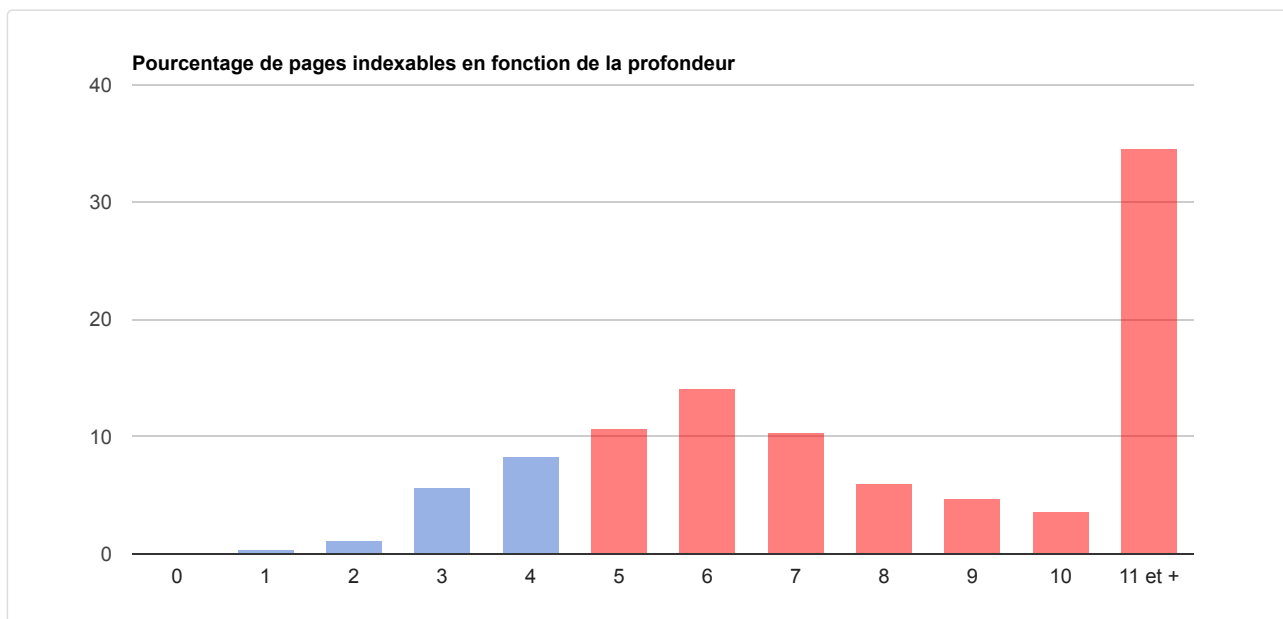
L'étude a porté sur les pages HTML indexables.

Voici les résultats selon les niveaux :

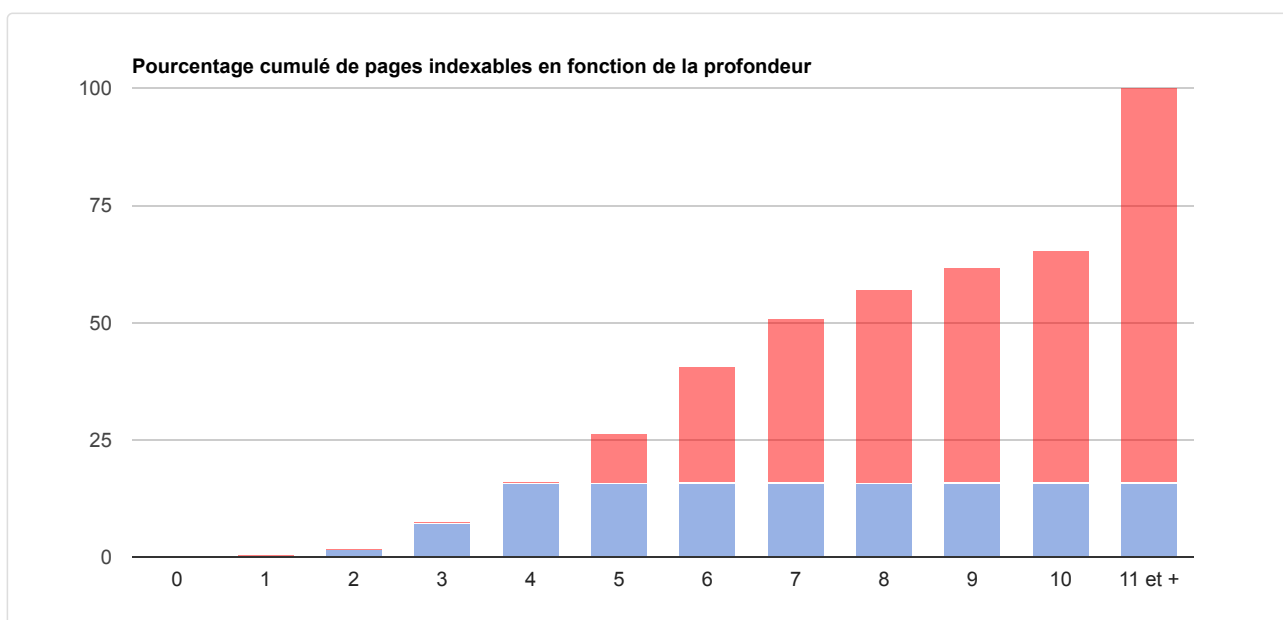
Profondeur	Nb de pages	Nb cumulé de pages	Pourcentage des pages	Pourcentage cumulé
0	1	1	0,01 %	0,01 %
1	34	35	0,44 %	0,45 %
2	96	131	1,24 %	1,70 %
3	437	568	5,66 %	7,36 %
4	643	1 211	8,33 %	<b>15,70 %</b>
5	822	2 033	10,65 %	26,35 %
6	1 090	3 123	14,13 %	40,48 %
7	809	3 932	10,49 %	50,97 %
8	468	4 400	6,07 %	57,03 %
9	366	4 766	4,74 %	61,78 %
10	282	5 048	3,66 %	65,43 %
11 et +	2 667	7 715	34,57 %	100 %

Dans votre cas, nous avons identifié **6 504 pages trop profondes**, soit un **taux d'erreur de 84,30 %**. Vous devriez corriger ceci car les pages trop profondes sont plus difficilement consultées par les internautes et ont des moins bonnes performances en termes de référencement. Cette erreur peut s'expliquer par une mauvaise structure du site et/ou un mauvais maillage interne.

Le graphique ci-dessous reprend les éléments du tableau précédent :



Le graphique ci-dessous montre le pourcentage cumulé :



## 5.2 Analyse du maillage interne

Pour qu'une page soit efficace en référencement naturel, il faut qu'elle reçoive des liens (on appelle ça des backlinks ou liens entrants). Ils peuvent venir d'autres pages du site (backlinks internes) ou pas (backlinks externes).

**Le minimum est d'avoir 1 backlink interne** (sinon comment Google l'indexerait ? avec les sitemaps ce n'est pas assez efficace pour un bon positionnement Google). Mais c'est tout de même préférable d'en avoir plusieurs, surtout pour les pages stratégiques. Bien sûr, cela dépend de la taille du site, mais d'après notre expérience il faut éviter d'avoir très peu de backlinks internes.

Toujours concernant les backlinks internes, il est intéressant d'étudier aussi les ancres de liens (c'est-à-dire le texte cliquable qui mène à la page, ou bien l'attribut ALT dans le cas d'un lien sur une image). Là aussi, il est préférable **d'éviter d'avoir toujours exactement le même texte dans les liens qui pointent vers une page donnée**. En variant les ancres de liens, vous rendez les choses plus naturelles ; en plus, vous optimisez pour la longue traîne car la page de destination sera mieux positionnée sur un plus grand nombre de requêtes.

Au sujet des ancres de liens, évitez les ancres vides : il s'agit souvent de liens placés sur des images sans attribut ALT, mais parfois aussi de liens texte vide. C'est déconseillé car cela ne permet pas d'optimiser le référencement de la page recevant le lien.

## Remarques importantes :

1. Si une page A fait plusieurs liens fait une même page B, seul le 1er lien (dans l'ordre du code source) est pris en compte dans notre analyse des ancres.
2. Comme pour la profondeur, nous ne recommandons pas d'augmenter artificiellement le nombre d'ancres dans les liens uniquement pour réduire le taux d'erreur. Nous recommandons de privilégier les liens utiles à l'internaute, c'est-à-dire situés à des endroits faciles à identifier, entre des pages de thématique proche. Par exemple, c'est une mauvaise solution que de mettre des liens vers (presque) toutes les pages du site dans une page plan du site.
3. Si vous utilisez un menu listant une grande partie de vos pages, il sera sans doute difficile de varier leurs ancres (car seul le 1er lien est pris en compte), mais ne changez pas votre menu si vous estimez qu'il est utile tel quel pour vos visiteurs.
4. En résumé, il n'est pas toujours possible ou facile d'éviter que certaines pages n'aient qu'une seule ancre dans leurs backlinks internes, par contre il faut essayer de le faire pour les pages stratégiques.

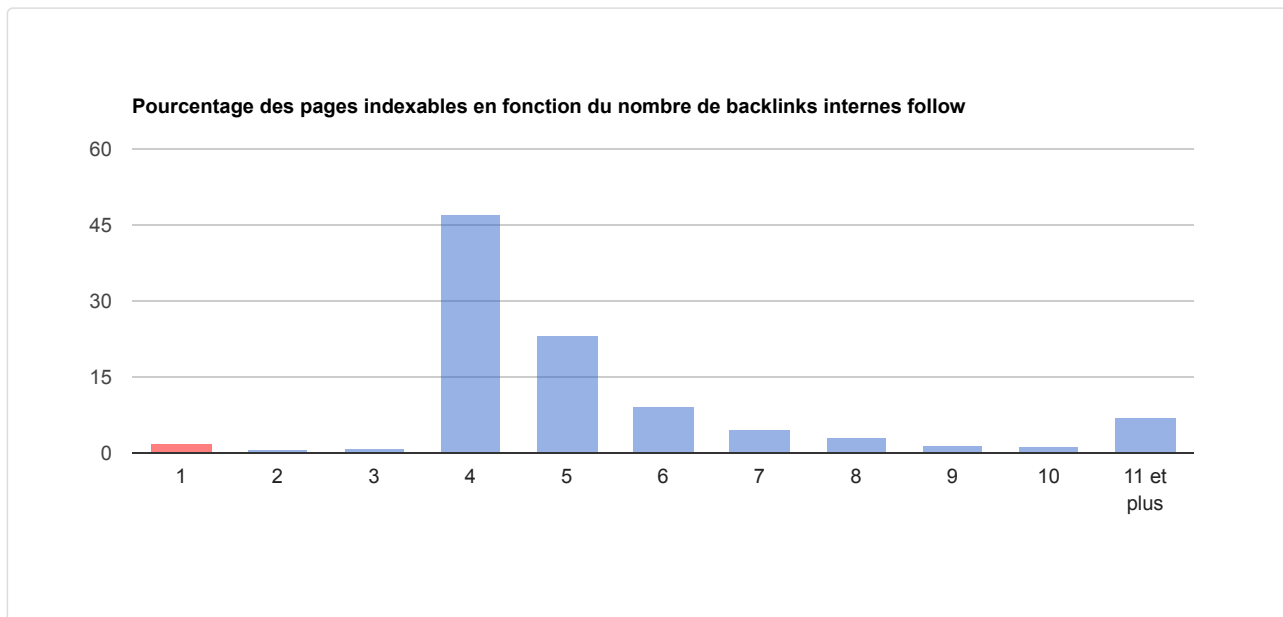
Nous avons étudié les pages qui reçoivent trop peu de backlinks internes ou qui ont trop peu de variété dans les ancres de liens internes, ou qui ont des ancres vides (liens follow uniquement).

Pour savoir en détail comment nous identifions et comptons les liens, [cliquez ici](#). L'étude a porté sur les pages HTML indexables. Vous trouverez [dans la conclusion](#) une annexe qui liste tous les backlinks internes follow de toutes les URL étudiées, avec l'ancre de chacun. Quand une page fait plusieurs liens vers la même page de destination, tous les liens sont listés, même si seul le premier est pris en compte dans l'audit.

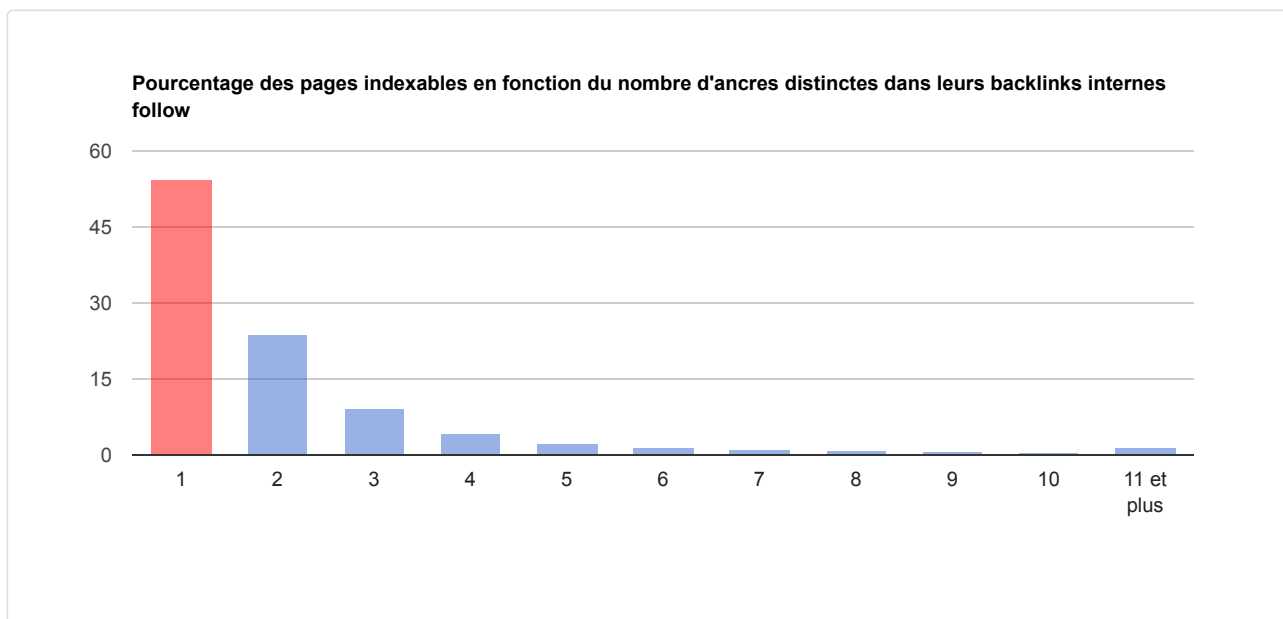
Voici les résultats de notre analyse :

	<b>Erreurs</b>	<b>Avertissements</b>	
<b>Nb de pages avec seulement 1 backlink interne</b>	137	n/a	
<b>Nb de pages avec seulement 2 backlinks internes</b>	n/a	38	
<b>Nb de pages avec toujours la même ancre (backlinks internes)</b>	4 192	n/a	
<b>Nb de pages avec seulement 2 ancres distinctes (backlinks internes)</b>	n/a	1 836	
<b>Nb de pages avec un ou des backlinks internes à ancre vide</b>	n/a	10 099	
<b>Nb total d'erreurs et d'avertissements</b>	4 192	1 901	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>	<b>54,34 %</b>	24,64 %	

Le graphique ci-dessous illustre l'analyse du nombre de backlinks par page :



Le graphique ci-dessous illustre l'analyse du nombre d'ancres distinctes par page :



### 5.3 Conclusion sur la profondeur et le maillage interne

Pour optimiser votre référencement, vos pages ne doivent pas être trop profondes (nombre de clics depuis la page d'accueil) et leur maillage doit être efficace (plusieurs liens entrants avec des ancres variées, non vides).

Pour cette conclusion, nous tenons compte uniquement des erreurs (pas des avertissements).

Voici les résultats de notre analyse :

	Erreurs	
<b>Nb de pages trop profondes ou avec un maillage interne insuffisant</b>	7 115	
<b>Nb de pages indexables</b>		7 715
<b>Taux d'erreur</b>	<b>92,22 %</b>	



**Annexe**

Le fichier annexe `exemple_com_05-3_profondeur-maillage.xlsx` contient les statistiques sur la profondeur, le nombre de backlinks internes et d'ancres distinctes.

