

DIGITAL SEEDeR

Audit SEO

Site analysé : [parisitour.fr](https://www.parisitour.fr)

Digital Seeder

26 rue de la république, 95650 Boissy l'Aillie France

<https://www.digitalseeder.com/>

N° tél. : 06 86 93 88 39

N° RCS : 81993240100014 - N° TVA : FR17819932401

Score SEO



Très forte marge de progression potentielle !

Meilleure note possible : 100

Score performance SEO



Très forte marge de progression potentielle !

Meilleure note possible : 100

L'audit du site a eu lieu le **31 janvier 2024** et a porté sur **243 URL** ainsi que sur **238 images distinctes** (238 images sont hébergées sur parisitour.fr, aucune sur d'autres sites).

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 : parisitour.fr
- URL de départ : <https://parisitour.fr/>
- Protocoles : aucune restriction (HTTP et HTTPS)
- Taille maximale du code HTML à télécharger : 700 Ko
- Nombre maximal d'URL à explorer : 2 000
- Motif d'inclusion : aucun
- Motif d'exclusion : .jpg OR .jpeg OR .png OR .gif OR .webp
- User-Agent pour le site : Mozilla/5.0 (Linux; Android 10) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/111.0.5563.57 Mobile Safari/537.36
- User-Agent pour le robots.txt : Googlebot
- Vitesse de l'exploration : 1.0 requête par seconde
- Ignorer le robots.txt : non
- Ignorer le contenu dans les formulaires pour l'évaluation de la taille du contenu : oui
- Suivre les liens internes nofollow : non
- Suivre les liens externes nofollow : oui
- Motif d'exclusion pour les liens externes : aucun
- Limiter l'exploration au domaine parisitour.fr : oui
- CMS déclarés : WordPress
- Analyse des images : oui
 - Nb maximal d'images à explorer : 1 000
 - 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
- Paramètres d'URL utilisés pour la pagination : aucun
- Regex d'URL pour la pagination : aucune
- Données Google Search Console : compte "digitalseeder@gmail.com", propriété <https://parisitour.fr/>, sur tous types d'appareil, sur les 365 derniers jours (du 30 janvier 2023 inclus au 29 janvier 2024 inclus)

Les données sont disponibles sous forme exhaustive dans des fichiers annexes au format Excel (et sur demande au format TSV Tab-separated values).

Annexes

Les annexes sont disponibles dans le fichier ZIP annexes-parisitour_fr-2024-01-31.zip.

1 Synthèse des résultats

1.1 Scores SEO

Score technique SEO :



- Il représente le niveau d'optimisation technique du référencement naturel des pages crawlées, basé sur l'indice QualityRisk ainsi que le crawl et l'indexabilité. Meilleure note possible : 100.
- Concernant QualityRisk : Indice moyen = 36,9. Taux d'erreur = 78,7 % (177 URL ont un QualityRisk > 20). [Voir les détails](#).

Score performance SEO :



- Il représente le niveau de performance des pages crawlées, basé sur l'indice Zombie. Meilleure note possible : 100.
- Concernant Indice Zombie : Indice moyen = 58,6. Taux d'erreur = 99,6 % (223 URL ont un indice Zombie > 20). [Voir les détails](#).

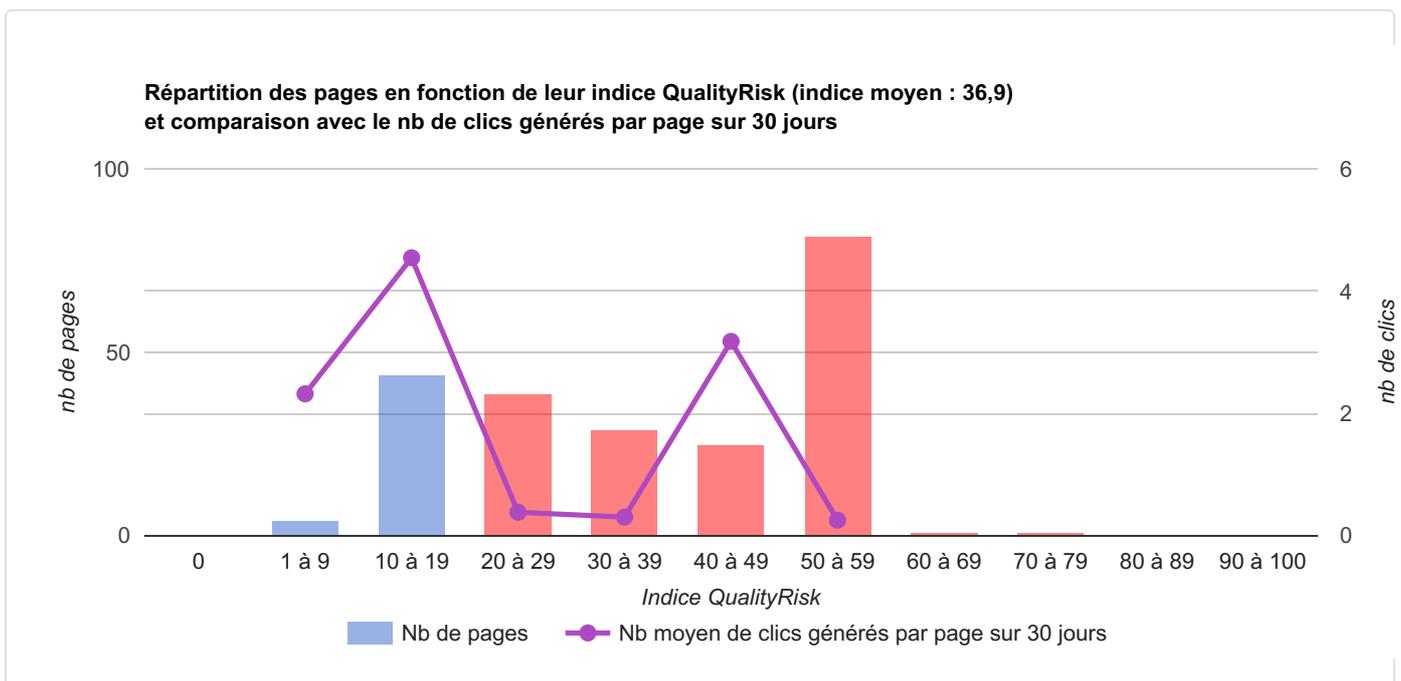
Pour ces 2 scores, il est recommandé d'atteindre 95 (score en vert) ; entre 70 et 95 le score est en jaune et en-dessous de 70 il est en rouge.

Grâce aux données de Google Search Console, les performances des pages ont pu être récupérées et utilisées dans tout l'audit. La principale métrique est le trafic organique généré par chaque page (via le référencement naturel Google), exprimé en nombre de visites (ou "clics"). Les pages qui ont généré des clics sont dites "actives" et les autres "inactives".

L'histogramme ci-dessous montre la répartition des pages du site en fonction de leur indice QualityRisk. Comme il est important de voir en même temps si les pages sont performantes (en mesurant le nombre de visites organiques qu'elles génèrent), les deux informations sont superposées (l'une en barres verticales, l'autre en courbe).

Voilà la question à vous poser :

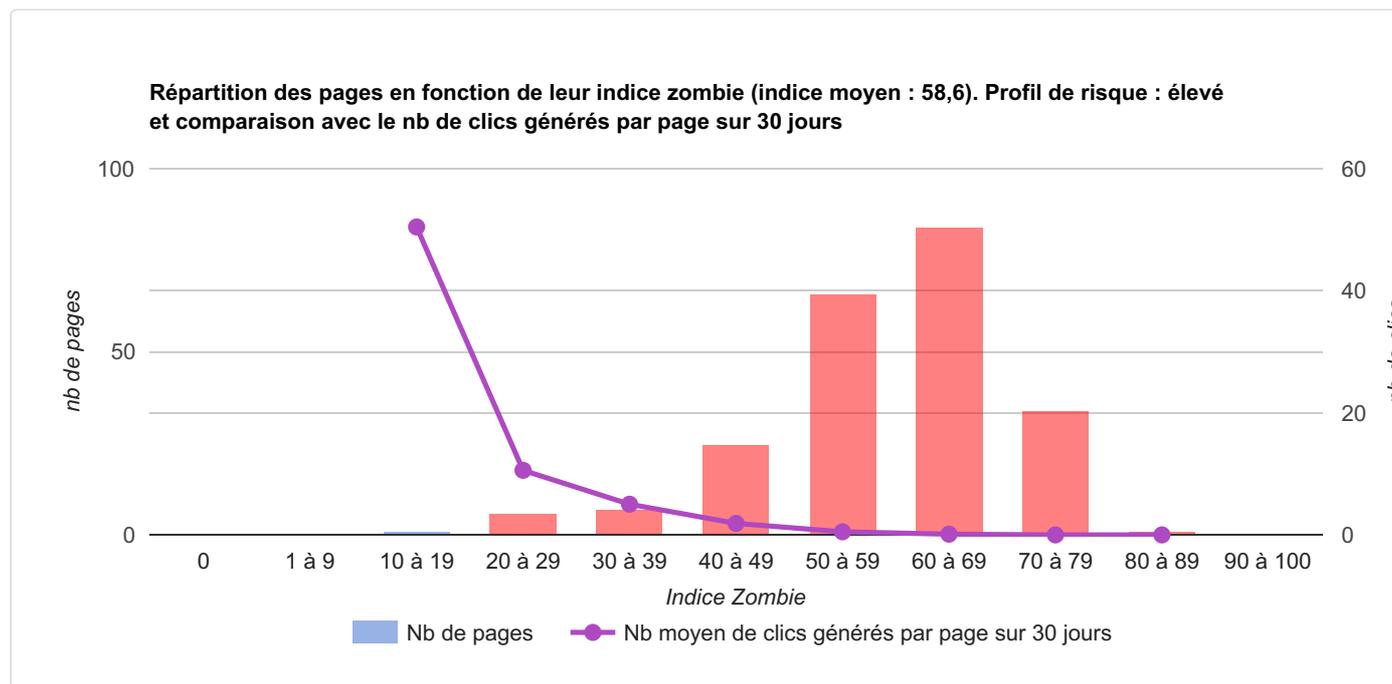
Est-ce là où le site contient le plus de pages (en termes d'indice QualityRisk) qu'elles génèrent le plus de trafic ?