**Annexe**

Le fichier annexe `exemple_com_05-3_ancres-internes.xlsx` contient pour chaque URL la liste des ancres avec pour chaque ancre le nombre de backlinks internes ainsi qu'un exemple d'URL de backlink interne.

**Annexe**

Le fichier annexe `exemple_com_05-3_liens-ancres-vides.xlsx` contient pour chaque URL la liste des pages recevant au moins un lien interne avec une ancre vide.

## 6 Balise title

Toutes les analyses qui suivent portent sur les pages HTML indexables.

### 6.1 Taille de la balise Title

Pour optimiser la balise title, il faut exploiter son potentiel, c'est-à-dire utiliser au mieux la place disponible. Mais il faut également éviter les titres trop longs qui risquent de ne pas s'afficher en entier dans les résultats de recherche. L'idéal est de ne pas dépasser ~70 caractères ; au-delà de 80, nous considérons ça une erreur.

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
<b>Nb de pages sans balise title ou avec une balise title vide</b>	0	n/a	
<b>Nb de pages avec seulement 1 mot dans title</b>	0	n/a	
<b>Nb de pages avec seulement 2 mots dans title</b>	3	n/a	
<b>Nb de pages avec seulement 3 mots dans title</b>	21	n/a	
<b>Nb de pages avec entre 71 et 80 caractères dans title</b>	n/a	14	
<b>Nb de pages avec entre 81 et 100 caractères dans title</b>	0	n/a	
<b>Nb de pages avec + de 100 caractères dans title</b>	0	n/a	
<b>Nb total d'erreurs et d'avertissements</b>	24	14	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>	<b>0,31 %</b>	0,18 %	

#### Annexe

Le fichier annexe `exemple_com_06-1_taille-balise-title.xlsx` contient les données associées.

### 6.2 Doublons dans la balise title

Pour optimiser la balise title, il faut que chaque page ait une balise title unique (pas utilisée ailleurs sur le site), sinon nous considérons cela comme une erreur.

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
<b>Nb de balises title utilisées plus d'une fois sur le site (hors balise vide)</b>	3	n/a	
<b>Nb de pages indexables par Google concernées (balise vide incluse)</b>	6	n/a	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>	<b>0,08 %</b>	n/a	

**Annexe**

Le fichier annexe example\_com\_06-2\_doublons-title.xlsx contient les données associées.

### 6.3 Répétition de mots dans la balise Title

Pour éviter la sur-optimisation de la balise title, il faut éviter de répéter plus de 2 fois le même mot (et encore, c'est souvent trop de répéter 2 fois un même mot, c'est pourquoi nous considérons ce cas comme un avertissement). Néanmoins, il peut y avoir des cas particuliers, notamment avec les articles et autres mots très courants. Voici les résultats de notre analyse :

	<b>Erreurs</b>	<b>Avertissements</b>	
<b>Nb de pages avec 2 fois un même mot dans title</b>	n/a	238	
<b>Nb de pages avec 3 fois un même mot dans title</b>	3	n/a	
<b>Nb de pages avec 4 fois un même mot dans title</b>	0	n/a	
<b>Nb de pages avec 5 fois ou plus un même mot dans title</b>	0	n/a	
<b>Nb total d'erreurs et d'avertissements</b>	3	238	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>	<b>0,04 %</b>	3,08 %	

**Annexe**

Le fichier annexe example\_com\_06-3\_repetition-mots-dans-title.xlsx contient les données associées.

### 6.4 Votre marque dans la balise Title

Pour optimiser la balise title, même s'il peut y avoir des exceptions, nous recommandons de ne pas inclure votre marque (faites-le dans la meta description).

Voici les résultats de notre analyse :

	<b>Erreurs</b>	<b>Avertissements</b>	
<b>Nb de pages qui contiennent la marque "example" dans title</b>	n/a	4	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>		0,05 %	

**Annexe**

Le fichier annexe example\_com\_06-4\_marque-dans-title.xlsx contient les données associées.

### 6.5 Conclusion sur la balise Title

Pour optimiser votre référencement, aucune balise title ne doit comporter une ou plusieurs des erreurs décrites précédemment. Nous avons identifié les cas où une page HTML contient au moins une de ces erreurs dans la balise Title.

Pour cette conclusion, nous tenons compte uniquement des erreurs (pas des avertissements).

Voici les résultats de notre analyse :

	<b>Erreurs</b>	
<b>Nb de pages avec une ou plusieurs erreurs dans la balise Title</b>	33	
<b>Nb de pages indexables</b>		7 715
<b>Taux d'erreur</b>	<b>0,43 %</b>	

## 7 Balise meta description

### 7.1 Taille de la balise meta description

Pour optimiser la balise meta description, il faut exploiter son potentiel, c'est-à-dire utiliser au mieux les ~230 caractères disponibles (environ 3 lignes). Nous conseillons de rédiger un texte d'au moins 90 caractères.

Voici les résultats de notre analyse :

	Erreurs	
<b>Nb de pages avec balise meta description vide ou absente</b>	0	
<b>Nb de pages avec une meta description de moins de 90 caractères</b>	84	
<b>Nb total</b>	84	
<b>Nb de pages indexables</b>		7 715
<b>Taux d'erreur</b>	<b>1,09 %</b>	

#### Annexe

Le fichier annexe `exemple_com_07-1_taille-meta-description.xlsx` contient les données associées.

### 7.2 Doublons dans la balise meta description

Pour optimiser la balise meta description, il faut qu'elle soit unique (pas utilisée ailleurs sur le site).

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
<b>Nb de balises meta description utilisées plus d'une fois (hors balise vide)</b>	3	n/a	
<b>Nb de pages concernées</b>	6	n/a	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>	<b>0,08 %</b>	n/a	

#### Annexe

Le fichier annexe `exemple_com_07-2_doublons-meta-description.xlsx` contient les données associées.

### 7.3 Conclusion sur la balise meta description

Pour optimiser votre référencement, aucune balise meta description ne doit comporter une ou plusieurs des erreurs décrites précédemment. Nous avons identifié les cas où une page HTML contient au moins une de ces erreurs dans la balise meta description.

Pour cette conclusion, nous tenons compte uniquement des erreurs (pas des avertissements).

Voici les résultats de notre analyse :

	<b>Erreurs</b>	
<b>Nb de pages avec une ou plusieurs erreurs dans la balise meta description</b>	90	
<b>Nb de pages indexables</b>		7 715
<b>Taux d'erreur</b>	<b>1,17 %</b>	

## 8 Contenu des pages

### 8.1 Balises H1-H6

Pour vous donner le plus de chances de réussir votre référencement, nous vous recommandons d'utiliser les balises de titres H1, H2, H3... H6 (notées "Hn"). Attention, ces balises ne doivent être intégrées en théorie que sur les pages à fort contenu éditorial : sur les autres (accueil, listings, etc.) nous conseillons généralement uniquement une balise H1.

Voici les règles que nous vous recommandons de suivre (considérées sinon comme des erreurs) :

- au minimum, une page doit avoir au moins avoir une H1 (sauf la page d'accueil ou celle de chaque sous-domaine où ce n'est pas indispensable et donc considéré comme un avertissement)
- une balise H1 ne doit pas être vide

Voici d'autres règles que nous vous recommandons également de suivre (considérées sinon comme des avertissements) :

- la 1ère balise d'une page doit être une H1
- il faut éviter les sauts de niveau : quand il s'agit de descendre dans les détails, une balise Hn ne peut être suivie que par une balise H(n+1) dans l'ordre du code source, mais pas H(n+2) ou H(n+3) : on ne doit pas sauter de niveau. Par exemple, après une H1, il ne peut y avoir qu'une H2. Après une H3, il ne peut y avoir qu'une H4 (ou bien une autre H3, ou bien on remonte et c'est une H2), mais pas une H5.
- la taille d'une balise Hn (nb de caractères) ne doit pas dépasser 80 caractères (même si ce chiffre est arbitraire, fixé selon notre expérience, l'idée est qu'un titre doit rester un titre et non un paragraphe)
- il est recommandé de limiter à 1 le nombre de balises H1 par page
- pour les balises H2-H6 "isolées" : dans un niveau donné, le nombre de titres du niveau inférieur doit être 0 ou au moins 2 (donc pas seulement 1). La raison est que si vous utilisez une balise H2 (ou H3, H4, H5, H6), c'est a priori pour découper en plusieurs sous-parties le bloc de niveau supérieur. Il doit donc y en avoir soit aucune, soit au moins 2 dans une page.
- dans la mesure du possible, il faut éviter d'utiliser un même libellé dans plusieurs balises H1-H6 d'une même page
- pour profiter de la longue traîne (et d'une limitation moins forte en nombre de caractères), essayez d'avoir un titre H1 légèrement différent de celui utilisé pour la balise Title de la (même) page

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
<b>Nb de pages sans balise H1</b>	150	0	
<b>Nb de pages avec une balise H1 mais vide</b>	0	n/a	
<b>Nb de pages où la 1ère balise Hn n'est pas une H1</b>	n/a	150	
<b>Nb de pages avec saut de niveau entre les Hn</b>	n/a	939	
<b>Nb de pages avec une balise Hn trop longue</b>	n/a	1 204	
<b>Nb de pages avec plusieurs balises H1</b>	n/a	0	
<b>Nb de pages avec balise H2-H6 isolée</b>	n/a	6 697	
<b>Nb de pages avec le même libellé dans plusieurs Hn</b>	n/a	67	
<b>Nb de pages avec le même libellé dans H1 et Title</b>	n/a	7 321	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>	<b>1,94 %</b>	n/a	

Vous trouverez dans l'annexe non seulement la liste des pages en erreur (sans balise H1), avec leur balise TITLE, mais aussi le libellé de toutes les balises H1 trouvées sur les autres pages.

#### Annexe

Le fichier annexe example\_com\_08-1\_balises-h1-h6.xlsx contient les données associées.

## 8.2 Doublons dans la balise H1

Pour optimiser la balise H1, il faut que chaque page ait une balise H1 unique (pas utilisée ailleurs sur le site), sinon nous considérons cela comme une erreur.

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
<b>Nb de balises H1 utilisées plus d'une fois sur le site (hors balise vide)</b>	4	n/a	
<b>Nb de pages indexables par Google concernées (balise vide incluse)</b>	171	n/a	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>	<b>2,22 %</b>	n/a	

#### Annexe

Le fichier annexe example\_com\_08-2\_doublons-h1.xlsx contient les données associées.

## 8.3 Taille du texte (zone principale de la page)

Comme vous le savez sans doute, l'objectif de l'algorithme Panda de Google est d'identifier les pages à faible valeur ajoutée. Panda est intégré au coeur de l'algorithme principal : même si Google n'indique plus de dates de mises à jour, il fonctionne chaque jour !

Nous y revenons plus loin dans ce rapport, mais un élément essentiel de la qualité est le contenu texte. Même si on trouve des contre-exemples, une page avec un contenu texte court est rarement de haute qualité (de nature à satisfaire tous les internautes). C'est ce qu'on appelle "thin content" dans le jargon SEO en anglais.

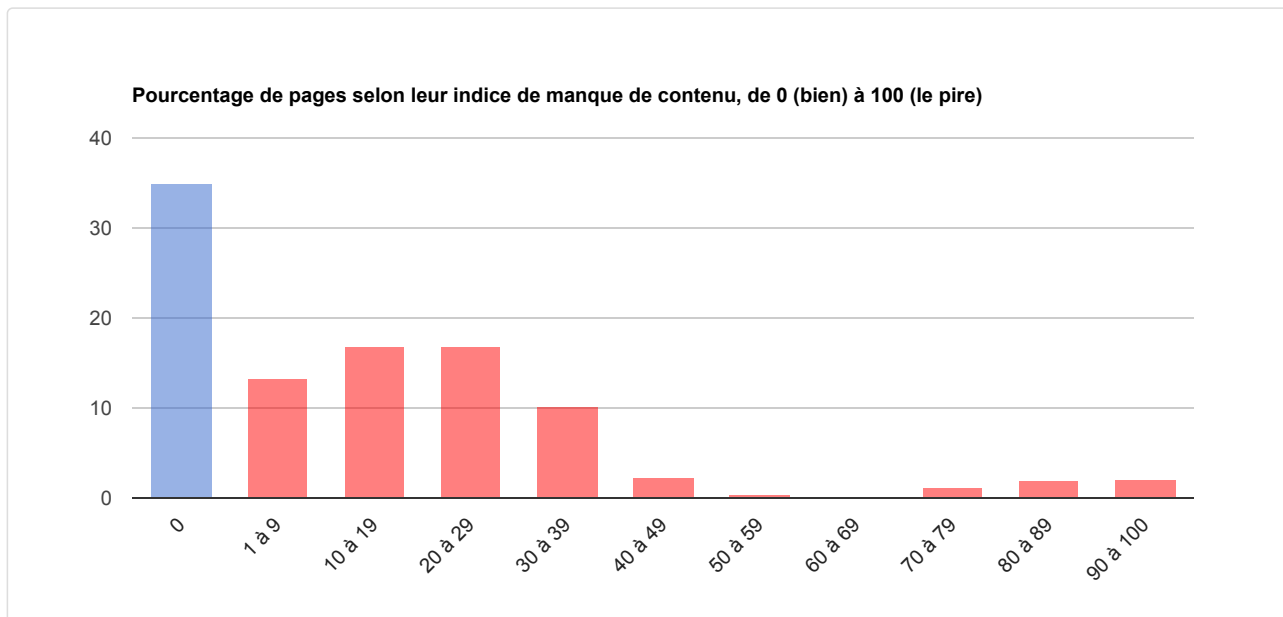
Une approche simpliste consiste à compter le nombre de mots dans l'ensemble de la page HTML (incluant l'entête, les menus, les barres latérales et le pied de page). Comme vous vous en doutez, cela ne fonctionne pas car ce qui compte, c'est la taille du contenu dans la zone principale de la page.

C'est pourquoi nous avons développé un algorithme spécifique, afin de vous aider à identifier les pages sans doute trop courtes.

Dans l'annexe ci-dessous, nous listons les pages dont le contenu texte est estimé trop court. Pour vous aider à les trier, nous fournissons un indice de gravité du problème, de 0 (aucun problème) à 100 (gros problème).

L'histogramme ci-dessous montre la répartition des pages de ce site selon leur indice de manque de contenu :





Le tableau ci-dessous donne la synthèse :

	<b>Erreurs</b>	<b>Avertissements</b>	
<b>Nb de pages sans doute trop courtes</b>	5 020	n/a	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>	<b>65,07 %</b>	n/a	

#### Annexe

Le fichier annexe `exemple_com_08-3_taille-texte.xlsx` contient les données associées.

## 8.4 Analyse du nombre de liens par page

Pour vous donner le plus de chances de réussir votre référencement, il faut que chaque page fasse suffisamment de liens vers les autres pages du site, mais pas trop.

Nous cherchons à identifier 4 types de problèmes :

- Les pages avec très peu de liens internes follow, a priori des erreurs : elles font moins de 5 liens
- Les pages avec trop de liens internes follow, sans doute également des erreurs : elles en font plus de 300
- Les pages avec des liens internes en nofollow (sauf cas particulier, cela n'a pas d'intérêt)
- Les pages qui font plus de 10 liens follow vers d'autres sites (à vous de valider que c'est bien votre choix). Nous avons exclu de nos analyses les réseaux sociaux suivants : LinkedIn, Twitter, Facebook, Pinterest, Viadeo, Myspace, Youtube, Dailymotion, Instagram, Vimeo, Flickr.

Nos analyses portent sur l'ensemble de la page, en incluant le haut de page, les éventuelles marges et le pied de page.

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
<b>Nb de pages avec en tout moins de 5 liens</b>	0	n/a	
<b>Nb de pages avec en tout plus de 300 liens</b>	n/a	0	
<b>Nb de pages avec plus de 10 liens externes</b>	n/a	35	
<b>Nb de pages avec 1 ou plusieurs liens internes nofollow</b>	n/a	3	
<b>Nb total</b>	0	38	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>	<b>0 %</b>	0,49 %	

#### Annexe

Le fichier annexe exemple\_com\_08-4\_nb-liens.xlsx contient les données associées.

## 8.5 Analyse des contenus dupliqués internes

Pour optimiser votre référencement, aucun contenu ne doit être disponible en interne à plusieurs URL différentes. En effet, si cela se produit, vous être confrontés aux problèmes de contenus dupliqués internes. Même si Google indique qu'il ne pénalise pas pour ça, avoir des contenus dupliqués internes dégrade les performances de votre référencement :

- le crawl est moins efficace car Google doit crawler des URL en trop. S'il y en a peu ce n'est pas bien grave, mais parfois cela impose à Google de crawler bien plus d'URL inutiles que d'URL importantes correspondant aux vrais contenus à faire indexer.
- vous ajoutez de la concurrence entre vos propres pages là où il ne devrait pas y en avoir. Laquelle des différentes URL Google va-t-il choisir de mettre en avant dans ses résultats ?
- vous éparpillez vos efforts, car les internautes qui souhaitent partager vos contenus ou vous faire des liens n'utiliseront pas une unique URL pour un contenu donné, mais un peu de chaque. Conséquence : vous aurez un score faible sur plusieurs URL (donc très peu visibles dans Google) au lieu d'avoir un bon score sur une seule URL (ce qui donne des chances d'être bien visible).

Nous avons analysé l'ensemble des pages HTML indexables afin d'identifier les cas de contenus dupliqués à 100%. Il s'agit donc des cas où l'ensemble du contenu HTML est strictement identique pour des URL différentes. Nous n'étudions pas les cas de contenus presque dupliqués, où la plupart du contenu est identique, mais pas la totalité. Les cas listés sont donc tous considérés comme des erreurs.

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
<b>Nb d'URL de pages HTML ayant un contenu strictement identique</b>	0	n/a	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>	<b>0 %</b>	n/a	

#### Annexe

Dans votre cas, comme aucune erreur n'a été détectée, le fichier annexe qui liste les erreurs n'a pas été généré.

## 8.6 Identifiants de session dans les URL

Il ne faut surtout pas que les robots trouvent des identifiants de session dans vos URL, car cela génère de nombreux problèmes.

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
<b>Nb de pages avec un ID de session</b>	0	n/a	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>	<b>0 %</b>	n/a	

#### Annexe

Dans votre cas, comme aucune erreur n'a été détectée, le fichier annexe qui liste les erreurs n'a pas été généré.

## 8.7 Paramètres de tracking dans les URL

En interne, il ne devrait pas y avoir de liens incluant des paramètres de tracking dans les URL (passés derrière le signe ?).

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
<b>Nb de pages avec une variable de tracking</b>	0	n/a	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>	<b>0 %</b>	n/a	

#### Annexe

Dans votre cas, comme aucune erreur n'a été détectée, le fichier annexe qui liste les erreurs n'a pas été généré.

## 8.8 Conclusion sur le contenu des pages

Pour optimiser votre référencement, aucune page ne doit comporter une ou plusieurs des erreurs décrites précédemment. Nous avons identifié les cas où une page HTML contient au moins une de ces erreurs.

Pour cette conclusion, nous tenons compte uniquement des erreurs (pas des avertissements).

Voici les résultats de notre analyse :

	Erreurs	
<b>Nb de pages avec une ou plusieurs erreurs dans le contenu</b>	5 024	
<b>Nb de pages indexables</b>		7 715
<b>Taux d'erreur</b>	<b>65,12 %</b>	

## 9 Liens sortants externes

Dans cette rubrique, nous faisons le point sur tous les liens qui sortent de votre site (ils pointent donc vers un autre nom de domaine). Nous les appelons des « liens externes », qu'il ne faut pas confondre avec des backlinks (qui peuvent être eux aussi internes ou externes).

Nous avons ignoré les liens nofollow et avons exclu de nos analyses les réseaux sociaux suivants : LinkedIn, Twitter, Facebook, Pinterest, Viadeo, Myspace, Youtube, Dailymotion, Instagram, Vimeo, Flickr.

### 9.1 Nombre total de liens sortants externes

Nous avons identifié 22 662 liens externes pointant vers 6 161 URL distinctes.

#### Annexe

Le fichier annexe `example_com_09-1_liens-sortants-externes.xlsx` contient les données associées.

### 9.2 Erreurs sur les liens sortants externes

Nous nous intéressons ici uniquement au cas où vos liens pointent vers un document en erreur. Comme vous le savez, il vaut mieux éviter ce type de situation, aussi bien pour les internautes que pour le référencement.

Dans certains cas, l'URL de destination se fait rediriger : si l'URL finale est accessible, ce n'est pas un problème pour l'internaute mais dans l'idéal il ne devrait y avoir que des liens directs. C'est pourquoi dans ce cas nous considérons qu'il s'agit simplement d'un avertissement.

Voici le bilan des codes HTTP obtenus pour ces URL en erreur :

Code HTTP	Nb d'URL	% des URL externes
301	1 967	31,93 %
302	406	6,59 %
303	1	0,02 %
307	3	0,05 %
308	5	0,08 %
400	14	0,23 %
401	4	0,06 %
402	1	0,02 %
403	219	3,55 %
404	1 156	18,76 %
408	2	0,03 %
410	29	0,47 %
500	980	15,91 %
503	589	9,56 %
504	1	0,02 %
526	1	0,02 %

	Erreurs	Avertissements	
<b>Nombre total de liens sortants externes redirigés</b>	n/a	2 382	
<b>Nombre total de liens sortants externes en erreur</b>	2 996	n/a	
<b>Nombre total de liens sortants externes</b>			6 161
<b>Taux d'erreur</b>	48,63 %	38,66 %	

Tous les détails (liste des codes HTTP pour chacune des URL de destination externe en erreur ainsi que l'URL du site qui fait un lien vers une page en erreur) sont dans le fichier annexe.

## 10 Analyse des images

L'analyse des images consiste à étudier les balises <img> sur l'ensemble des pages HTML qui ont pu être crawlées (code HTTP 200).

**25 560 images** (autorisées au crawl) ont été trouvées et analysées sur **10 100 pages** HTML.

### Annexe

Le fichier annexe example\_com\_images.xlsx contient la liste des images.

### Annexe

Le fichier annexe example\_com\_pages-avec-images.xlsx contient la liste des pages HTML contenant au moins 1 image.

Une image est parfois utilisée sur plusieurs pages HTML. Dans ces cas, il peut arriver qu'une même image ne soit pas intégrée de la même façon dans ces pages : parfois avec un attribut ALT, parfois sans, etc. Toutes les analyses associées sont présentées ci-après.

### Annexe

Le fichier annexe example\_com\_pages-et-images.xlsx contient la liste des couples Page / Image.

### 10.1 Images bloquées par le(s) fichier(s) robots.txt

En analysant les 10 100 pages HTML accessibles, **182 images** ont été identifiées mais non analysées car bloquées dans le fichier robots.txt. Vérifiez si c'était bien voulu.

Aucun crédit n'a été compté pour ces URL d'images.

### Annexe

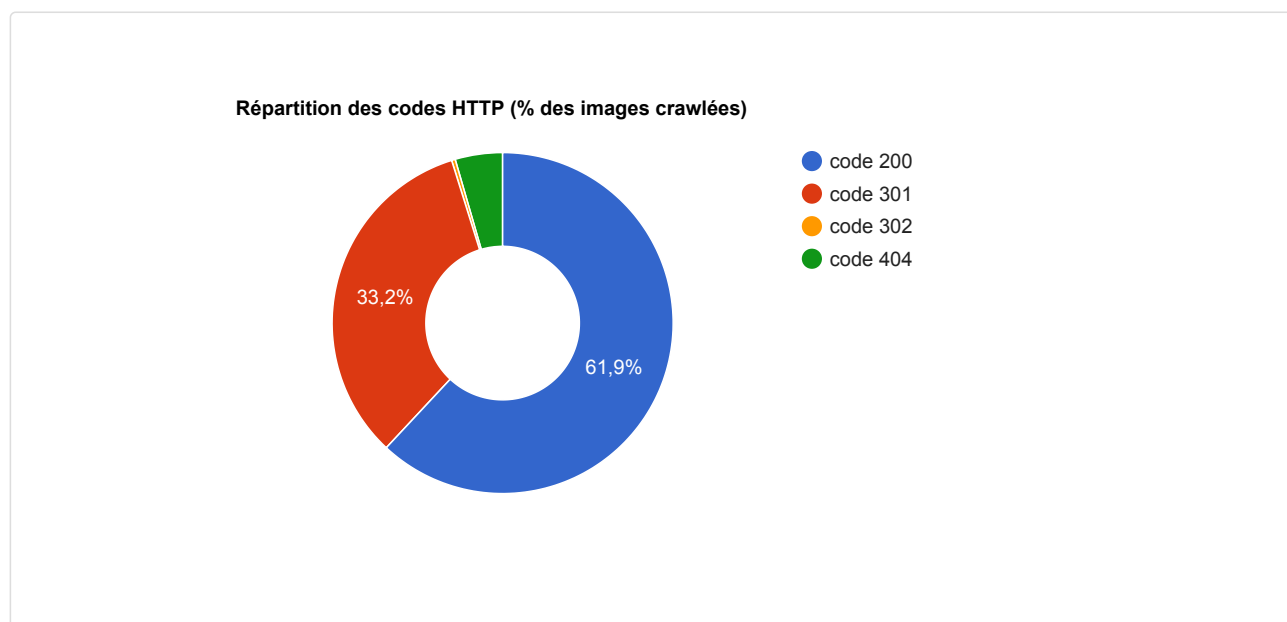
Le fichier annexe example\_com\_images-bloquees.xlsx contient les données associées.

### 10.2 Codes HTTP des URL des images

Pour qu'une image soit correctement indexée, elle doit renvoyer un code HTTP 200. Voici le bilan des codes HTTP obtenus pour les URL des images (lisez leur signification [en fin de rapport](#)) :

Code HTTP	Nb d'URL en HTTP	Nb d'URL en HTTPS	Nb d'URL (HTTP et HTTPS)	% des URL crawlées
200	350	15 584	15 834	61,95 %
301	1	8 489	8 490	33,22 %
302	69	19	88	0,34 %
404	451	697	1 148	4,49 %
<b>Total erreurs 3xx</b>	70	8 508	8 578	33,56 %
<b>Total erreurs 4xx</b>	451	697	1 148	4,49 %
<b>Total erreurs</b>	521	9 205	9 726	<b>38,05 %</b>

Voici ce que cela donne sous forme graphique :



Toutes ces données sont disponibles dans l'annexe des images fournie précédemment.

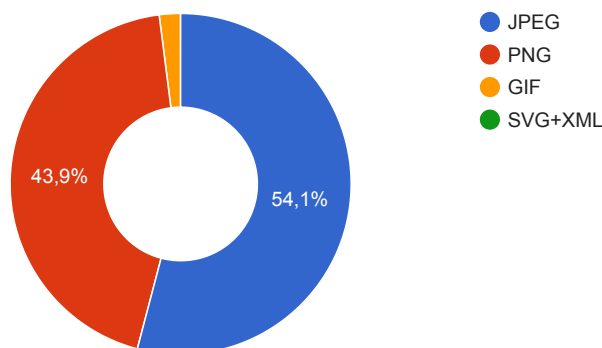
### 10.3 Types MIME (formats d'images)

Pour connaître les différents formats d'images sur le site, il faut analyser les types MIME (lisez leur signification en fin de rapport). Voici le bilan de ceux rencontrés pour les images répondant correctement (code HTTP 200) :

Type MIME	Nb d'URL	% des URL crawlées en code 200
image/jpeg	8 564	54,09 %
image/png	6951	43,90 %
image/gif	318	2,01 %
image/svg+xml	1	0,01 %

Voici ce que cela donne sous forme graphique :

Répartition des types MIME des images



Toutes ces données sont disponibles dans l'[annexe des images](#) fournie précédemment.

## 10.4 Images indexables

Les **images indexables** sont toutes celles :

- qui renvoient un code 200
- et qui ne sont pas interdites d'indexation (via un en-tête HTTP X-Robots-Tag)
- et qui ne contiennent pas d'URL canonique différente de l'URL crawlée

Voici le décompte pour votre site :

	Nb d'images	% des images code 200
<b>Images avec code HTTP 200 (A)</b>	15 834	
<b>Dont images interdites d'indexation</b>	390	2,5 %
<b>Dont images avec URL canonique différente de l'URL crawlée</b>	15	0,1 %
<b>Nb total d'images non-indexables (B)</b>	405	2,6 %
<b>Nombre d'images indexables (A - B)</b>	15 429	97,4 %

Toutes ces données sont disponibles dans l'[annexe des images](#) fournie précédemment.

## 10.5 Images trop lourdes

Pour assurer une bonne vitesse d'affichage, il faut **éviter les images trop lourdes**. Pour cet audit, les images de + de 100 Ko sont considérées trop lourdes (valeur paramétrée au lancement).

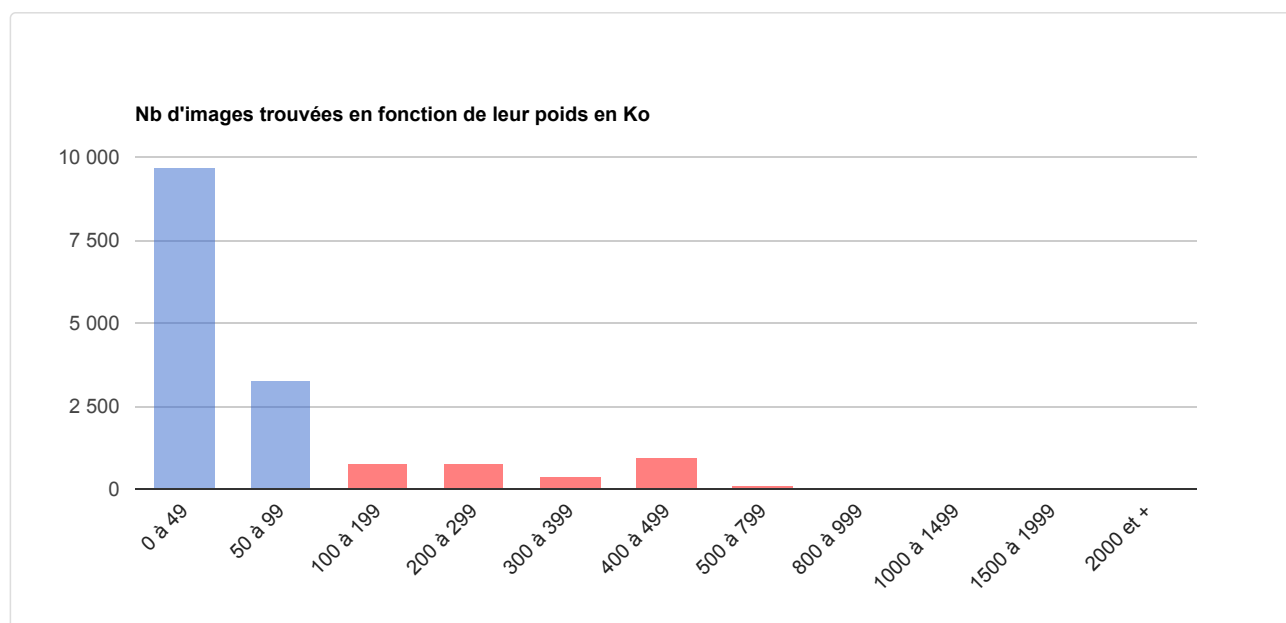
**2 889 images trop lourdes** ont été trouvées.

Voici la répartition des images trouvées (code 200) en fonction de leur poids (en Ko) :



Poids (Ko)	Nb d'images	Pourcentage des images
0 à 49	9 690	61,20 %
50 à 99	3 255	20,56 %
100 à 199	731	4,62 %
200 à 299	754	4,76 %
300 à 399	387	2,44 %
400 à 499	923	5,83 %
500 à 799	76	0,48 %
800 à 899	13	0,08 %
1000 à 1499	5	0,03 %
1500 à 1999	0	0 %
2000 et +	0	0 %

Voici ce que cela donne sous forme graphique :



Toutes ces données sont disponibles dans [l'annexe des images](#) fournie précédemment.