L'histogramme ci-dessous montre la répartition des pages du site en fonction de leur indice Zombie. Comme il est important de voir en même temps combien le site contient de pages (selon leur indice Zombie), les deux informations sont superposées (l'une en barres verticales, l'autre en courbe).

Voilà la question à vous poser :

Est-ce là où le site contient le plus de pages (en termes d'indice Zombie) qu'elles génèrent le plus de trafic ?



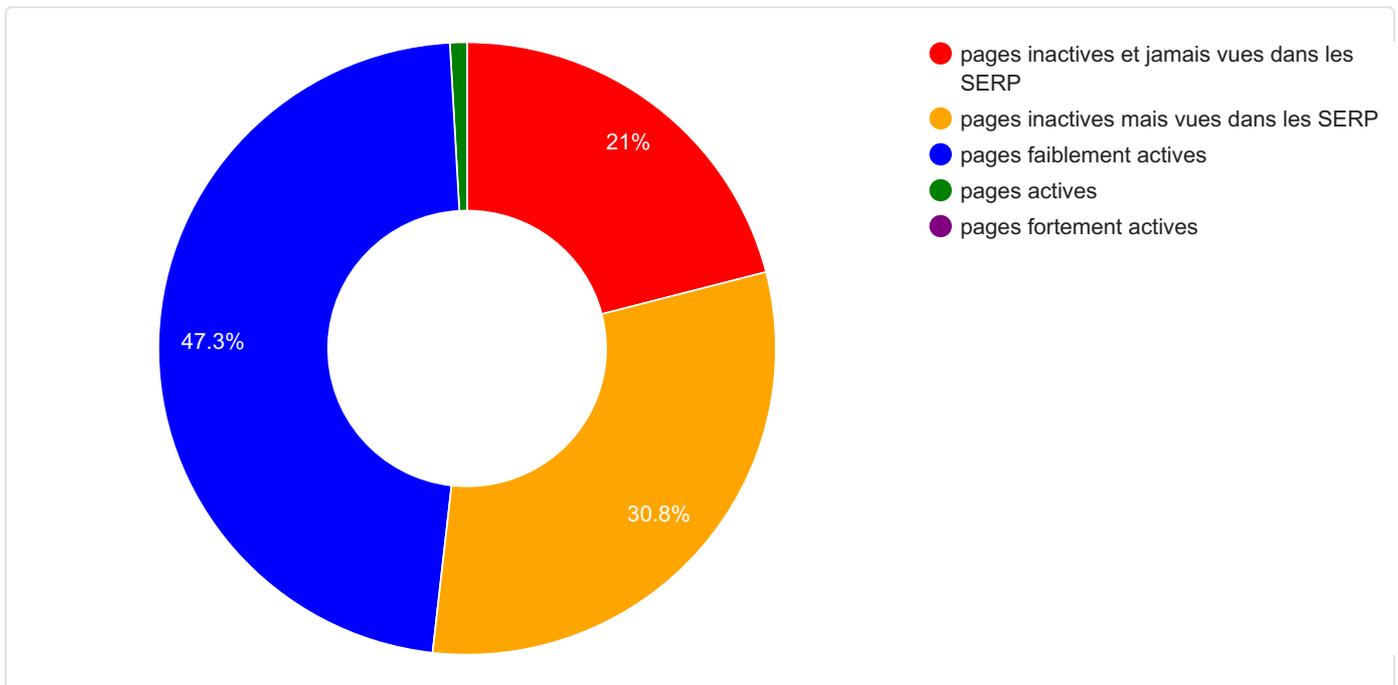
1.2 Présence dans les SERP Google et trafic généré

Sur la période étudiée, selon Google Search Console, l'URL de départ a généré **1 130 clics**. **2 923 clics** ont été générés par les 225 autres pages indexables découvertes par cet audit. Voici le bilan :

- Nb de pages inactives et jamais vues dans les SERP : 47
- Nb de pages inactives mais vues dans les SERP : 69
- Nb de pages faiblement actives (1 à 364 clics) : 106
- Nb de pages actives (365 à 3649 clics) : 2
- Nb de pages fortement actives (3650 clics et +) : 0

En résumé, il y a donc :

- 108 pages actives, soit 48,2 %
- 116 pages inactives, soit 51,8 %



1.3 Taux d'erreurs SEO

Ce tableau regroupe en 9 familles les dizaines d'éléments audités sur le site :

| Élément analysé | Optimisation | Erreurs constatées sur les pages | |
|-----------------------------|--------------|----------------------------------|--------------------|
| | | en pourcentage | en nombre de pages |
| URL répondant en erreur | ● | 2,47 % ⁽¹⁾ | 6 |
| Indexabilité des pages HTML | ● | 5,06 % ⁽²⁾ | 12 |
| Temps de téléchargement | ● | 0,42 % ⁽²⁾ | 1 |
| Profondeur | ● | 57,33 % ⁽³⁾ | 129 |
| Maillage interne | ● | 98,22 % ⁽³⁾ | 221 |
| Balise title | ● | 96 % ⁽³⁾ | 216 |
| Balise meta description | ● | 80,44 % ⁽³⁾ | 181 |
| Contenu des pages | ● | 68 % ⁽³⁾ | 153 |
| Liens sortants externes | ● | 13,79 % ⁽⁴⁾ | 4 |

- (1) pourcentage calculé par rapport au nombre d'URL explorées, soit ici 243 URL
- (2) pourcentage calculé par rapport au nombre d'URL de pages HTML sans erreur (code 200), soit ici 237 URL
- (3) pourcentage calculé par rapport au nombre d'URL de pages HTML indexables, soit ici 225 URL
- (4) pourcentage calculé par rapport au nombre d'URL externes distinctes trouvées dans les liens sortants follow, soit ici 29 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 qui est considéré comme des **erreurs**, mais vous verrez que par endroits dans ce rapport sont également listés 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).

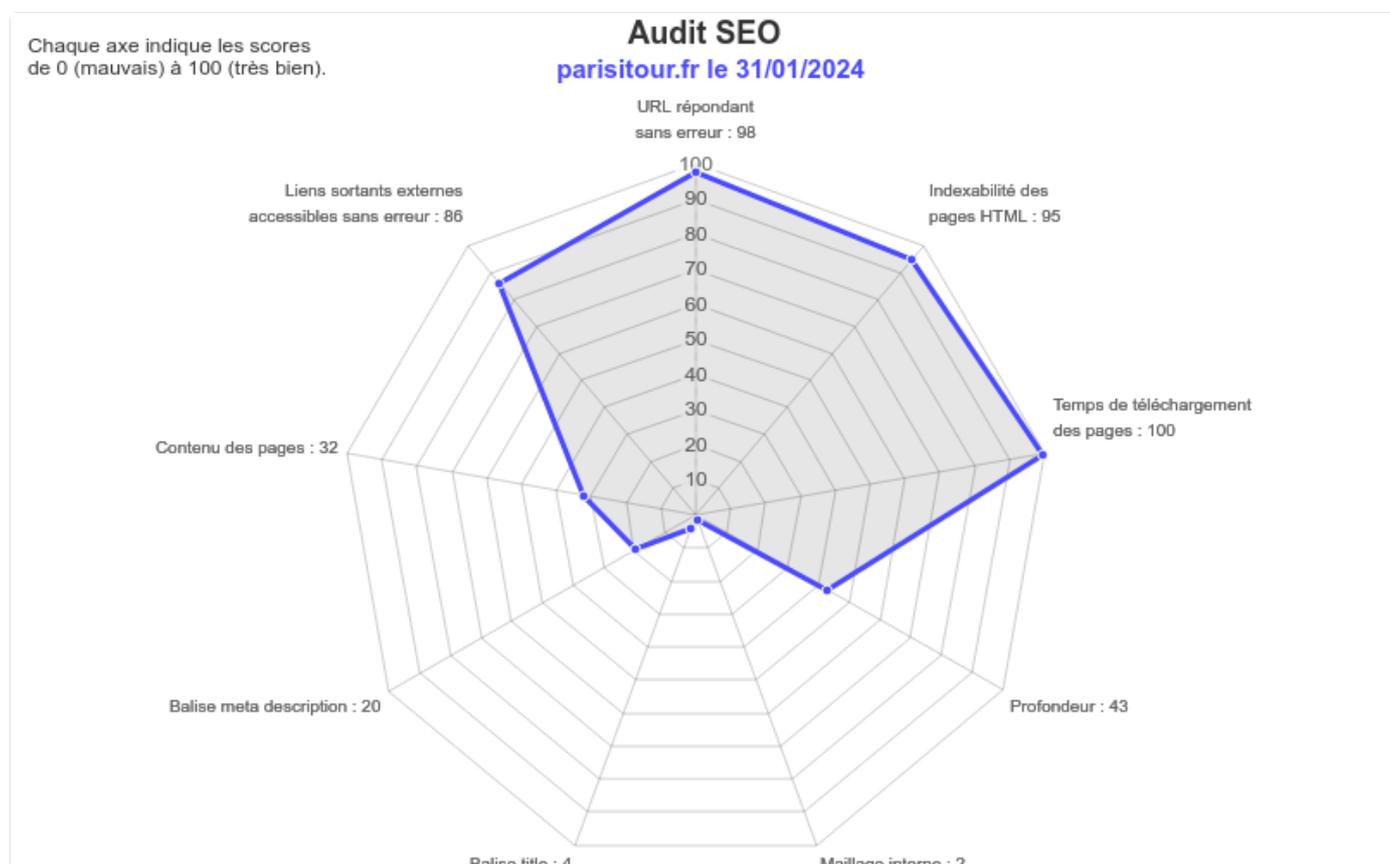
Certaines cases des tableaux contiennent le libellé "n/a" ce qui signifie "non applicable" pour indiquer qu'il est normal que la case ne contienne pas de valeur.