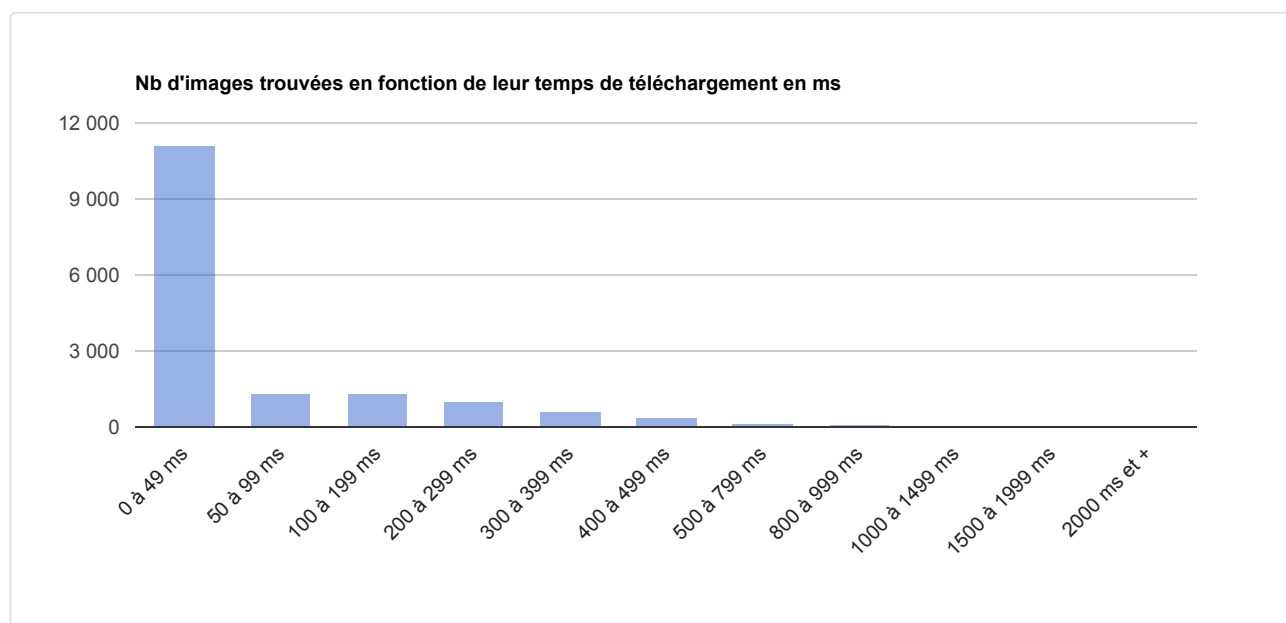
## 10.6 Temps de téléchargement des images

En complément de l'analyse du poids du fichier présentée ci-avant, le temps de téléchargement a été mesuré pour chaque image analysée en code 200.

Voici la répartition par tranches de 200 ms

Temps de téléchargement (ms)	Nb d'images	Pourcentage des images
0 à 49 ms	11 054	69,81 %
50 à 99 ms	1 310	8,27 %
100 à 199 ms	1 282	8,10 %
200 à 299 ms	953	6,02 %
300 à 399 ms	601	3,80 %
400 à 499 ms	357	2,25 %
500 à 799 ms	122	0,77 %
800 à 999 ms	52	0,33 %
1000 à 1499 ms	34	0,21 %
1500 à 1999 ms	41	0,26 %
2000 ms et +	28	0,18 %

Voici ce que cela donne sous forme graphique :



## 10.7 Attribut ALT des images

Pour un meilleur référencement, les images doivent avoir un **attribut alt rempli** (avec une courte description de l'image), sauf les images décoratives (qui doivent avoir un attribut alt présent mais vide).

**86 images** sont utilisées dans des pages avec un attribut alt toujours absent, ce qu'il faut corriger.

**452 images** sont utilisées dans des pages avec un attribut alt présent mais toujours vide. C'est OK s'il s'agit d'images décoratives (design) mais pas pour les autres.

Toutes ces données sont disponibles dans [l'annexe des images](#) fournie précédemment.

## 10.8 Images avec dimensions forcées en HTML

Parfois, une image est affichée dans une page HTML avec des "dimensions forcées". Cela signifie que les dimensions indiquées dans les attributs width et height de la balise img diffèrent des valeurs réelles (largeur et hauteur) du fichier image. L'erreur la plus répandue est d'afficher une image de façon réduite par rapport à sa taille

réelle, ce qui oblige pourtant le navigateur à télécharger la grande taille "pour rien". L'analyse effectuée ne tient pas compte des CSS.

**7 885 images** ont été trouvées toujours affichées avec des dimensions forcées dans le code HTML.

Dans l'annexe, la colonne "ratio taille HTML / taille image (optimal : 1)" vous permet de repérer les cas les plus extrêmes :

- une valeur élevée indique que l'image est artificiellement zoomée (elle va pixelliser) ;
- une valeur faible indique que l'image est affichée bien plus petite que la réalité (ce n'est pas optimal en termes de poids). La valeur 0 (zéro) correspond à un arrondi, cela signifie que le ratio est très faible.

Toutes ces données sont disponibles dans l'[annexe des images](#) fournie précédemment.

## 10.9 Images HTTP sur pages HTTPS

Sur une page en HTTPS, toutes les ressources utilisées doivent être également en HTTPS, sinon le navigateur affiche un message d'avertissement. On parle de "mixed content" ou de "insecure content" quand certaines ressources sont en HTTP sur une page HTTPS. Parmi toutes les ressources, ceci s'applique bien évidemment aux images.

Remarque : une image en HTTP qui est redirigée vers du HTTPS est considérée comme HTTP.

**482 pages** HTML en HTTPS contiennent au moins une image HTTP, ce qu'il faut corriger.

Toutes ces données sont disponibles dans l'[annexe des images](#) fournie précédemment.

## 10.10 Bilan des pages comportant des images

Voici les résultats du bilan de notre analyse des pages avec images :

	Erreurs	Avertissements	
<b>Nb de pages avec image(s) inaccessible(s)</b>	5 502	n/a	
<b>Nb de pages en HTTPS avec image(s) en HTTP</b>	482	n/a	
<b>Nb de pages avec image(s) sans attribut ALT</b>	248	n/a	
<b>Nb de pages avec + de 50 images</b>	74	n/a	
<b>Nb de pages avec image(s) trop lourde(s)</b>	2 889	n/a	
<b>Nb de pages avec + de 1 000 Ko d'images</b>	41	n/a	
<b>Nb de pages avec image(s) dont la taille est forcée dans le code HTML</b>	n/a	2 615	
<b>Nb total d'erreurs et d'avertissements</b>	5 823	2 615	
<b>Nb de pages avec images</b>			10 100
<b>Taux d'erreur</b>	<b>57,65 %</b>	78,07 %	

## 11 Analyse de la qualité

Google améliore sans arrêt son algorithme afin de favoriser les contenus de qualité et pénaliser les autres. C'est un objectif majeur pour l'algorithme Panda, mais aussi pour d'autres (Phantom et dans une certaine mesure Penguin, ainsi que des modules sans nom intégrés au cœur de l'algorithme général).

Nous ne prétendons pas pouvoir analyser la qualité des pages comme peut le faire l'algorithme de Google, mais nous avons tout de même développé un algorithme maison qui repère certains problèmes sur vos pages. Nous nous basons sur des aspects techniques mais aussi des éléments du contenu qui nuisent à l'expérience utilisateur.

Nous calculons le risque pour chaque page, noté QR pour "Quality Risk" : plus QR est élevé (le maximum étant 100), plus vous devriez étudier nos recommandations pour améliorer la qualité de la page.

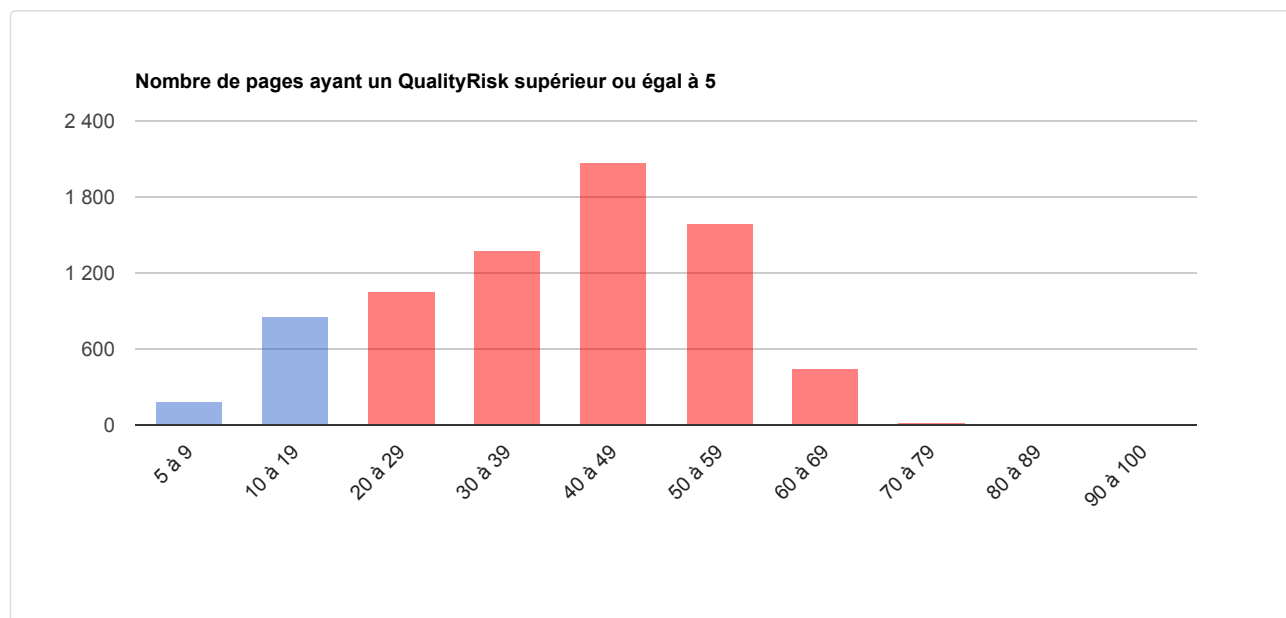
Pour réduire le QualityRisk d'une page, suivez tous les conseils déjà listés précédemment dans ce rapport d'audit. En voici le résumé :

- temps de téléchargement (page brute hors images, JS, CSS...) : ne dépassez pas 600ms (essayez même d'atteindre 200ms)
- pas trop de profondeur (nombre de clics depuis l'accueil)
- un bon maillage interne (si possible plusieurs backlinks internes avec des textes d'ancrage variés)
- balise Title : spécifique (unique) à chaque page, ni trop courte (au moins 3 mots) ni trop longue (moins de 80 caractères), sans répétition de mots
- une balise meta description spécifique (unique) à chaque page, pas trop courte (au moins 90 caractères)
- au moins une balise H1 spécifique (unique) à chaque page, et s'il y a des balises H2-H6, elles doivent décrire l'architecture du contenu principal de la page
- le contenu : spécifique (unique) à chaque page, de taille suffisante
- les liens internes présents dans la page : il vaut mieux éviter d'en faire trop (300 liens follow nous semble un maximum)
- les liens sortants externes doivent pointer vers des pages accessibles

Comme vous l'avez compris, il ne s'agit que d'une première piste : à vous de compléter par d'autres analyses !

Pour l'ensemble des pages indexables de ce site, **le QualityRisk moyen est de 38,3** (le pire possible étant 100).

L'histogramme ci-dessous montre la répartition des pages de ce site ayant un QualityRisk  $\geq 5$  (entre 0 et 4, les pages sont considérées sans risque) :



Par expérience, nous vous conseillons de ne pas avoir de pages avec un QualityRisk supérieur ou égal à 20. Le tableau ci-dessous donne la synthèse :

	<b>Erreurs</b>	<b>Avertissements</b>	
<b>Nb de pages avec un QualityRisk &gt;= 20</b>	6 538	n/a	
<b>Nb de pages avec un QualityRisk entre 5 et 19</b>	n/a	1 042	
<b>Nb de pages indexables</b>			7 715
<b>Taux d'erreur</b>	<b>84,74 %</b>	13,51 %	

Remarque : avant de supprimer une page identifiée comme étant à risque en termes de qualité, veuillez vérifier qu'elle ne correspond pas à l'un des 3 cas suivants (pour lesquels il ne faut pas supprimer la page mais l'améliorer) :

- si la page génère du trafic via le référencement naturel
- ou si la page a reçu des backlinks de qualité
- ou si après analyse humaine, vous estimez qu'en réalité la page n'est pas en risque

#### Annexe

Dans l'annexe, nous listons uniquement les pages ayant un QualityRisk supérieur ou égal à 5. Le fichier annexe `exemple_com_10_analyse-qualite.xlsx` contient les données associées.

## 12 Analyse Google Analytics

Dans tout ce qui précède, nous avons étudié le site lui-même (par un crawl et l'analyse qui a suivi) pour identifier des erreurs ou des améliorations possibles concernant le référencement naturel. L'indice QualityRisk regroupe tout ça par page et permet d'**estimer l'efficacité d'une page** (ou son potentiel d'amélioration).

Grâce à l'accès au compte Google Analytics associé au site, des données mesurant les performances SEO effectives des pages ont pu être récupérées. En croisant les 2 analyses, il est possible d'obtenir des informations à haute valeur ajoutée, de prioriser les actions et de gagner du temps et de l'efficacité.

Remarque : pour toutes les analyses liées à Google Analytics, les URL de pagination sont exclues, ainsi que l'URL de départ (souvent la page d'accueil) car elle génère du trafic sur le nom du site, ce qui n'est pas comparable.

7 715 pages indexables ont été trouvées lors du crawl ; parmi elles, 1 est exclue des analyses croisées avec Google Analytics.

Dans la suite de cette section liée à Analytics, le terme "page" fait référence aux 7 714 pages indexables trouvées lors du crawl excluant ces pages spéciales.

Aucun échantillonnage n'a été effectué par Google Analytics pour fournir ces données ([explications](#)). Si en allant dans votre compte Google Analytics vous n'obtenez pas les mêmes valeurs pour les mêmes conditions que cet audit, cela peut provenir de l'échantillonnage. En effet, la récupération des données est réalisée via une méthode basée sur l'API officielle qui réduit les cas où l'échantillonnage se produit.

À savoir : les performances des pages AMP ne sont actuellement pas prises en compte (seules celles de la version HTML le sont).

Les pages qui ont généré des visites via le référencement naturel sur cette période sont appelée **actives**, les autres **inactives**. Dans Google Analytics, il s'agit du nombre de sessions enregistrées pour le support "organic" pour chaque page de destination.

Bien entendu, il est préférable d'avoir le maximum de pages actives. Et même parmi les pages actives, certaines sont très faiblement actives, ce qui revient presque au même (générer 1 ou 2 visites par an, c'est presque comme en générer 0). À chaque fois que vous arriverez à rendre active une page qui ne l'est pas actuellement, vous augmenterez votre trafic SEO. Idem pour les pages faiblement actives qui peuvent être optimisées.

Pour identifier tout ça, examinez l'analyse des pages inactives ci-après ainsi que [l'analyse des pages zombies](#) qui va plus loin pour vous aider...

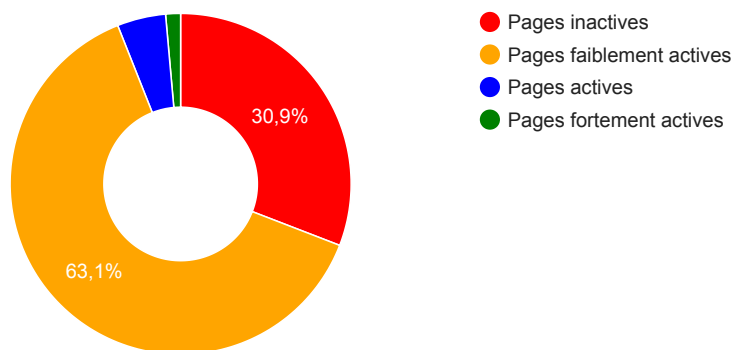
### 12.1 Pages actives et inactives

Voici les détails :

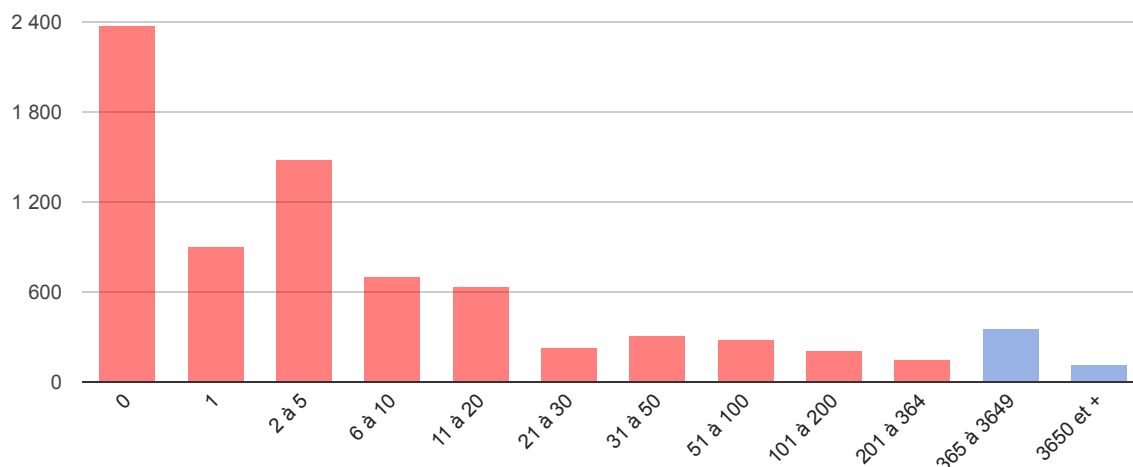
- Nb de sessions : **1 599 785** sur les 365 jours de la période, soit en moyenne 131 489 pour 30 jours, générées via le référencement naturel par 5 331 pages
- Nb de pages analysées : 7 714 qui se décomposent ainsi :
  - Nb de pages inactives : **2 380**
  - Nb de pages faiblement actives (1 à 364 sessions) : **4 867**
  - Nb de pages actives (365 à 3649 sessions) : **355**
  - Nb de pages fortement actives (3650 sessions et +) : **109**
- Au global, Score Pages Actives : **21,4 %**

Voici ce que cela donne sous forme graphique :

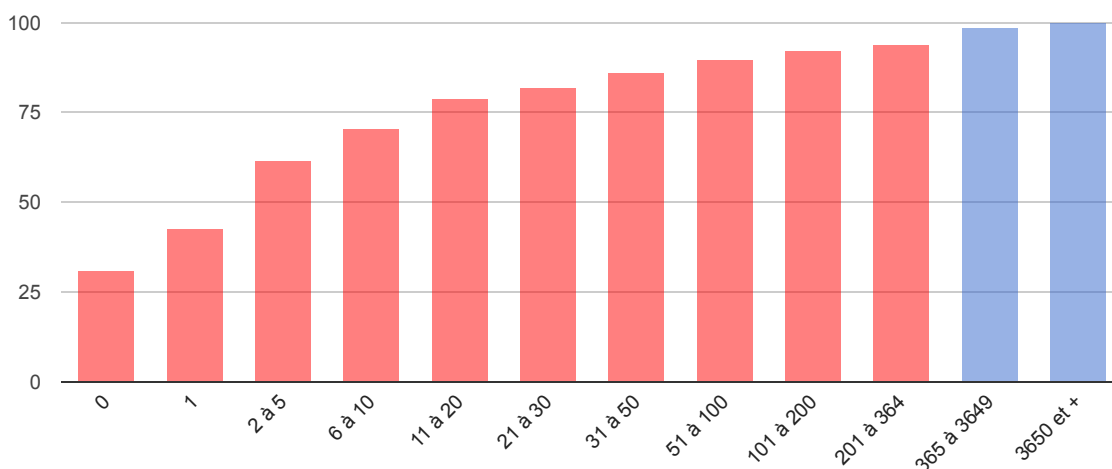
Répartition des pages selon leur performance SEO (nombre de sessions "organic" générées) sur 365 jours



Nombre de pages selon le nombre de sessions "organic" générées sur 365 jours



Pourcentage cumulé de pages indexables selon le nombre de sessions "organic" générées sur 365 jours



Comme vous pouvez le constater sur le graphique ci-dessus, **94,1 %** de vos pages génèrent (chacune) moins d'une visite par jour !

Les pages inactives (ou faiblement actives) et pourtant trouvées par le crawl (et indexables) doivent attirer votre attention. En effet, il s'agit de pages avec lesquelles vous avez tout fait pour qu'elles génèrent du trafic, en vain (sur la période étudiée).

On peut même trouver des pages qui n'ont jamais été consultées pendant 1 an, toutes sources de trafic confondues ! Cette information est fournie dans l'annexe de la conclusion sous le libellé "Nb de vues de la page (GA)".

C'est justement tout l'intérêt de l'analyse des pages zombies ci-après !

## 12.2 Pages zombies

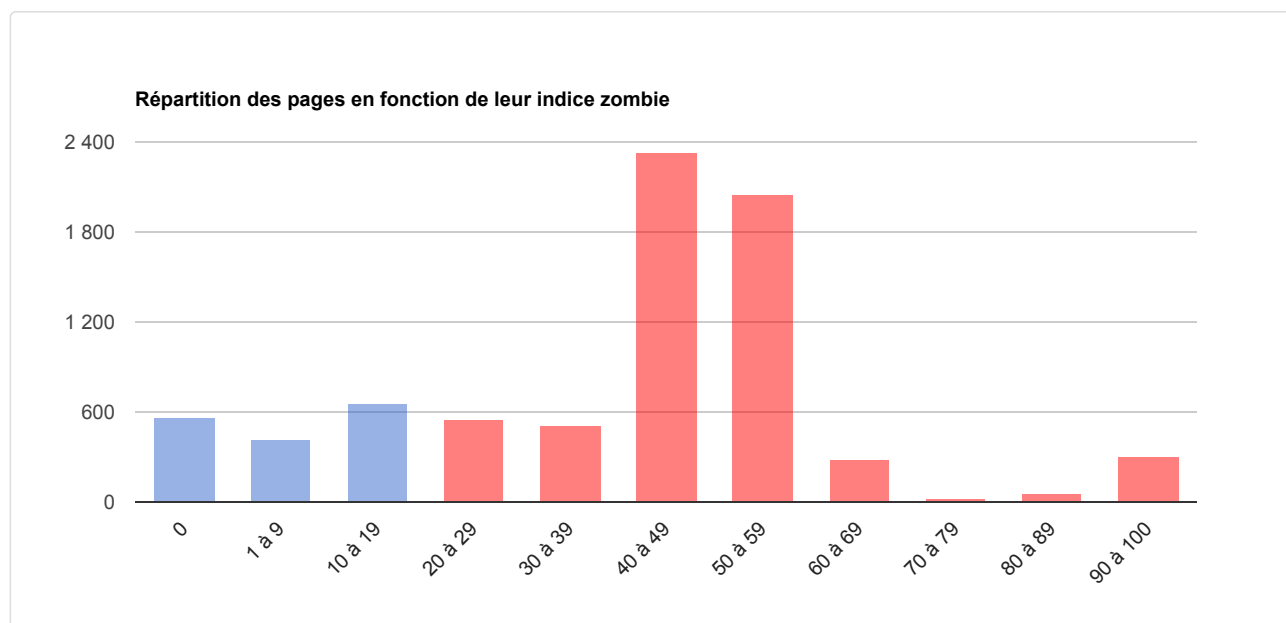
Avez-vous sur votre site des pages zombies ?

Une page zombie est une page qui non seulement cumule énormément de problèmes (et risque de décevoir l'internaute) mais aussi qui n'a pas de succès (génère pas ou très peu de visites SEO sur 1 an, est rarement consultée, rank très mal...).

L'indice Zombie de chaque page est calculé avec des valeurs entre 0 et 100, la pire valeur étant 100. **Ces valeurs sont disponibles dans l'annexe en conclusion.**

Au global, l'indice zombie moyen est de **40,6**.

L'histogramme ci-dessous montre la répartition des pages du site en fonction de leur indice Zombie :



Par expérience, nous vous conseillons d'éviter d'avoir trop de pages avec un indice Zombie supérieur ou égal à 20. Le tableau ci-dessous donne la synthèse :

	Erreurs	Avertissements	
<b>Nb de pages avec un indice Zombie &gt;= 20</b>	6 092	n/a	
<b>Nb de pages avec un indice Zombie entre 1 et 19</b>	n/a	1 068	
<b>Nb de pages avec un indice Zombie de 0</b>	n/a	n/a	554
<b>Nb de pages analysées</b>			7 714
<b>Taux d'erreur</b>	78,97 %	13,84 %	



Que faire des pages zombies ?

Plus l'indice zombie d'une page est élevé, plus il est probable qu'elle soit "irrécupérable".

Plus votre site a de pages à fort indice zombie, plus c'est risqué et inefficace :

- "risqué" car pour juger votre site, Google se base sur l'ensemble des pages que vous demandez de faire indexer. Laisser se faire indexer des pages "zombies" peut pénaliser vos bonnes pages
- "inefficace" car ces pages génèrent aucune ou très peu de visites et sont pas ou très peu consultées par vos visiteurs

Il est recommandé d'effectuer une analyse humaine suivant la méthode ci-dessous. Attention, le but est d'essayer de "sauver" le maximum de pages et de supprimer ou désindexer seulement en dernier recours.

Pour chaque page zombie identifiée, nous vous conseillons de tenter les solutions dans cet ordre (passez à une solution seulement si la précédente n'est pas applicable).

#### 1. Traitez les cas particuliers

- la page est-elle récente ? Si elle est indexée depuis quelques mois à peine, c'est peut-être normal qu'elle n'ait pas encore eu un grand succès, surtout quand on compare avec les autres pages étudiées sur 1 an.
- La page est-elle sur un sujet qui n'intéresse plus ? S'il s'agit d'une actualité, vérifiez que la date de publication de votre article est bien visible. Une analyse humaine s'impose ici pour savoir s'il faut conserver la page. Parfois une mise à jour est possible (sans modifier l'URL). A vous de décider si vous souhaitez conserver vos "archives".
- La page concerne-t-elle une niche très petite ? Si très peu de monde s'intéresse au sujet abordé, même sur 1 an, mais que le contenu est excellent et les (rares) conversions sont très rentables, alors vous pouvez décider de la conserver.

#### 2. Améliorez la page

Il s'agit essentiellement de son contenu, mais parfois aussi de critères techniques (vitesse, profondeur et maillage...). Complétez les informations, mettez-les à jour, vérifiez si vous ciblez les bons mots-clés et que la richesse sémantique du texte est supérieure à celle de vos concurrents dans Google, améliorez l'UX, ajoutez des liens vers d'autres pages internes, précisez vos sources (avec des liens sortants externes), mettez en avant l'auteur, illustrez par des images ou des vidéos, etc. Ajoutez-lui des liens entrants internes et faites-en la promotion sur les réseaux sociaux ou dans vos mailings.

#### 3. Regroupez des pages

Si plusieurs pages ont un contenu trop faible traitant d'un même sujet, regroupez-les. Choisissez la meilleure de ce groupe (cluster), fusionnez les contenus et reformatez le résultat final. N'oubliez pas de supprimer les autres pages du cluster et de faire des redirections 301 vers l'URL qui les regroupera. Parfois, aucune URL existante n'est adaptée pour les regrouper : vous pouvez en créer une toute nouvelle et rediriger toutes les autres vers elle. En aucun cas il faut faire des redirections massives vers une même URL, surtout n'ayant pas de contenu équivalent aux pages supprimées (l'erreur classique étant de rediriger en masse vers l'accueil).

#### 4. Désindexez la page

Si vous pensez que le contenu peut malgré tout intéresser des internautes ou générer des revenus, désindexez la page mais laissez-la en ligne. Sauf cas particulier, cela signifie que vous ne touchez pas au maillage interne : les liens vers la page restent en place. A vous d'évaluer si, une fois la page désindexée, il faut aussi la bloquer au crawl (fichier robots.txt).

#### 5. Supprimez (et désindexez) la page

Il s'agit ici d'une page irrécupérable ou d'une URL qui n'aurait jamais dû être indexée ("masse noire"). La page doit donc être supprimée (et désindexée). Si l'URL bénéficie de bons backlinks externes, faites une redirection 301 vers la meilleure page ayant un contenu sémantiquement proche. Sinon, votre serveur doit renvoyer un code 410. Si vous ne pouvez pas faire la 410, renvoyez un 404. Ca ne change pas grand chose mais c'est pour éviter de polluer votre rapport d'erreurs 404 dans Search Console notamment. Vérifiez que plus aucun lien interne ne pointe vers elle.

## 12.3 Pages les + rentables à travailler en premier

Parmi toutes les pages présentes sur votre site, en considérant l'ensemble des corrections d'erreurs et les optimisations possibles, par quoi devez-vous commencer pour améliorer votre référencement naturel ?

Nous savons très bien que votre temps (et vos ressources) est compté et vous devez vous concentrer sur le plus efficace : c'est pourquoi cette analyse repose sur un algorithme entièrement dédié à cela ! Basé sur des dizaines de critères et le fruit de nombreuses recherches, il détermine les pages les plus rentables à travailler, avec un indice de rentabilité pour chacune. Plus l'indice est élevé, plus la page est rentable à travailler. C'est-à-dire que le ratio "augmentation du trafic SEO" / "ressources à mettre en oeuvre (temps, argent, ...)" sera élevé.

Pour que ce soit efficace, il faut rester pragmatique : c'est pourquoi nous vous proposons seulement une sélection des pages de votre site. Quand vous aurez pu les travailler toutes, votre trafic SEO devrait avoir augmenté. Vous pourrez alors faire une nouvelle analyse pour étudier les nouvelles pages rentables à travailler, en tenant compte de toutes les évolutions (sur votre site mais aussi des internautes ou de Google).

Il est possible que certaines pages listées ici soient déjà performantes et qu'aucun problème n'ait été détecté dessus. Cela signifie que vous avez bien travaillé ! Et qu'il faut poursuivre avec les autres leviers : enrichissement de l'éditorial et des médias, amélioration de l'expérience utilisateur notamment sur mobile, obtention de backlinks de qualité, partages sur les réseaux sociaux, etc.

#### Annexe

Le fichier annexe `exemple_com_analytics-pages-rentables-a-travailler.xlsx` contient les données associées.

## 12.4 Analyse de la satisfaction utilisateur

L'objectif de Google est que les internautes soient satisfaits de leur recherche, et donc du site (de la page) qui leur est proposée. Son algorithme est donc logiquement centré sur la notion de qualité des pages proposées à l'internaute. Pour avoir un bon référencement aujourd'hui et de façon durable, vous devez absolument vous concentrer sur cette notion.

Grâce aux données récupérées dans Google Analytics, nous calculons par page un **indice évaluant la gravité que l'internaute ne soit pas satisfait**. Il est basé sur 2 types d'analyses :

1. satisfaction de l'internaute : semble-t-il satisfait par cette page (comparativement aux autres pages de ce site)
2. importance de la page : génère-t-elle beaucoup de visites en SEO ?

Cet indice prend des valeurs entre 0 (aucun problème) et 1000 (page la plus prioritaire à améliorer en termes de satisfaction utilisateur). Comme précédemment (Pages les + rentables à travailler en premier), **nous sélectionnons les pages les plus stratégiques à améliorer**.

L'analyse de la satisfaction utilisateur se base sur 2 indicateurs liés au comportement de l'internaute disponibles dans Google Analytics :

- le taux de rebond
- le nombre de pages vues par session

Nous avons choisi (pour l'instant ?) de ne pas nous baser sur les métriques liées à des durées, car leur mesure ne nous semble pas assez fiable ou représentative de l'ensemble de ce qui se passe sur un site.

Concernant les 2 indicateurs sélectionnés, nous considérons qu'il n'existe pas de valeur optimale universelle. Par exemple, le taux de rebond est forcément très différent sur un article d'un site d'actualité et sur une fiche produit d'un site e-commerce.

Par contre, nous avons remarqué qu'il était pertinent de comparer les pages entre elles, sur un même site. Ceci permet de repérer les pages qui semblent plus décevoir les internautes que d'autres pages.

En général, les pages qui cumulent ces 2 caractéristiques sont celles qui déçoivent le plus l'internaute :

- taux de rebond plus élevé que les autres
- nombre de pages vues par session plus faible que les autres

Notre indice final tient également compte du trafic : plus une page génère de trafic (SEO), plus c'est important qu'elle satisfasse l'internaute. Toute notre analyse est basée sur la notion de page de destination pour les visites générées par le référencement naturel.

Bien entendu, il ne s'agit que d'un indice : il vous aidera à repérer des problèmes parmi la masse des pages de votre site, mais au final c'est à vous de conclure si une page est décevante pour l'internaute.