Les résultats présentés ci-dessus (et dans le reste du rapport) sont exprimés en termes de taux d'erreur. Retrouvez ci-dessous 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 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.



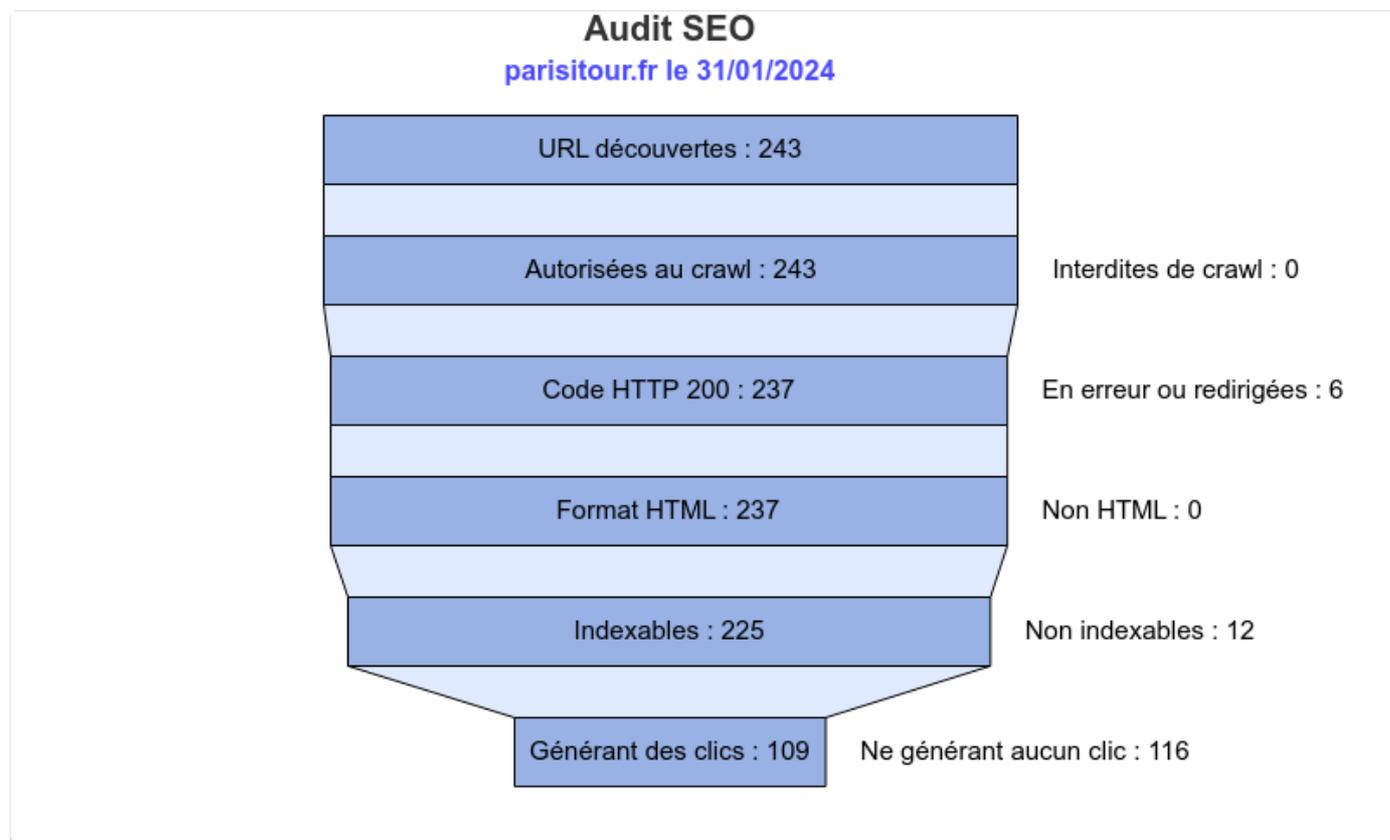
1.4 Notion d'entonnoir des URL

Un audit comme celui-ci est construit sur une notion d'entonnoir :

1. d'abord il faut lister toutes les URL trouvables par des liens (ou en suivant des redirections ou URL canoniques)
2. ensuite il faut se concentrer uniquement sur celles qui sont autorisées au crawl (non bloquées dans le robots.txt)
3. parmi elles, il faut étudier celles qui sont accessibles (code HTTP 200)
4. parmi elles, il faut se concentrer sur celles au format HTML
5. parmi elles, il faut étudier celles qui sont indexables
6. enfin, parmi elles, il faut analyser celles qui ont généré au moins 1 clic (selon GSC, y compris l'URL de départ).

À chaque niveau de l'entonnoir, la proportion d'URL OK est calculée à partir du nombre d'URL du niveau précédent.

Ce schéma résume cette notion d'entonnoir :



Voici les mêmes informations sous forme de tableau :

| Type d'analyse | Optimisation | Nb d'URL OK | % d'URL OK |
|-----------------------|--------------------|-------------|--------------------|
| URL découvertes | | 243 | |
| Autorisées au crawl ? | ● | 243 | 100 % |
| Code HTTP 200 ? | ● | 237 | 97,53 % |
| Format HTML ? | n/a ⁽¹⁾ | 237 | n/a ⁽¹⁾ |
| Indexables ? | ● | 225 | 94,94 % |
| Génération de clics | ● | 109 | 48,44 % |

(1) On ne peut pas dire si c'est bien ou pas d'avoir autre chose que des pages HTML (par exemple, il peut être légitime d'avoir des PDF). Le niveau d'optimisation n'a donc pas de sens.

1.5 Images

Le tableau ci-dessous présente le bilan de l'analyse des images elles-mêmes. Les pourcentages sont calculés par rapport aux 238 images distinctes (URL des images) autorisées au crawl.

| Élément analysé | Optimisation | Erreurs constatées sur les images | |
|-----------------------|-----------------------------------------------------------------------------------|-----------------------------------|--------------------|
| | | en pourcentage | en nombre d'images |
| Images inaccessibles |  | 0 % | 0 |
| Images non indexables |  | 0 % | 0 |
| Images trop lourdes |  | 89,08 % | 212 |

Le tableau ci-dessous présente le bilan de l'analyse des pages contenant des images (pages HTML code 200). Les pourcentages sont calculés par rapport aux 237 pages contenant au moins une image.

| Élément analysé | Optimisation | Erreurs constatées sur les pages | |
|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|----------------------------------|--------------------|
| | | en pourcentage | en nombre de pages |
| Nb de pages avec image(s) inaccessible(s) |  | 0 % | 0 |
| Nb de pages en HTTPS avec image(s) en HTTP |  | 0 % | 0 |
| Nb de pages avec image(s) sans attribut ALT |  | 0 % | 0 |
| Nb de pages avec + de 50 images |  | 0 % | 0 |
| Nb de pages avec image(s) trop lourde(s) |  | 100 % | 237 |
| Nb de pages avec + de 1 000 Ko d'images |  | 54,43 % | 129 |
| Nb de pages avec image(s) n'ayant pas les 2 dimensions indiquées dans le code HTML |  | 0 % | 0 |

2 Crawl et indexabilité

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

1. d'abord, le point est fait sur les (éventuelles) URL qui n'ont pas été crawlées en raison d'une interdiction figurant dans un fichier robots.txt
2. ensuite, l'ensemble des URL crawlées sont listées (qu'elles soient accessibles ou pas)
3. enfin, parmi les pages HTML accessibles, celles qui sont indexables sont étudiées (sans vérifier si Google les a indexées, mais aucun critère technique n'empêche qu'elles le soient)

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

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

Il est conseillé 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 le code HTTP du fichier robots.txt de chaque sous-domaine crawlé est indiqué (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 |
|---------------------------------------------------------------------------------|-----------|-----------------------|
| https://parisitour.fr/robots.txt | 200 | 0 |
| Total | | 0 |

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

Annexe

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

Annexe

Le fichier annexe [parisitour_fr-2024-01-31-robotstxt.zip](#) contient l'ensemble des fichiers robots.txt récupérés.

2.2 Nombre total d'URL crawlées

A partir d'ici et pour tout le reste de ce rapport, seules sont prises en compte les URL autorisées au crawl (non bloquées dans un fichier robots.txt).

Dans votre cas, 100% des URL du site (trouvables par des liens depuis la page de départ) ont pu être analysées. Le nombre d'URL crawlées a été de **243**.

Voici la répartition selon les protocoles HTTP et HTTPS :

| Protocole | Nb d'URL | % des URL crawlées |
|-----------|----------|--------------------|
| HTTP | 0 | 0 % |
| HTTPS | 243 | 100 % |

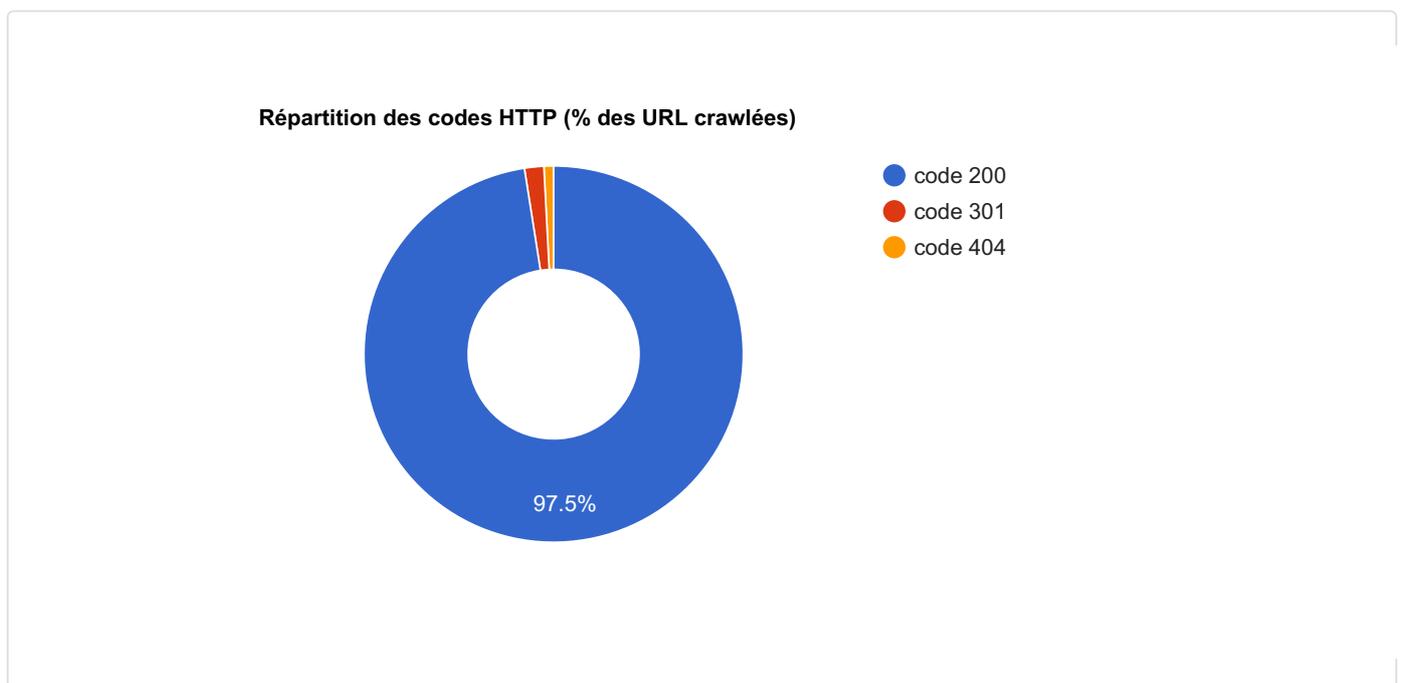
2.3 État des URL crawlées

Voici le bilan des codes HTTP rencontrés sur le 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 | 237 | 237 | 97,53 % |
| 301 | 0 | 4 | 4 | 1,65 % |
| 404 | 0 | 2 | 2 | 0,82 % |
| Total erreurs 3xx | 0 | 4 | 4 | 1,65 % |
| Total erreurs 4xx | 0 | 2 | 2 | 0,82 % |
| Total erreurs | 0 | 6 | 6 | 2,47 % |

Remarque : même si les redirections ne sont pas à proprement parler des erreurs techniques, c'est pénalisant pour le référencement. C'est pourquoi elles ont été incluses dans le calcul des erreurs. Il est recommandé 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 parisitour_fr-2024-01-31_03-3_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 parisitour_fr-2024-01-31_03-3_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 parisitour_fr-2024-01-31_03-3_liens-vers-erreurs-3xx.xlsx contient les URL des pages contenant les liens en erreur.

2.4 URL trouvables uniquement par des liens internes nofollow

Depuis le 01/03/2020, Google peut décider de suivre un lien nofollow, qu'il soit interne ou externe ([source](#)). Pour comprendre les implications SEO, lisez [cet article](#) d'Olivier Duffez sur WebRankInfo.

L'option de suivi des liens internes nofollow n'a pas été activée pour cet audit.

2.5 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 | 237 | 100 % |

Annexe

Le fichier annexe parisitour_fr-2024-01-31_03-5_types-mime.xlsx contient les données associées.

2.6 Pages HTML crawlées sans erreur

Le référencement naturel se travaille essentiellement avec les pages HTML. Pour vous aider à avoir une vue d'ensemble, les principales informations disponibles pour l'ensemble de vos pages HTML crawlées sans erreur (code HTTP 200) ont été regroupées dans les annexes :

- 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 parisitour_fr-2024-01-31_03-6_pages-sans-erreur.xlsx contient les données associées.

2.7 Pages HTML indexables par les moteurs

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

Il s'agit ici d'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 le site :

| | Nb d'URL | % des URL HTML code 200 |
|------------------------------------------------------------------|----------|-------------------------|
| Pages HTML avec code HTTP 200 (A) | 237 | |
| Dont pages interdites d'indexation | 2 | 0,84 % |
| Dont pages avec URL canonique différente de l'URL crawlée | 10 | 4,22 % |
| Nb total de pages non-indexables (B) | 12 | 5,06 % |
| Nombre de pages HTML indexables (A - B) | 225 | 94,94 % |

Sauf cas particuliers dont vous connaissez bien l'existence sur ce 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, 2 pages HTML interdites d'indexation ont été identifiées, représentant 0,84 % des pages HTML en code 200. Il faudrait vérifier si c'est normal d'interdire l'indexation de ces pages. Vous trouverez la liste des URL concernées dans les annexes.

Dans votre cas, 10 pages HTML avec une URL canonique différente de l'URL crawlée ont été identifiées, représentant 4,22 % des pages HTML en code 200. Il faudrait vérifier si c'est normal pour ces pages d'indiquer une URL canonique différente de l'URL crawlée. Vous trouverez la liste des URL concernées dans les annexes.

Annexe

Le fichier annexe parisitour_fr-2024-01-31_03-7_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 parisitour_fr-2024-01-31_03-7_liens-vers-noindex.xlsx contient les URL des pages contenant les liens en erreur.

Annexe

Afin de vous permettre de localiser les erreurs potentielles liées à l'URL canonique sur le site, le fichier annexe parisitour_fr-2024-01-31_03-7_liens-vers-erreurs-url-canonique.xlsx contient les URL des pages contenant les liens vers ces URL.

3 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

Il est recommandé d'optimiser la conception du site, le contenu de vos pages ainsi que le serveur d'hébergement afin d'avoir des pages qui se téléchargent rapidement. Dans cette partie, l'analyse concerne le temps de téléchargement de chaque page HTML répondant sans erreur (code HTTP 200). Il est recommandé que chaque page soit téléchargée en 600 ms maximum.

Pour l'ensemble des pages HTML code 200, la moyenne du temps de téléchargement des pages (code HTML brut) est de **363 ms**.