Quelques explications sur les données fournies dans l'annexe :

- nous listons uniquement les pages HTML indexables pour lesquelles nous disposons d'assez de données fiables. Faites un audit avec la plus grande période Analytics possible.
- l'indice est un nombre entre 0 (semble satisfaire l'internaute) et 1000 (semble décevoir l'internaute)
- le nb de sessions est le nombre de sessions (visites) que la page a réussi à générer en référencement naturel sur la période étudiée
- le taux de rebond est le taux de rebond moyen de cette page, pour les sessions qu'elle a générées en référencement naturel sur la période étudiée
- le titre est le contenu de la balise title de la page

#### Annexe

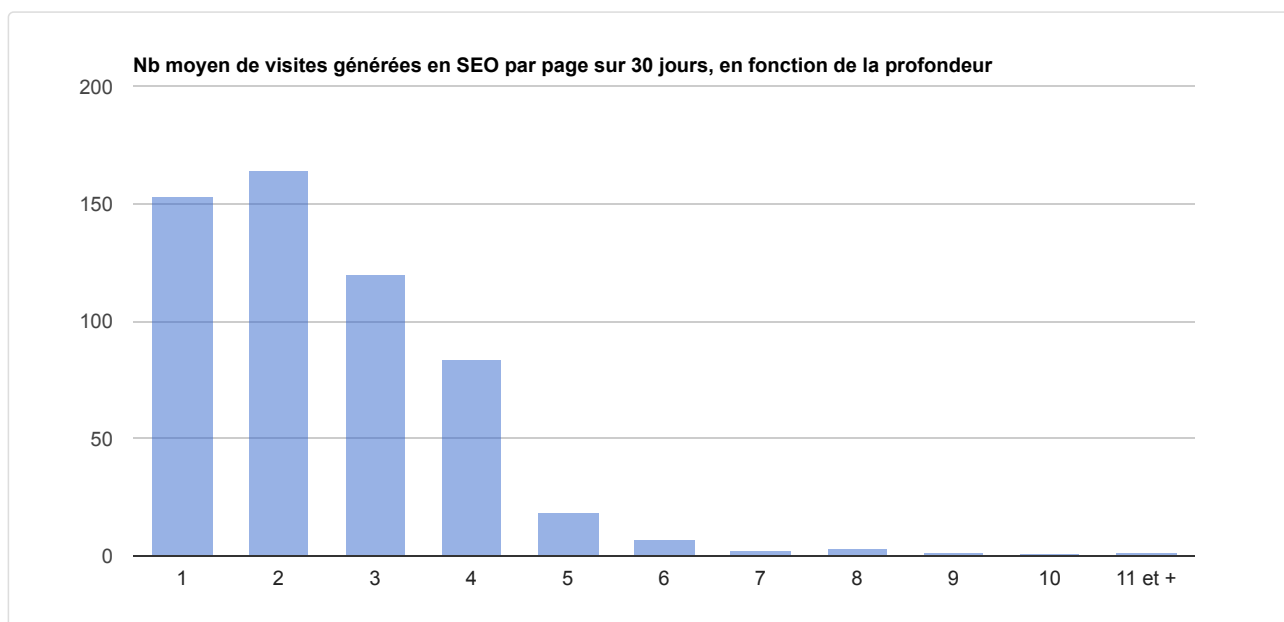
Le fichier annexe `example_com_analytics-satisfaction-utilisateur.xlsx` contient les données associées.

## 12.5 Autres analyses basées sur Google Analytics

### Analyse n°1 : profondeur

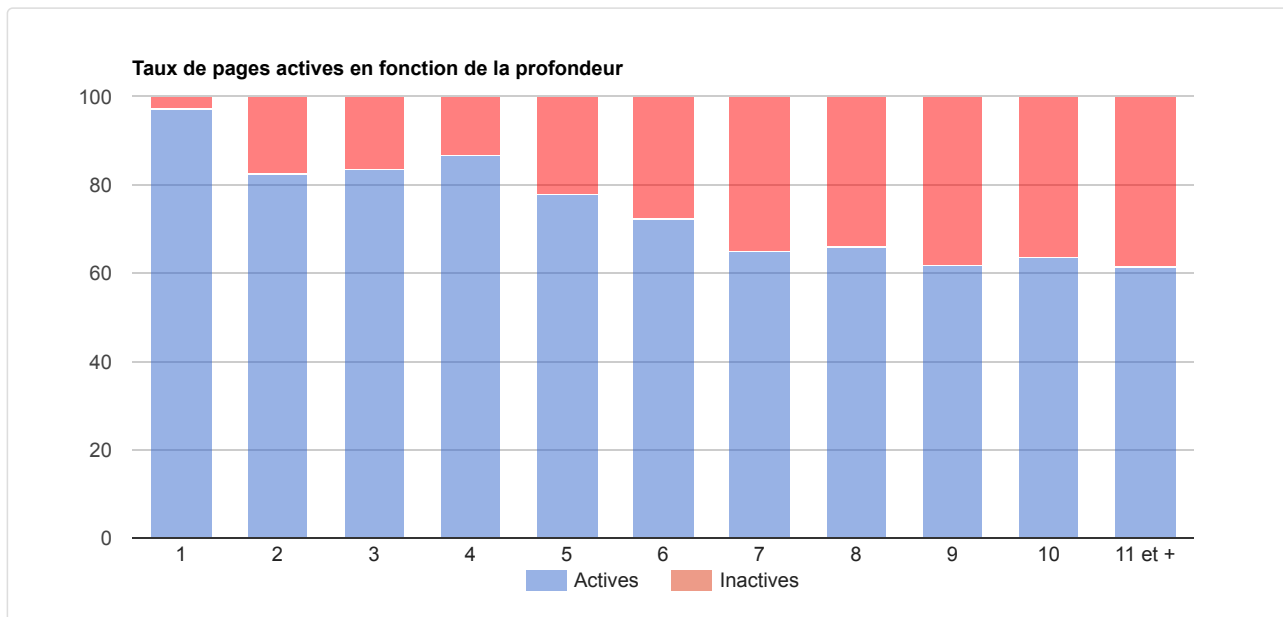
Comme indiqué précédemment dans ce rapport, il est préférable de limiter la profondeur de vos pages, surtout celles qui sont stratégiques.

Pour commencer, le plus important est de **mesurer les performances des pages** selon leur profondeur, c'est-à-dire le nombre de visites que chaque page arrive à générer en moyenne, via le référencement naturel. Pour faciliter les comparaisons, ce nombre est indiqué pour une période de 30 jours, quelle que soit la période d'analyse choisie pour l'audit.



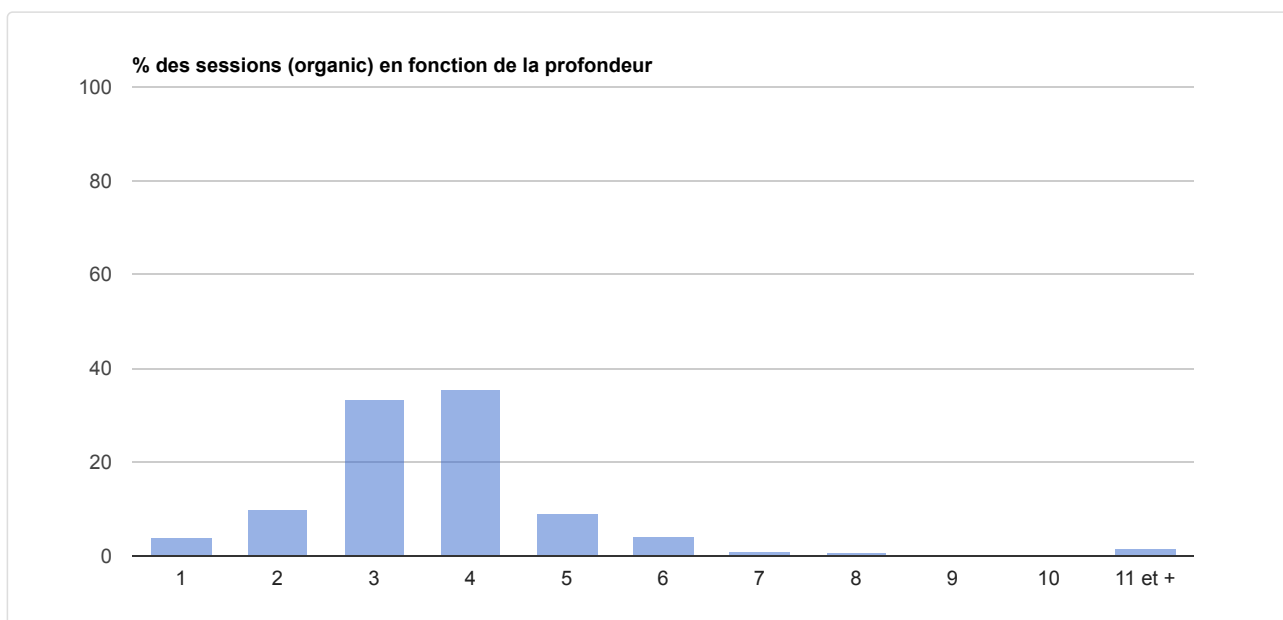
Le graphique qui suit indique le pourcentage de pages actives (en bleu) et de pages inactives (en rouge) par niveau de profondeur.

Attention : si dans un intervalle il y a peu de pages (actives + inactives), alors le taux de pages actives pour cet intervalle n'est pas significatif.



Pour compléter cette mesure du taux de pages actives, voici une étude de la répartition du trafic généré en référencement naturel (= "organic") selon le niveau de profondeur.

Attention : si vous avez beaucoup de pages trop profondes, elles peuvent malgré tout générer une part importante du trafic total, mais ce n'est pas pour autant recommandé ! Dans ce cas étudiez bien le 1er graphique ci-dessus.

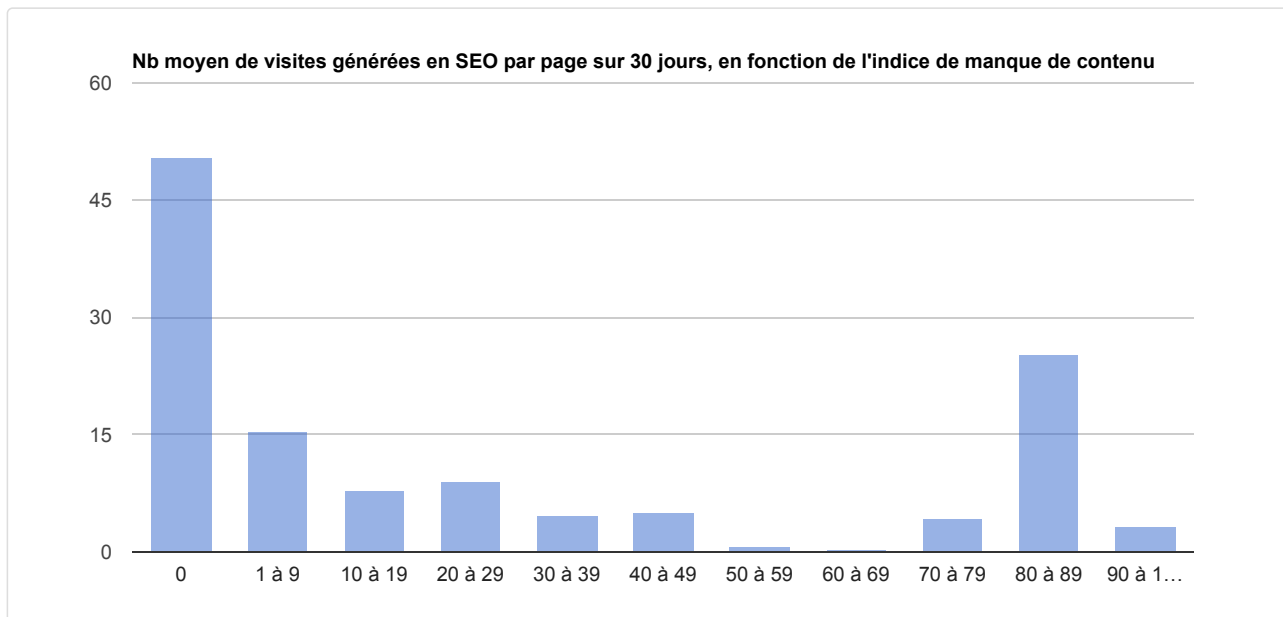


## Analyse n°2 : taille du contenu texte

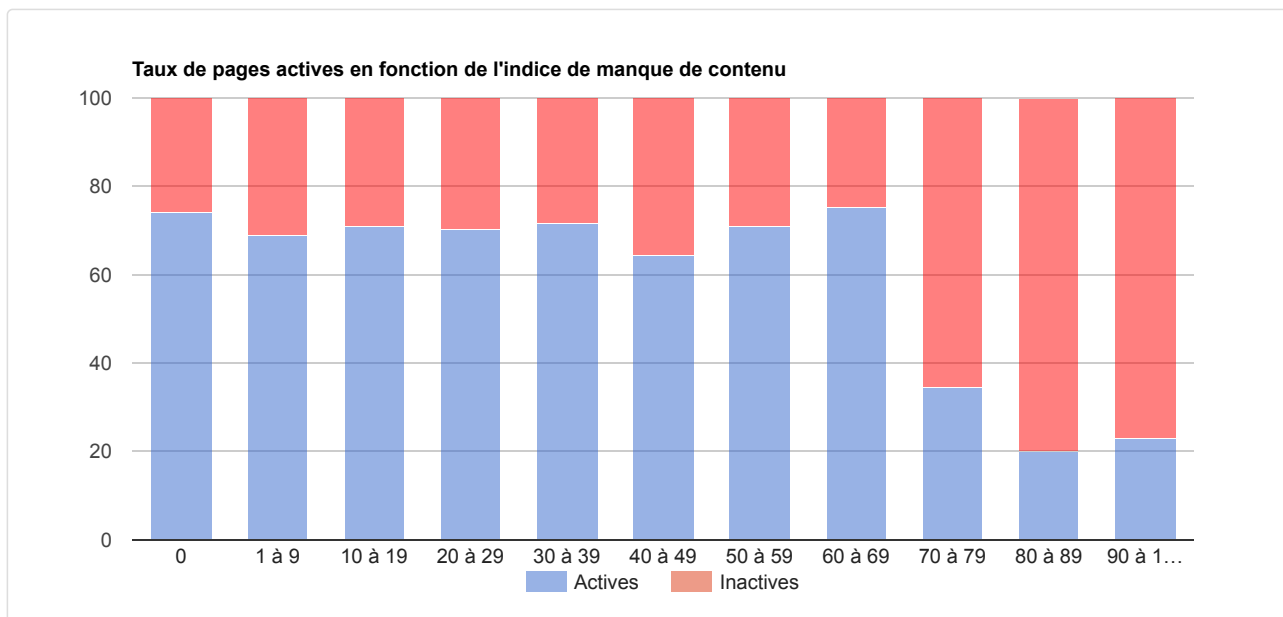
Comme indiqué précédemment dans ce rapport, les pages avec trop peu de contenu texte sont généralement moins efficaces en référencement naturel. Pour maximiser vos chances de succès, il vaut mieux bien travailler le contenu éditorial, surtout sur vos pages stratégiques.

Pour chaque page indexable, un indice de gravité du manque de contenu a été calculé : 0 signifie qu'aucun problème n'a été détecté et 100 est le maximum.

Pour commencer, le plus important est de **mesurer les performances des pages** selon leur manque éventuel de contenu, c'est-à-dire le nombre de visites que chaque page arrive à générer en moyenne, via le référencement naturel. Pour faciliter les comparaisons, ce nombre est indiqué pour une période de 30 jours, quelle que soit la période d'analyse choisie pour l'audit.