Voici les résultats de l'analyse :

| | Nb d'URL | % des URL HTML code 200 |
|---------------------------------------------------------|----------|-------------------------|
| Pages trop lentes (> 600 ms) | 1 | 0,42 % |
| dont : pages vraiment trop lentes (> 1000 ms) | 0 | 0 % |
| Nb de pages HTML code 200 | 237 | |
| Taux d'erreur | | 0,42 % |

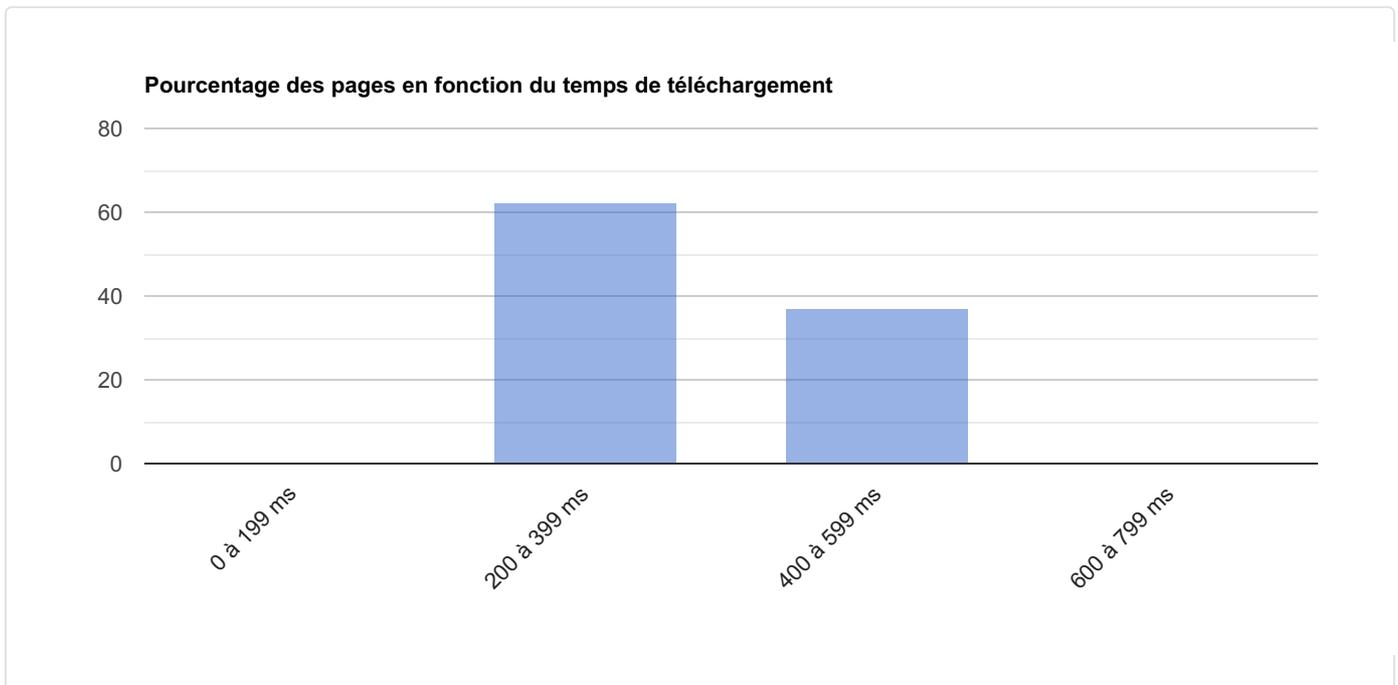
Remarque : **il ne s'agit pas ici du temps de chargement complet de la page**, puisque tous les fichiers externes sont ignorés pour ce calcul (Javascript, CSS, images, etc.). Il s'agit du temps mis par le serveur pour répondre à la requête, générer la page HTML et la télécharger. C'est donc le même principe que le temps indiqué dans Google Search Console : rubrique Paramètres > Exploration > Statistiques sur l'exploration > Temps de réponse moyen (ms). Pour être précis, n'oubliez pas de filtrer pour ne prendre en compte que les fichiers HTML (cliquez sur "HTML" dans "Par type de fichier"). En effet, par défaut Google tient compte aussi des 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 | 0 | 0 % |
| de 200 ms à 399 ms | 148 | 62,45 % |
| de 400 ms à 599 ms | 88 | 37,13 % |
| de 600 ms à 799 ms | 1 | 0,42 % |

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, après décompression). Attention, ceci n'inclut aucune ressource de la page : ni les images, ni les javascript, CSS et autres.

Il est déconseillé d'avoir un poids (code HTML brut) trop important car cela augmente le temps de téléchargement. En outre, même si Google crawle des documents très volumineux (plusieurs Mo), Bing s'arrête autour de 700 Ko (ce qui est déjà très gros). C'est un cas très grave puisque le contenu et les liens de la page risquent de ne pas être vus !

La limite pour cet audit est de 700 Ko. Si certaines pages dépassent ce seuil, une partie seulement de leur contenu est prise en compte.

Pour l'ensemble des pages HTML code 200, la moyenne du poids du code HTML brut est de **117 Ko**.

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 | |
|-------------------------------------------------------------------|---------|----------------|-----|
| Nb de pages dont le poids du code HTML brut dépasse 700 Ko | n/a | 0 | |
| Nb de pages HTML en code 200 | | | 237 |
| Taux d'erreur | | 0 % | |

Annexe

Le fichier annexe `parisitour_fr-2024-01-31_04_temps-telechargement.xlsx` contient les données associées.

4 Profondeur des pages et maillage interne

4.1 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, **la profondeur maximale recommandée est de 2**. Concrètement, cela signifie qu'il est préférable que la majorité des URL soit à une profondeur inférieure ou égale à 2. 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, il n'est pas conseillé de réduire artificiellement la profondeur des pages uniquement pour ne pas dépasser la profondeur recommandée. Il est préférable 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.

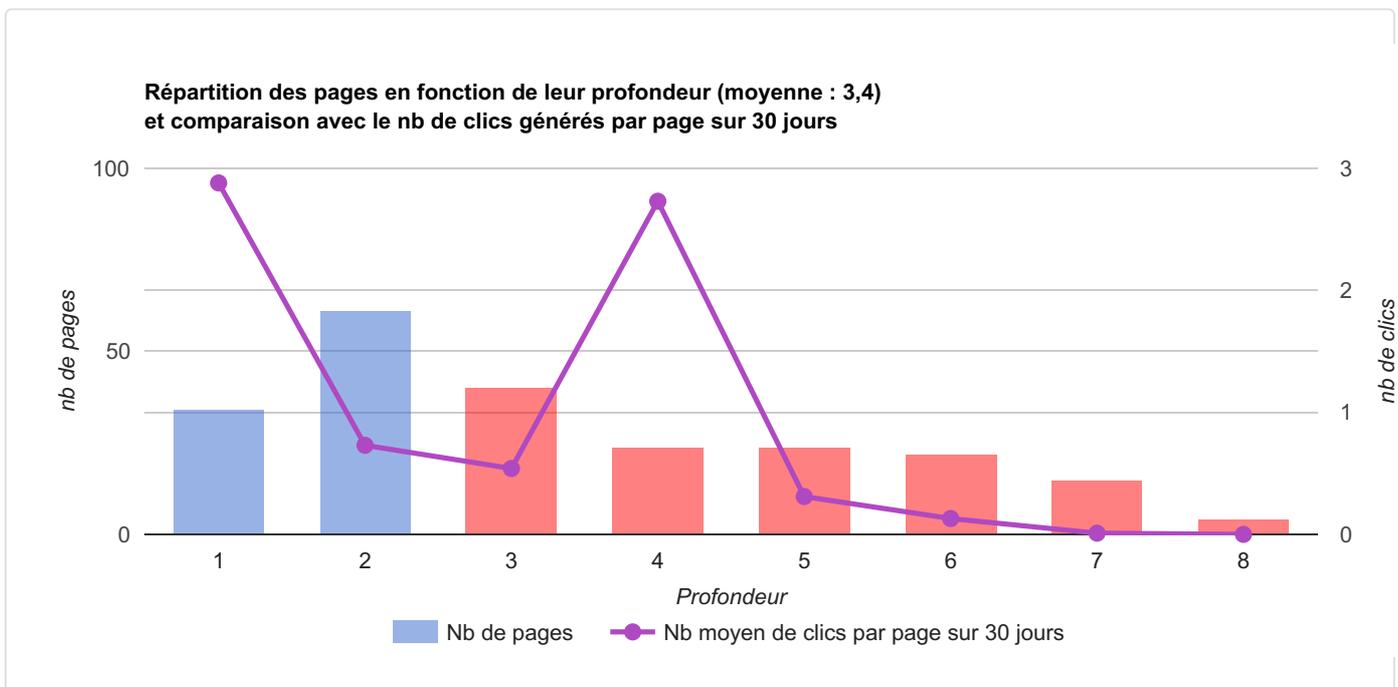
La profondeur moyenne des pages est de 3,4.