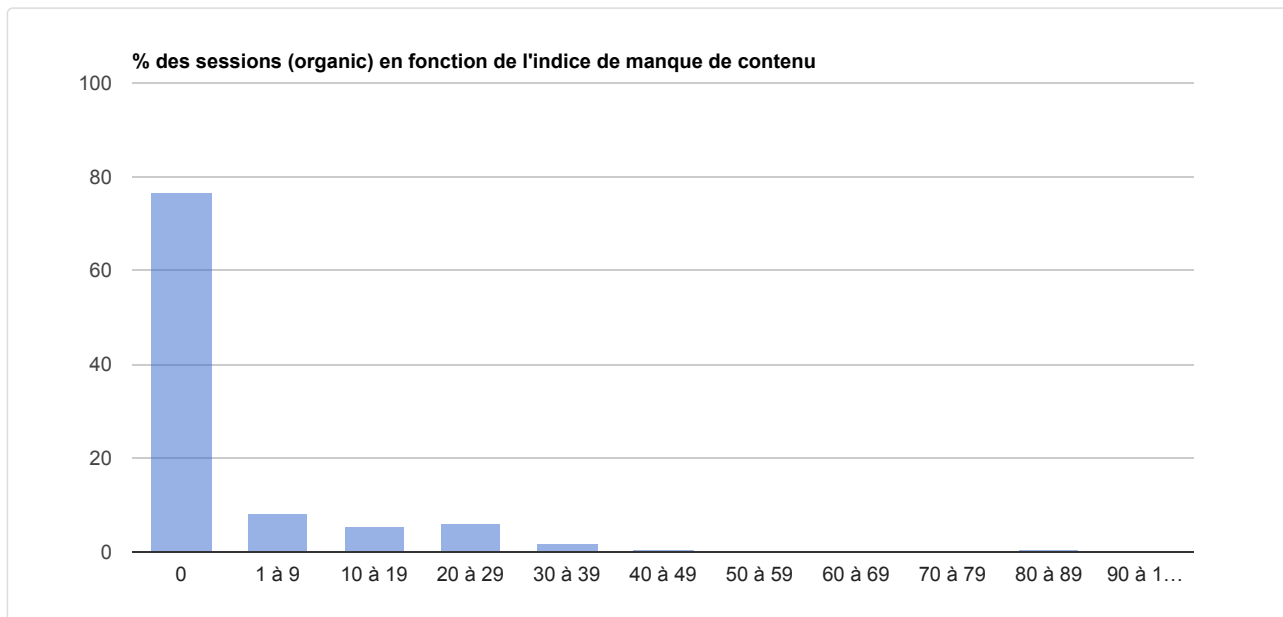


Le graphique qui suit indique le pourcentage de pages actives (en bleu) et de pages inactives (en rouge) selon cet indice de gravité :



Pour compléter cette mesure du taux de pages actives, voici une étude de la répartition du trafic généré en référencement naturel (= "organic") selon l'indice de manque de contenu

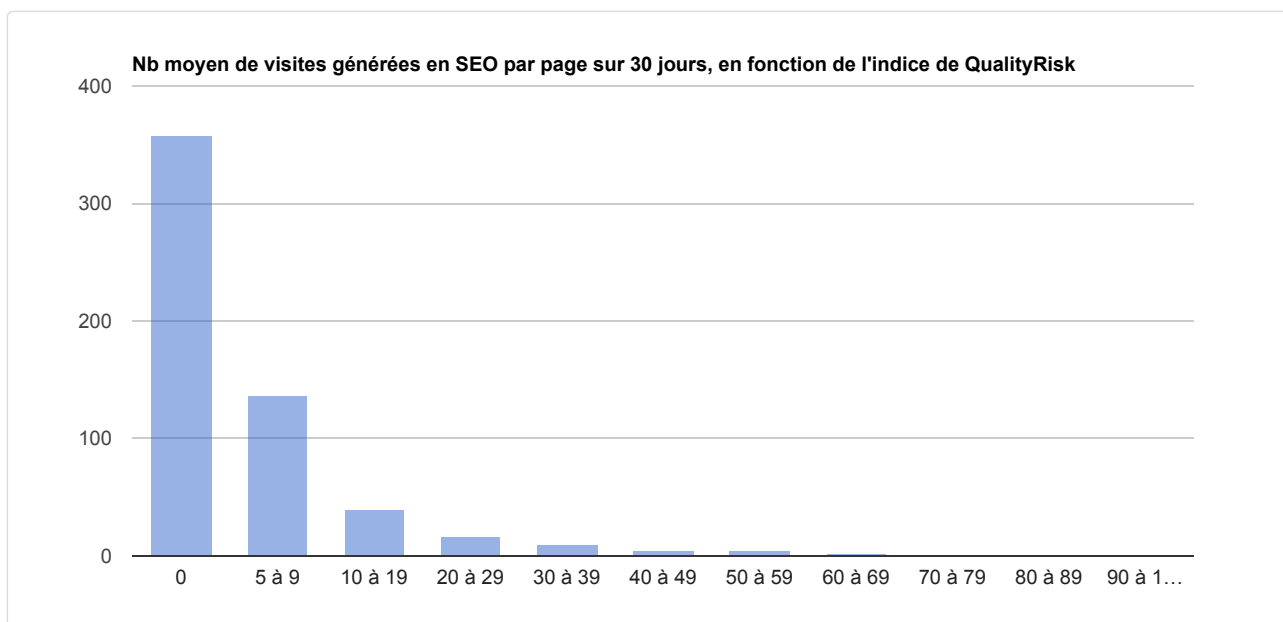
Attention : si vous avez beaucoup de pages manquant de contenu, elles peuvent malgré tout générer une part importante du trafic total, mais ce n'est pas pour autant recommandé ! Dans ce cas étudiez bien le 1er graphique ci-dessus.



### Analyse n°3 : QualityRisk

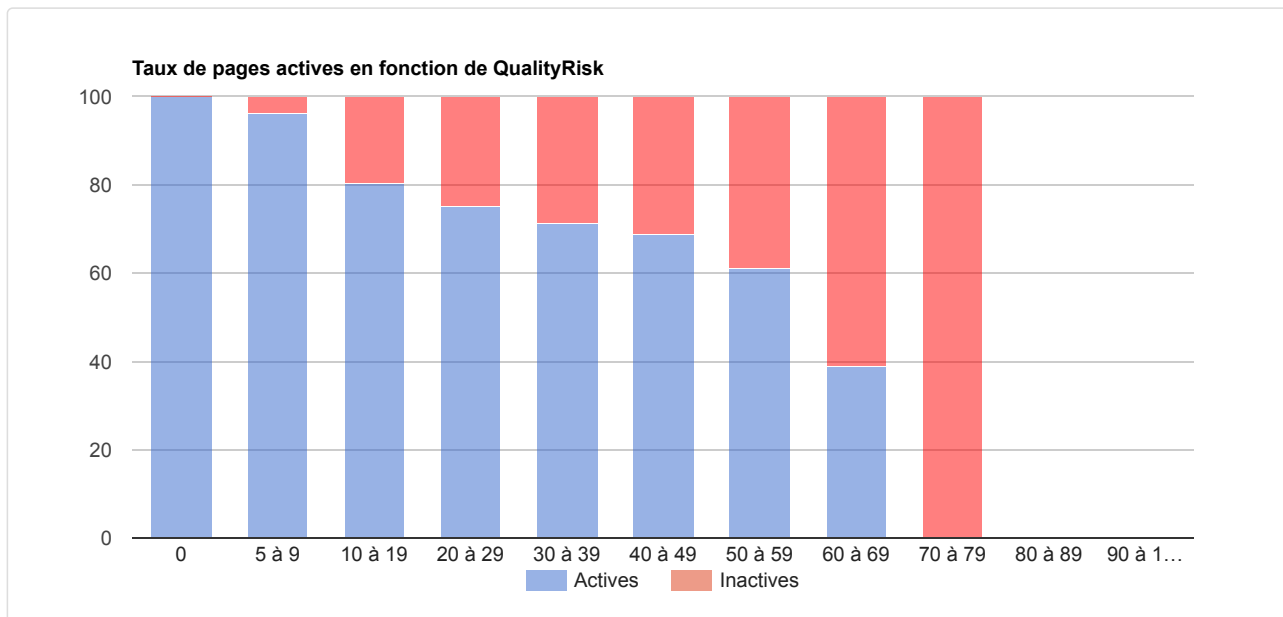
Comme indiqué précédemment dans ce rapport, pour chaque page indexable, son QualityRisk a été calculé. Plus il est élevé, plus la page risque de décevoir l'internaute et d'être pénalisante pour votre référencement naturel. Vous devriez améliorer la qualité de vos pages afin d'avoir un QualityRisk de 0, surtout sur vos pages stratégiques.

Pour commencer, le plus important est de **mesurer les performances des pages** selon leur QualityRisk, c'est-à-dire le nombre de visites que chaque page arrive à générer en moyenne, via le référencement naturel. Pour faciliter les comparaisons, ce nombre est indiqué pour une période de 30 jours, quelle que soit la période d'analyse choisie pour l'audit.



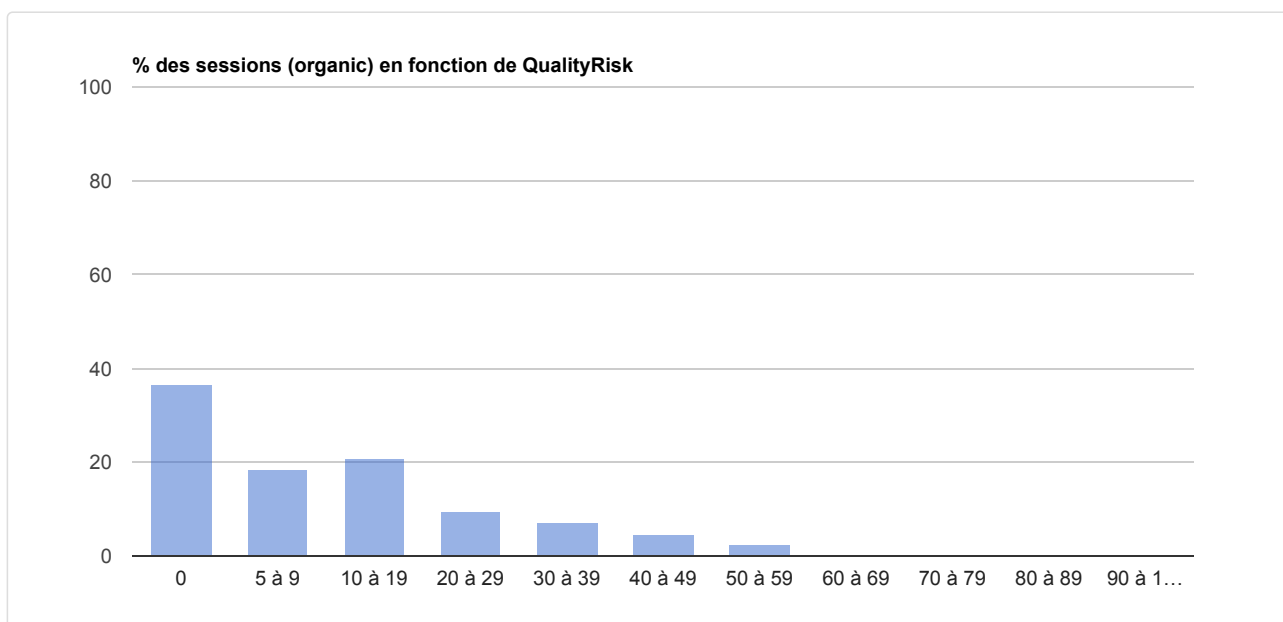
Le graphique qui suit indique le pourcentage de pages actives (en bleu) et de pages inactives (en rouge) selon QualityRisk.

Attention : si dans un intervalle il y a peu de pages (actives + inactives), alors le taux de pages actives pour cet intervalle n'est pas significatif.



Pour compléter cette mesure du taux de pages actives, voici l'étude de la répartition du trafic généré en référencement naturel (= "organic") selon QualityRisk.

Attention : si vous avez beaucoup de pages à trop fort QualityRisk, elles peuvent malgré tout générer une part importante du trafic total, mais ce n'est pas pour autant recommandé ! Dans ce cas étudiez bien le 1er graphique ci-dessus.



## 13 Analyse Google Search Console

En connectant le compte Google Search Console associé au site (dans les conditions rappelées dans l'introduction), il est possible de récupérer des données mesurant les performances SEO des pages du site dans les pages de résultats de recherche (partie référencement naturel).

Non seulement nous fournissons des données brutes (plus complètes que celles disponibles dans l'interface web de Search Console), mais aussi les analyses avancées décrites ci-après.

Toutes les analyses qui suivent sont basées uniquement sur les pages indexables découvertes par RM Tech lors du crawl.

### 13.1 Performances du site dans les SERP de Google

Grâce au couplage avec Google Search Console, nous récupérons pour chaque page indexable du site les données suivantes :

- **nombre d'impressions** : nombre de fois où la page a été affichée dans les résultats quand un internaute a fait une recherche Google
- **nombre de clics** : nombre de fois où l'internaute a cliqué sur un résultat menant à la page
- **taux de clic** : (nombre de clics) / (nombre d'impressions) en pourcentage. En anglais on parle de Click Through Rate (CTR)
- **position moyenne** : position moyenne obtenue par la page pendant la période étudiée, pour l'ensemble des fois où elle a été affichée dans les résultats

Contrairement à l'interface web qui vous limite à 1000 lignes, **nous récupérons toutes les données**, pour l'ensemble des pages indexables ayant eu des impressions. C'est un avantage énorme apporté par l'utilisation de l'API officielle.

Sur la période étudiée, avec les paramètres de crawl du site et les paramètres Google Search Console rappelés dans l'introduction, **4 166 541 clics** ont été générés par **4789 pages** (autres que l'URL de départ) via le référencement naturel sur Google.

L'URL de départ a généré 179 176 clics.

Voici les détails (en excluant l'URL de départ, les pourcentages étant exprimés par rapport au nombre de pages indexables autres que l'URL de départ) :

- Nb de clics : **4 166 541** sur les 365 jours de la période, soit en moyenne **342 455** pour 30 jours
- Nb de pages ayant généré des clics : **4 789** soit **62,08 %**
- Nb de pages ayant généré des impressions mais aucun clic : **2 651** soit **34,37 %**
- Nb de pages ayant généré ni clic ni impression : **274** soit **3,55 %**

Retrouvez ces données dans l'annexe située [dans la conclusion de l'audit](#).

### 13.2 Pages du site pouvant passer de la page 2 à la page 1 des résultats Google

En référencement naturel sur Google, passer de la position #2 à la position #1 est souvent très difficile et nécessite de progresser sur de nombreux critères. Le rapport (temps passé ou ressources utilisées) / (hausse de trafic) n'est donc pas toujours très bon.

Par contre, passer de la page 2 à la page 1 est très rentable. En effet, très peu d'internautes consultent la page 2, donc arriver en position #11 ou #12 génère très peu de visites. Pourtant, gagner 1 place de #11 à #10 est généralement bien plus simple que de #2 à #1.

Attention : il est inefficace de se baser uniquement sur Search Console pour tenter de faire passer d'une position moyenne 11 à 10. En couplant aux faiblesses de la page en termes de référencement, c'est bien plus efficace.



C'est pourquoi notre algorithme spécialement conçu pour cela repère les pages qui présentent le meilleur potentiel à ce niveau. Il est basé sur de nombreux critères issus aussi bien de Search Console que de l'audit RM Tech, dont les principaux sont les suivants :

- le nombre d'impressions : plus la page est affichée dans les résultats de Google, plus elle a du potentiel.
- la position moyenne : il faut chercher des pages qui s'affichent généralement en haut de la page 2. Si elles sont trop loin, ça risque d'être trop dur de les faire progresser.
- la marge de progression SEO : plus il est facile d'améliorer la page (balises ou contenu) ou son maillage interne (profondeur, liens internes), plus le potentiel est élevé.

Pour vous éviter d'être noyé sous une masse de pages à optimiser, nous sélectionnons seulement celles qui présentent le meilleur potentiel. Il est mesuré avec un indice (de 1 à 1000, le maximum possible).

Pour chaque page sélectionnée, l'annexe contient des recommandations sur ce qu'il faut précisément faire. Concrètement, elle contient les informations suivantes :

- URL de la page
- indice : potentiel à passer en page 1, le meilleur étant 1000
- explications : ce qu'il faut améliorer pour la page
- nombre d'impressions selon GSC
- nombre de clics selon GSC
- CTR : taux de clics dans les SERP selon GSC
- position moyenne selon GSC
- gravité du problème de taille du contenu (le pire étant 100)
- balise title
- balise meta description

Remarque : si vous souhaitez mettre au point votre propre algorithme pour sélectionner autrement les pages, vous disposez dans l'annexe en conclusion de l'ensemble des données récupérées ou calculées par RM Tech.