Voici les résultats selon les niveaux :

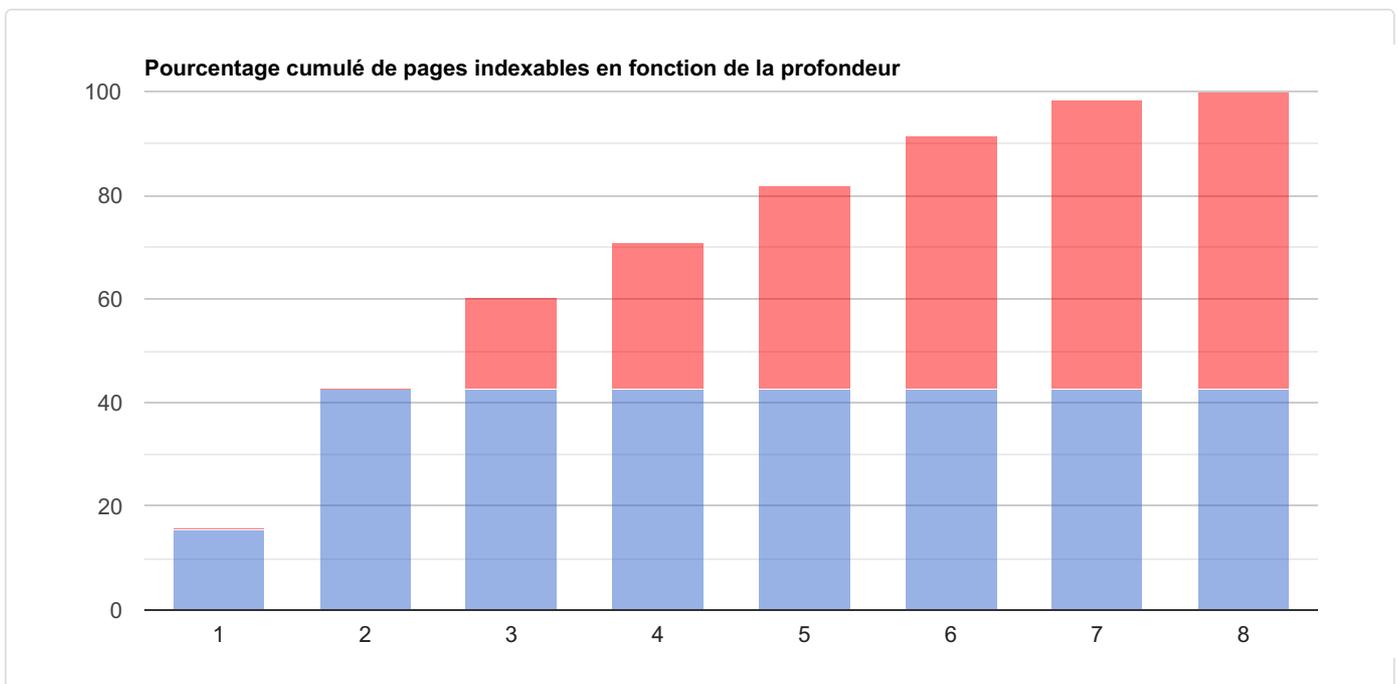
| Profondeur | Nb de pages | Nb cumulé de pages | Pourcentage des pages | Pourcentage cumulé |
|------------|-------------|--------------------|-----------------------|--------------------|
| 0 | 1 | 1 | 0,44 % | 0,44 % |
| 1 | 34 | 35 | 15,11 % | 15,56 % |
| 2 | 61 | 96 | 27,11 % | 42,67 % |
| 3 | 40 | 136 | 17,78 % | 60,44 % |
| 4 | 24 | 160 | 10,67 % | 71,11 % |
| 5 | 24 | 184 | 10,67 % | 81,78 % |
| 6 | 22 | 206 | 9,78 % | 91,56 % |
| 7 | 15 | 221 | 6,67 % | 98,22 % |
| 8 | 4 | 225 | 1,78 % | 100 % |

Dans votre cas, **129 pages trop profondes** ont été identifiées, soit un **taux d'erreur de 57,33 %**. 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 (URL de départ exclue), complétés par les statistiques sur le trafic généré :

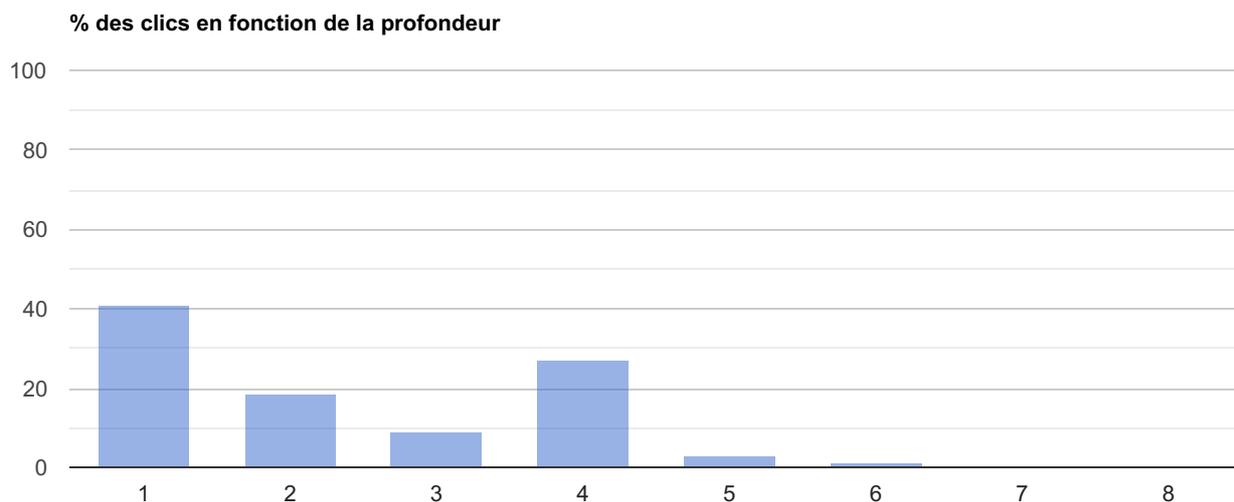


Le graphique ci-dessous montre le pourcentage cumulé (URL de départ exclue) :



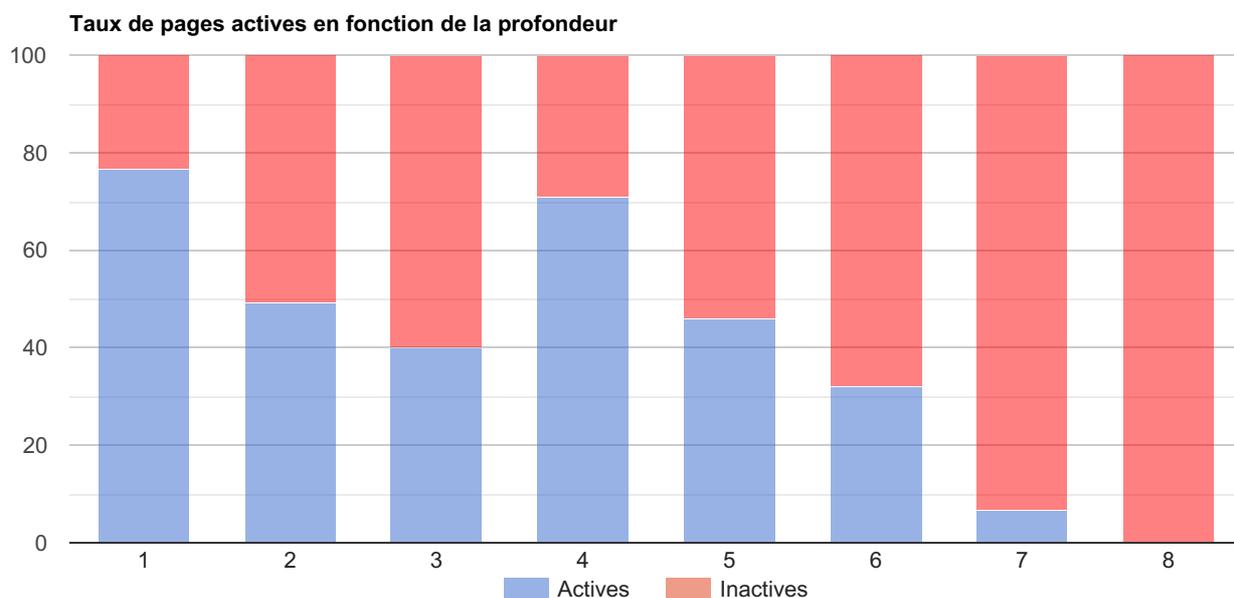
Pour compléter, voici une étude de la répartition du trafic généré en référencement naturel (= "organic") selon le niveau de profondeur.

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.



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.



4.2 Analyse du maillage interne

Pour qu'une page soit efficace en référencement naturel, il est impératif qu'elle reçoive des liens. On appelle ça des liens entrants internes (inlinks) quand ils viennent d'autres pages du site, et des backlinks (liens entrants externes) quand ils viennent d'autres sites.

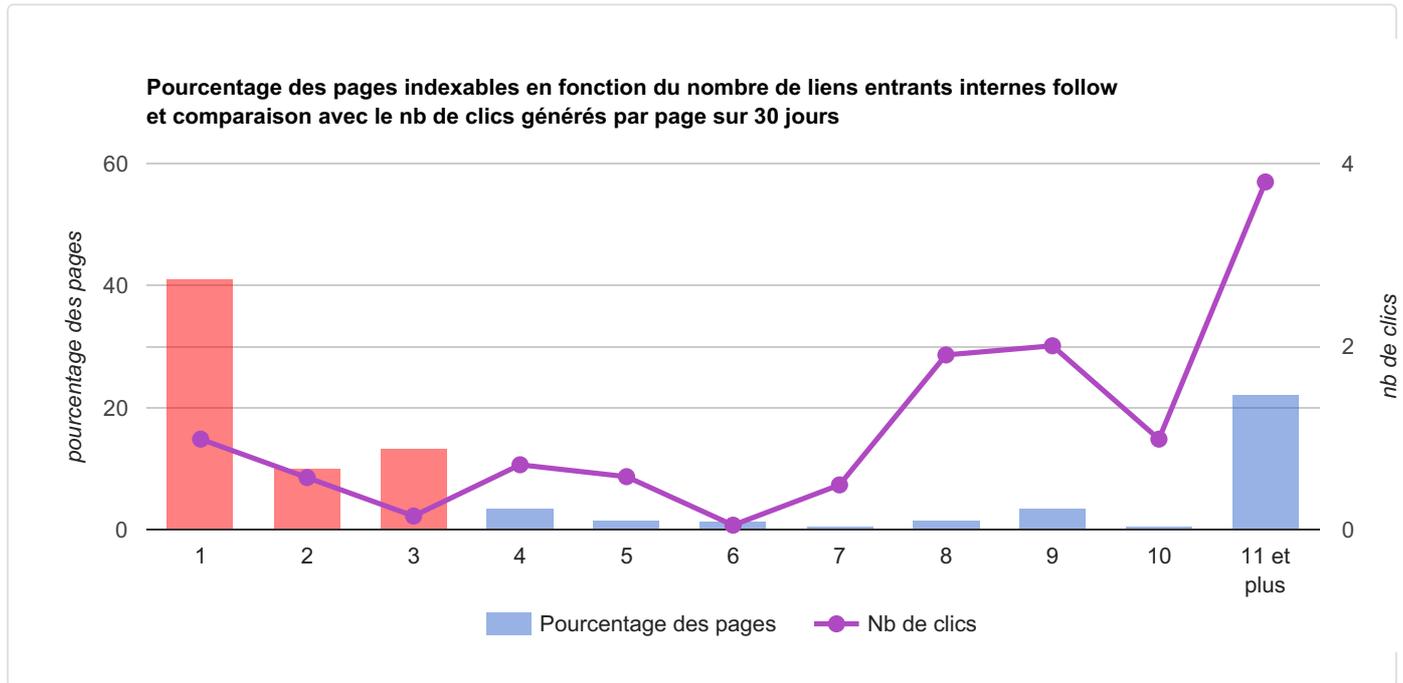
Évidemment, il faut au moins 1 lien entrant interne (sinon la page risque d'être orpheline et non indexée) mais c'est nettement insuffisant.

À part pour les petits sites, chaque page devrait recevoir au moins 10 liens entrants internes (avec des ancres variées, voir ci-dessous). RM Tech affiche un avertissement pour les pages ayant entre 4 et 9 liens entrants internes, et considère une erreur s'il y en a moins.

Dans ce graphique :

- la courbe en violet indique le trafic moyen généré par page en fonction du nombre de liens entrants
- les barres verticales montrent combien le site contient de pages selon leur nombre de liens entrants

Sur ce site, les pages avec au moins 10 liens entrants internes génèrent **4 fois +** de visites SEO que celles avec un seul lien.

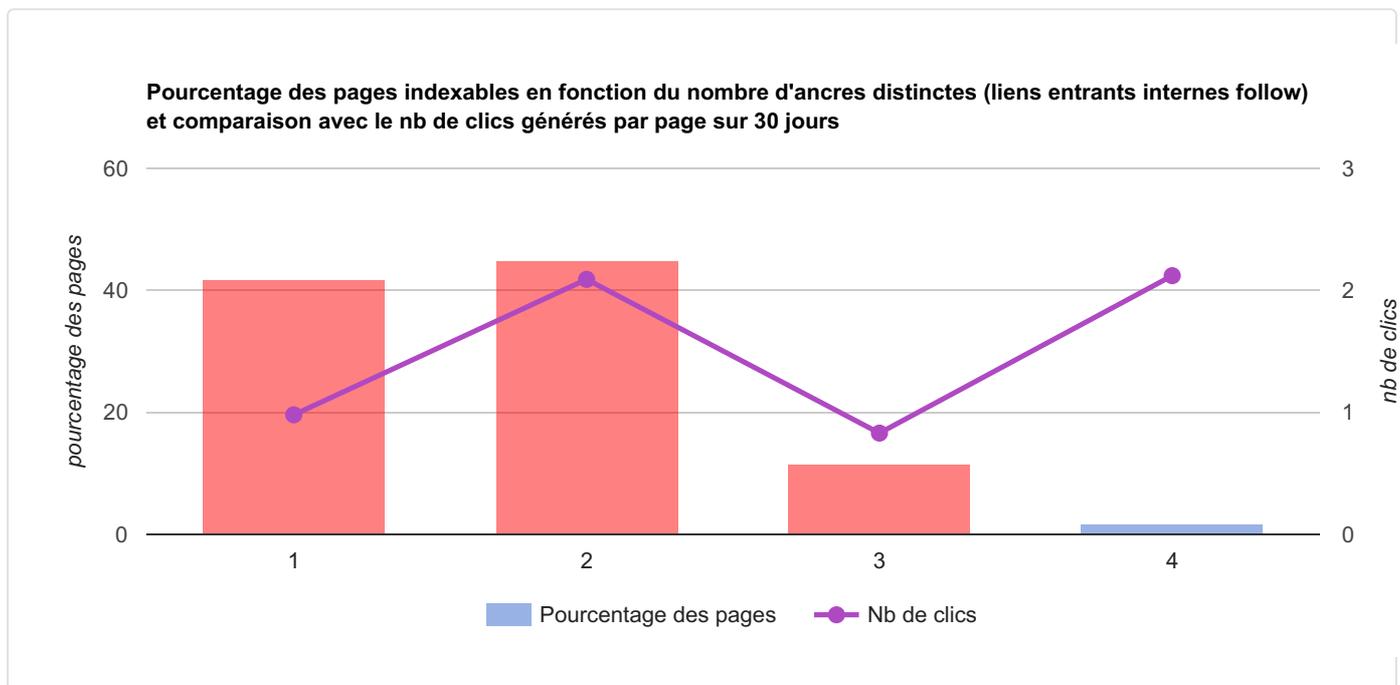


Toujours concernant les liens entrants internes, il est intéressant d'étudier aussi les ancres de liens (c'est-à-dire le texte cliquable qui mène à la page, ou bien l'attribut alt dans le cas d'un lien sur une image). Là aussi, il est préférable **d'éviter d'avoir toujours exactement le même texte dans 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.

À part pour les petits sites, chaque page devrait avoir au moins 10 ancres distinctes dans ses liens entrants internes. RM Tech affiche un avertissement pour les pages ayant entre 4 et 9 ancres, et considère une erreur s'il y en a moins.

Dans ce graphique :

- la courbe en violet indique le trafic moyen généré par page en fonction du nombre d'ancres dans leurs liens entrants
- les barres verticales montrent combien le site contient de pages selon leur nombre d'ancres de liens entrants



Voici le tableau récapitulatif :

| | Erreurs | Avertissements | |
|---------------------------------------------------------------------------------|----------------|----------------|-----|
| Nb de pages avec seulement 0 à 3 liens entrants internes | 146 | n/a | |
| Nb de pages avec seulement 4 à 9 liens entrants internes | n/a | 28 | |
| Nb de pages avec seulement 0 à 3 ancres dans ses liens entrants internes | 221 | n/a | |
| Nb de pages avec seulement 4 à 9 ancres dans ses liens entrants internes | n/a | 4 | |
| Nb de pages avec un ou des liens entrants internes à ancre vide | n/a | 2 | |
| Nb total d'erreurs et d'avertissements | 221 | 30 | |
| Nb de pages indexables | | | 225 |
| Taux d'erreur | 98,22 % | 13,33 % | |

Remarques importantes :

1. Au sujet des ancres de liens, **évités les ancres vides** : il s'agit souvent de liens placés sur des images sans attribut ALT, mais parfois aussi de liens texte vide. Il faut absolument l'éviter, car cela ne permet pas d'optimiser le référencement de la page recevant le lien.
2. 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. Mais si ce 1er lien est nofollow, alors aucun lien n'est pris en compte.
3. Comme pour la profondeur, il n'est pas conseillé d'augmenter artificiellement le nombre d'ancres dans les liens uniquement pour réduire le taux d'erreur. Il est préférable 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.
4. Si le site comporte un menu listant une grande partie des pages, il sera sans doute difficile de varier leurs ancres (car seul le 1er lien est pris en compte), mais il n'est pas nécessaire de changer le menu si celui-ci

est utile tel quel pour les visiteurs.

5. En résumé, il n'est pas toujours possible ou facile d'éviter que certaines pages n'aient qu'une seule ancre dans leurs liens entrants internes, par contre il faut essayer de le faire pour les pages stratégiques.

4.3 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).

Annexe

Le fichier annexe parisitour_fr-2024-01-31_05-3_profondeur-maillage.xlsx contient les statistiques sur la profondeur, le nombre de liens entrants internes et d'ancres distinctes.