#### Annexe

Annexe Pages qui pourraient passer en page 1 : example\_com\_search-console-potentiel-premiere-page.xlsx

## 14 Conclusion

Avoir toutes les données mesurées et calculées par RM Tech sur les pages indexables est très précieux pour faire des analyses encore plus avancées.

Grâce aux options Google Analytics et Google Search Console, les performances des pages (trafic généré en SEO et statistiques dans les SERP) ont pu être récupérées et agrégées avec les données du crawl et celles issues de l'analyse.

### Annexe

Le fichier annexe `exemple_com_conclusion.xlsx` contient les données associées.

L'annexe ci-dessous liste toutes les URL crawlées avec pour chacune la raison de ce crawl (la plupart du temps il s'agit d'un lien mais ça peut aussi être une URL canonique ou une redirection par exemple). Dit autrement, cette annexe liste tous les backlinks internes follow de toutes les URL étudiées, avec l'ancre de chacun.

### Annexe

Le fichier annexe `exemple_com_conclusion-maillage-interne.xlsx` contient les données associées.

Vous trouverez ci-après une liste des problèmes techniques du référencement en cours sur votre site (étudié dans les conditions précisées dans l'introduction) et qu'il faut corriger au plus vite.

### 14.1 Problèmes de crawl et d'indexation

Il faut éviter de faire des liens internes vers des URL qui répondent en erreur. À la date de notre analyse, votre taux d'erreur était de 1,65 % (détails).

Il faut éviter de faire des liens internes vers des URL non indexables. À la date de notre analyse, votre taux d'erreur était de 23,61 % (détails).

### 14.2 Problèmes de temps de téléchargement

Nous vous conseillons de faire en sorte que chaque page soit téléchargée en 600 ms maximum. À la date de notre analyse, nous avons identifié que 66,60 % des URL HTML code 200 sont trop lentes, ce qui n'est pas optimal (détails).

### 14.3 Problèmes de profondeur des pages ou de maillage interne

Il faut que la majorité des URL soit à une profondeur inférieure ou égale à 4. À la date de notre analyse, nous avons identifié que 84,30 % des URL indexables sont trop profondes (5 ou plus), ce qui n'est pas optimal (détails).

Il faut aussi que les pages reçoivent plusieurs liens entrants internes, avec des ancres distinctes. Notre analyse a montré que 54,34 % des pages n'ont pas des liens entrants internes assez efficaces.

### 14.4 Problèmes concernant le contenu

Dans un premier temps, nous faisons le point sur le plus important, à savoir l'optimisation du contenu des pages HTML indexables. Le référencement étant tellement concurrentiel, vous devriez vous débrouiller pour n'avoir aucun frein sur votre propre site. Voici les éléments techniques que vous devez corriger au plus vite, classés par taux d'erreur décroissant :

Élément d'analyse	Taux d'erreur
contenu (détails)	65,12 %
balise meta description (détails)	1,17 %
balise title (détails)	0,43 %

Pour réussir votre référencement, il faudrait donc dans l'idéal qu'aucune de vos pages n'ait de sous-optimisation. Nous avons calculé un taux d'erreur global en identifiant toutes les pages HTML indexables qui contiennent une ou plusieurs erreurs (concernant les éléments techniques du tableau précédent).

<b>Nb de pages ayant au moins une erreur technique</b>	5 034
<b>Nb de pages indexables</b>	7 715
<b>Taux d'erreur global</b>	<b>65,25 %</b>

Plus vous réussirez à réduire ce taux d'erreur global, plus vous vous donnerez de chances de réussir un bon référencement. Évidemment, cela ne suffit pas, il faut par exemple :

- obtenir régulièrement des liens de qualité ;
- avoir une bonne présence dans les réseaux sociaux ;
- que le comportement des internautes sur votre site montre qu'ils en sont satisfaits.

## 14.5 Problèmes sur les liens sortants externes

À la date de notre analyse, votre taux d'erreur était de 48,63 % (détails).

## 14.6 Les problèmes de masse noire

En plus de ces éléments techniques assez connus, nous avons évoqué au début de ce rapport la problématique de la "masse noire". C'est ainsi que nous appelons les URL qui sont crawlées et même souvent indexées par Google alors qu'elles ne devraient pas être indexées, ou même pas exister du tout. Vous pouvez également avoir le problème inverse, à savoir que Google n'indexe qu'une partie de vos pages.

Dans les deux cas c'est un problème que vous devriez corriger. Pour savoir si vous êtes concerné, suivez les étapes suivantes :

### Étape 1 : Identification de vos contenus

Si votre site contient une boutique en ligne :

- combien avez-vous de catégories et sous-catégories dans votre catalogue ?
- combien avez-vous de fiches produits actuellement en vente ?
- combien avez-vous de pages pour d'autres types de contenus (page marque, page catégorie filtrée selon un ou plusieurs critères, etc.) ?

Si votre site contient une rubrique éditoriale (blog, actualités, conseils) :

- combien avez-vous de catégories et sous-catégories ?
- combien avez-vous d'articles publiés ?

Si votre site contient un forum :

- combien avez-vous de catégories et sous-catégories de forums ?
- combien avez-vous de discussions en ligne ? parmi elles, combien s'étalent sur plusieurs pages ?

Concernant les autres pages de votre site :

- à part la page d'accueil, combien avez-vous de pages présentant votre entreprise et ses activités ?
- combien avez-vous de pages spécifiques type mentions légales, contact, etc. ?

### Étape 2 : Calcul du nombre de pages à indexer

Faites la somme de tous les éléments identifiés à l'étape 1 : il s'agit du nombre de pages que vous souhaitez faire indexer dans les moteurs de recherche. Nous l'appelons Nombre de pages à indexer.

### Étape 3 : Analyse

Comparez ce Nombre de pages à indexer au Nombre de pages indexables indiqué dans notre rapport (il était égal à 7 715 le jour de notre analyse) :

- si votre nombre de pages à indexer est égal à 7 715, alors vous n'avez aucun problème.
- si votre nombre de pages à indexer est inférieur à 7 715, alors vous avez sans doute des problèmes de masse noire et donc une sous-optimisation en termes de crawl et d'indexation. Google perd son temps (et ses ressources) à crawler et parfois à indexer des pages de votre site qui n'auraient pas dû l'être, au détriment du temps passé pour vos pages stratégiques.
- si votre nombre de pages à indexer est supérieur à 7 715, alors vous avez des problèmes d'indexation. Certains de vos contenus n'étant pas indexés, vous n'exploitez pas tout le potentiel de votre site dans Google. Cela peut être dû à de nombreux facteurs notamment des erreurs techniques, un mauvais maillage interne, une notoriété générale de votre site trop faible, des contenus dupliqués.

Pendant notre formation SEO, nous expliquons longuement comment éviter et corriger tous ces problèmes.

## 15 Pour aller plus loin

---

Voici de quelle façon nous pouvons vous accompagner dans le succès pour votre référencement.

### 15.1 Quand refaire un audit SEO ?

Ce type de rapport d'audit vous aide à identifier et à corriger les problèmes techniques qui freinent votre référencement naturel. Afin de détecter au plus vite les prochains problèmes techniques, nous vous conseillons de faire une nouvelle analyse :

- après avoir modifié ou mis à jour votre CMS
- après avoir ajouté de nouvelles rubriques ou pages importantes dans le site
- après avoir modifié le maillage interne
- après avoir ajouté des balises dans vos modèles de pages (exemple : URL canoniques)
- après avoir modifié les règles de crawl (robots.txt) ou d'indexation (balise meta robots)
- après des mises à jour de Google (algorithmiques ou d'interface de ses pages de résultats, par exemple la taille du titre)
- après avoir reçu une pénalité Google pour liens **sortants** externes artificiels (ou avant, pour l'anticiper)

D'une manière générale, refaire un audit SEO vous permet de surveiller et suivre l'avancée du travail effectué par vos équipes et/ou vos prestataires. Vous savez que vous pouvez compter sur notre expertise ainsi que sur l'exhaustivité et la fiabilité de notre outil indépendant.

### 15.2 Formation

Organisme de formation déclaré depuis 2006, nous avons formé plus de 4000 entreprises à de nombreuses composantes du webmarketing. Retrouvez le catalogue de formation sur [notre site formation](#)

### 15.3 Accompagnement personnalisé

Si vous avez besoin d'un accompagnement par un expert, vous pouvez :

- Acheter des heures de consulting (par e-mail ou téléphone) avec l'un de nos experts SEO. Cela peut servir à poser des questions sur le rapport d'audit SEO et aussi à vos questions liées au SEO (voir les détails de cette offre [ici](#)).
- Acheter des audits détaillés de template. Contrairement à l'audit SEO qui est automatisé, l'audit de modèles de pages est réalisé à la main par nos experts SEO (voir les détails de cette offre [ici](#)).

## 16 Explications complémentaires

### 16.1 Fonctionnement de nos outils

Grâce aux outils spécifiques développés dans notre centre de recherche, nous avons pu faire une analyse exhaustive de votre site, page par page, en partant de la page d'accueil, comme Google peut le faire.

Au fil des années, nous avons pu mettre au point un outil qui reproduit de manière assez fidèle le fonctionnement de Googlebot (le robot de Google). Par exemple :

- Nous respectons le fichier robots.txt de la même façon que Google (avec donc quelques particularités par rapport au standard)
- Nous tenons compte également :
  - des balises meta robots
  - des directives passées dans l'entête HTTP (pour l'URL canonique ou pour les directives d'indexation et de suivi des liens)
- Nous ne gérons pas :
  - les cookies
  - le javascript (et ajax ou équivalent)
  - le CSS
  - le fait d'avoir une version mobile d'une page à une URL distincte (les URL mobiles peuvent être crawlées mais risquent d'être indiquées comme non indexables en raison de leur URL canonique)
- Nous ne suivons pas :
  - les formulaires
  - les redirections meta
  - les liens nofollow
- Les images et les vidéos sont exclues de l'analyse
- Pour calculer le nombre de liens entrants internes d'une page (backlinks) :
  - nous cherchons les pages (indexables ou pas) qui ont une balise <a> qui amène au final vers la page qu'on étudie (soit de façon directe, soit via une ou plusieurs redirections)
  - si une page A fait un lien vers elle-même, cela ne compte pas comme un backlink
  - si une page A fait un lien vers B qui redirige vers C, alors A est comptée comme un backlink de C
  - si une page A fait plusieurs liens vers une même page B, alors cela compte comme un seul backlink. Attention : dans ce cas, seul le 1er lien est pris en compte (à condition qu'il soit follow)
  - l'URL canonique n'est pas prise en compte ici, uniquement les liens avec une balise <a href>

### 16.2 Signification des codes HTTP

Voici les principaux codes HTTP (retrouvez la liste complète [sur Wikipédia](#)).

Les principales familles de codes :

- Codes 2XX (commençant par "2") : succès
- Codes 3XX (commençant par "3") : redirection
- Codes 4XX (commençant par "4") : erreur du client
- Codes 5XX (commençant par "5") : erreur du serveur

Les codes les plus courants :

- 200 = *OK* : l'URL a été trouvée
- 301 = *Moved Permanently* : l'URL a été déplacée de façon permanente (redirection 301)
- 302 = *Moved Temporarily* : l'URL a été déplacée de façon temporaire (redirection 302)
- 400 = *Bad Request* : la syntaxe de la requête est erronée
- 401 = *Unauthorized* : une authentification est nécessaire pour accéder à la ressource
- 403 = *Forbidden* : accès interdit (droits d'accès insuffisants)
- 404 = *Not Found* : l'URL est introuvable
- 410 = *Gone* : la ressource est indisponible et aucune adresse de redirection n'est connue (en général, c'est le cas des documents supprimés définitivement)
- 500 = *Internal Server Error* : erreur interne du serveur

- 503 = *Service Unavailable* : serveur temporairement indisponible ou en maintenance
- 504 = *Gateway Time-out* : temps d'attente d'une réponse d'un serveur à un serveur intermédiaire écoulé

## 16.3 Signification des types MIME

Voici les principaux types MIME, c'est-à-dire les formats de données des documents situés sur un site Internet (retrouvez la liste complète [sur Wikipédia](#)) :

- text/css : CSS (feuilles de styles)
- text/csv : CSV (données séparées par une virgule)
- text/html : HTML
- text/plain : texte brut
- text/xml : XML
- image/bitmap : image au format bitmap
- image/gif : image au format GIF
- image/jpeg : image au format JPEG
- image/png : image au format PNG
- image/svg+xml : image au format SVG
- image/tiff : image au format TIFF
- application/javascript : JavaScript
- application/json : JSON (JavaScript Object Notation)
- application/msword : Microsoft Word
- application/pdf : PDF
- application/xhtml+xml : XHTML
- application/xml : XML
- application/x-shockwave-flash : Flash
- application/zip : ZIP

## 16.4 Signification des termes SEO

Voici les principaux termes utilisés dans le jargon du référencement naturel (SEO) et notamment dans ce rapport d'audit. Vous trouverez des liens vers des articles qui détaillent les notions correspondantes :

- **HTTP et HTTPS** : HTTP est le protocole standard d'accès à des documents sur le web, tandis que HTTPS est la version sécurisée (transmission cryptée). Une URL en HTTP et la même en HTTPS sont considérées comme 2 URL distinctes.
- **Google Search Console** : interface proposée par Google pour fournir au webmaster plusieurs informations concernant son site dans Google (problèmes de crawl ou d'indexation, pénalités, etc.). En savoir plus : [site officiel](#), [dossiers WebRankInfo](#). Bing propose la même chose (voir [ici](#)).
- **URL indexable** : il s'agit d'une URL qui ne présente aucun obstacle technique pour que Google l'indexe. Il faut qu'elle ne soit pas bloquée au crawl (dans le fichier robots.txt) et qu'elle soit accessible (donc que le code HTTP renvoyé par le serveur soit 200) et qu'elle ne soit pas interdite d'indexation (via la balise meta robots noindex ou un entête HTTP X-Robots-Tag) et qu'elle ne contienne pas d'URL canonique différente. En général, une URL indexable est indexée par Google (mais ce rapport ne le vérifie pas). En savoir plus : [dossiers WebRankInfo](#).
- **URL canonique** : il s'agit de l'URL officielle d'une page. On l'utilise surtout dans le cas où un même contenu est disponible à plusieurs URL (dans ce cas pour ces différentes URL on indique quelle est l'URL officielle). En savoir plus : [aide Google](#), [dossiers WebRankInfo](#).
- **Fichier robots.txt** : situé à la racine de chaque sous-domaine, il liste les éventuels endroits du site pour lesquels tous ou certains robots n'ont pas le droit de venir crawler. En savoir plus : [aide Google](#), [spécifications Google](#), [dossiers WebRankInfo](#).
- **Balise meta robots** : cette balise permet de restreindre les robots tels que Googlebot pour leur interdire par exemple d'indexer la page (noindex) ou de tenir compte des liens présents sur la page (nofollow). En savoir plus : [aide Google](#), [dossiers WebRankInfo](#).
- **En-tête HTTP X-Robots-Tag** : pour les documents non HTML, il n'est pas possible d'utiliser une balise meta robots. On peut dans ce cas passer des instructions par l'entête HTTP. En savoir plus : [aide Google](#), [dossiers WebRankInfo](#).
- **Redirection (301 ou 302)** : on parle de redirection d'une URL A vers une URL B quand, en cherchant à accéder à A, le serveur nous renvoie vers B. En référencement, il faut privilégier la redirection permanente

(code HTTP 301) dans la majorité des cas. Il existe de nombreux moyens de gérer les redirections. En savoir plus : [aide Google](#), [dossiers WebRankInfo](#).

- **Sitemap** : un fichier sitemap liste des URL que l'on souhaite faire crawler par les moteurs de recherche (en général pour qu'ils indexent les URL concernées). Ce n'est pas obligatoire mais ça peut aider à accélérer le crawl ou dans le suivi du référencement. En savoir plus : [aide Google](#), [dossiers WebRankInfo](#).
- **Lien nofollow** : si une page A fait un lien nofollow vers une page B, alors les moteurs de recherche feront comme si ce lien n'existe pas. En d'autres termes, la page B ne bénéficie d'aucun avantage (ou inconvénient) de recevoir un lien depuis A. Google impose que les liens commerciaux soient en nofollow, par le biais d'un attribut rel=nofollow sur la balise A (liens vendus, liens affiliés). En savoir plus : [aide Google](#), [dossiers WebRankInfo](#).