Annexe

Le fichier annexe parisitour_fr-2024-01-31_05-3_ancres-internes.xlsx contient pour chaque URL la liste des ancres avec pour chaque ancre le nombre de liens entrants internes ainsi qu'un exemple d'URL de lien entrant interne.

Annexe

Le fichier annexe parisitour_fr-2024-01-31_05-3_liens-ancres-vides.xlsx contient pour chaque URL la liste des pages recevant au moins un lien interne avec une ancre vide.

5 Balise title

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

5.1 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, c'est considéré comme une erreur.

Voici les résultats de l'analyse :

| | Erreurs | Avertissements | |
|--------------------------------------------------------------------|---------------|----------------|-----|
| Nb de pages sans balise title ou avec une balise title vide | 0 | n/a | |
| Nb de pages avec seulement 1 mot dans title | 0 | n/a | |
| Nb de pages avec seulement 2 mots dans title | 0 | n/a | |
| Nb de pages avec seulement 3 mots dans title | 0 | n/a | |
| Nb de pages avec entre 71 et 80 caractères dans title | n/a | 4 | |
| Nb de pages avec entre 81 et 100 caractères dans title | 2 | n/a | |
| Nb de pages avec + de 100 caractères dans title | 2 | n/a | |
| Nb total d'erreurs et d'avertissements | 4 | 4 | |
| Nb de pages indexables | | | 225 |
| Taux d'erreur | 1,78 % | 1,78 % | |

Annexe

Le fichier annexe parisitour_fr-2024-01-31_06-1_taille-balise-title.xlsx contient les données associées.

5.2 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 c'est considéré comme une erreur.

Voici les résultats de l'analyse :

| | Erreurs | Avertissements | |
|-------------------------------------------------------------------------------------|-------------|----------------|-----|
| Nb de balises title utilisées plus d'une fois sur le site (hors balise vide) | 94 | n/a | |
| Nb de pages indexables par Google concernées (balise vide incluse) | 216 | n/a | |
| Nb de pages indexables | | | 225 |
| Taux d'erreur | 96 % | n/a | |

Annexe

Le fichier annexe parisitour_fr-2024-01-31_06-2_doublons-title.xlsx contient les données associées.

5.3 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 ce cas est considéré comme un avertissement). Néanmoins, il peut y avoir des cas particuliers, notamment avec les articles et autres mots très courants. Voici les résultats de l'analyse :

| | Erreurs | Avertissements | |
|---------------------------------------------------------------|------------|----------------|-----|
| Nb de pages avec 2 fois un même mot dans title | n/a | 18 | |
| Nb de pages avec 3 fois un même mot dans title | 0 | n/a | |
| Nb de pages avec 4 fois un même mot dans title | 0 | n/a | |
| Nb de pages avec 5 fois ou plus un même mot dans title | 0 | n/a | |
| Nb total d'erreurs et d'avertissements | 0 | 18 | |
| Nb de pages indexables | | | 225 |
| Taux d'erreur | 0 % | 8 % | |

Annexe

Le fichier annexe parisitour_fr-2024-01-31_06-3_repetition-mots-dans-title.xlsx contient les données associées.

5.4 Conclusion sur la balise Title

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

Pour cette conclusion, seules les erreurs sont prises en compte (pas les avertissements).

Voici les résultats de l'analyse :

| | Erreurs | |
|-----------------------------------------------------------------------|----------------|-----|
| Nb de pages avec une ou plusieurs erreurs dans la balise Title | 216 | |
| Nb de pages indexables | | 225 |
| Taux d'erreur | 96 % | |

6 Balise meta description

6.1 Taille de la balise meta description

Même si la meta description n'a pas d'impact officiel direct sur le positionnement, il est conseillé de rédiger un texte d'au moins 90 caractères. Google n'indique aucun nombre maximal de caractères, mais vous pouvez cibler environ 160 caractères, voire 200. Il n'est pas garanti que Google l'utilise pour l'extrait dans les SERP car il le génère parfois à partir du texte présent dans le contenu de la page.

Voici les résultats de l'analyse :

| | Erreurs | |
|------------------------------------------------------------------------|---------------|-----|
| Nb de pages avec balise meta description vide ou absente | 18 | |
| Nb de pages avec une meta description de moins de 90 caractères | 3 | |
| Nb total | 21 | |
| Nb de pages indexables | | 225 |
| Taux d'erreur | 9,33 % | |

Annexe

Le fichier annexe parisitour_fr-2024-01-31_07-1_taille-meta-description.xlsx contient les données associées.

6.2 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 l'analyse :

| | Erreurs | Avertissements | |
|------------------------------------------------------------------------------------|----------------|----------------|-----|
| Nb de balises meta description utilisées plus d'une fois (hors balise vide) | 72 | n/a | |
| Nb de pages concernées | 160 | n/a | |
| Nb de pages indexables | | | 225 |
| Taux d'erreur | 71,11 % | n/a | |

Annexe

Le fichier annexe parisitour_fr-2024-01-31_07-2_doublons-meta-description.xlsx contient les données associées.

6.3 Conclusion sur la balise meta description

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

Pour cette conclusion, seules les erreurs sont prises en compte (pas les avertissements).

Voici les résultats de l'analyse :

| | Erreurs | |
|----------------------------------------------------------------------------------|----------------|-----|
| Nb de pages avec une ou plusieurs erreurs dans la balise meta description | 181 | |
| Nb de pages indexables | | 225 |
| Taux d'erreur | 80,44 % | |

7 Contenu des pages

7.1 Taille du texte (zone principale de la page)

Comme vous le savez sans doute, l'algorithme de Google cherche à identifier les pages à faible valeur ajoutée. Vous le verrez 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.

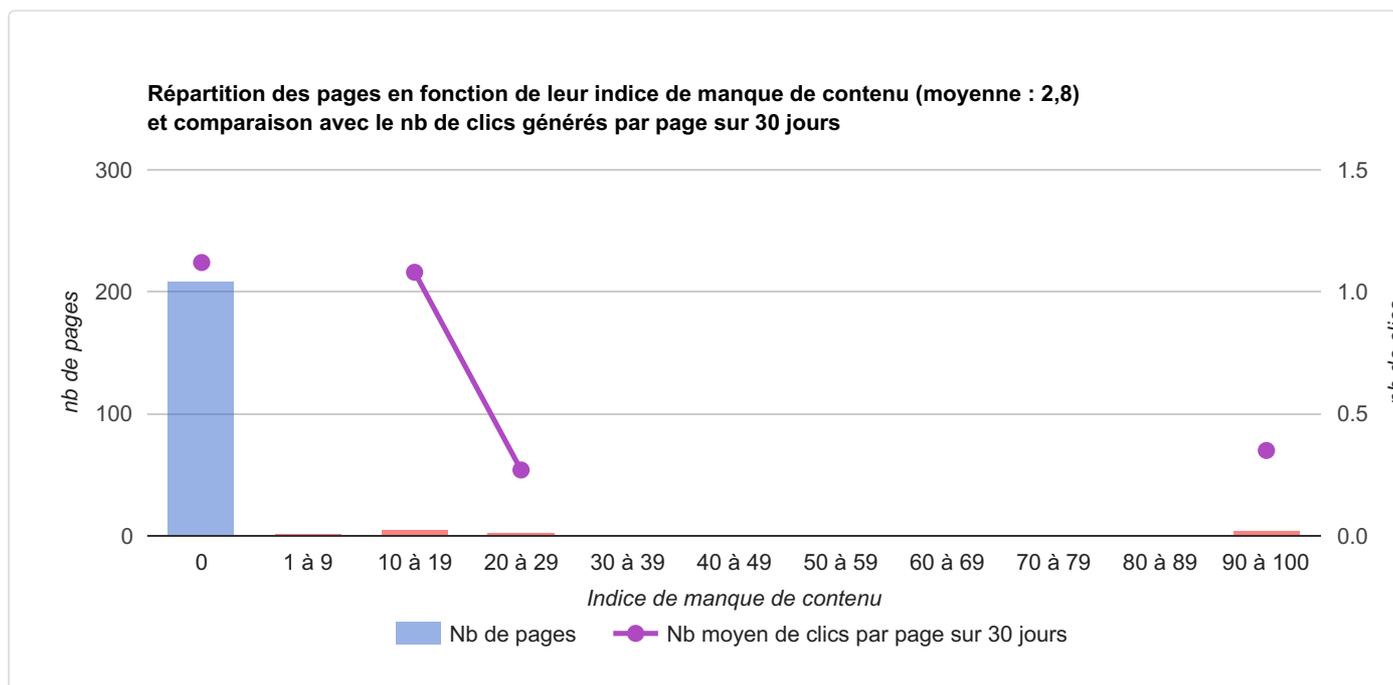
Une approche simpliste consiste à compter le nombre de mots dans l'ensemble de la page HTML (incluant l'entête, les menus, les barres latérales et le pied de page). Comme vous vous en doutez, cela ne fonctionne pas car ce qui compte, c'est la taille du contenu dans la zone principale de la page.

Grâce à un algorithme spécifique, ce rapport vous aide à identifier les pages sans doute trop courtes.

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

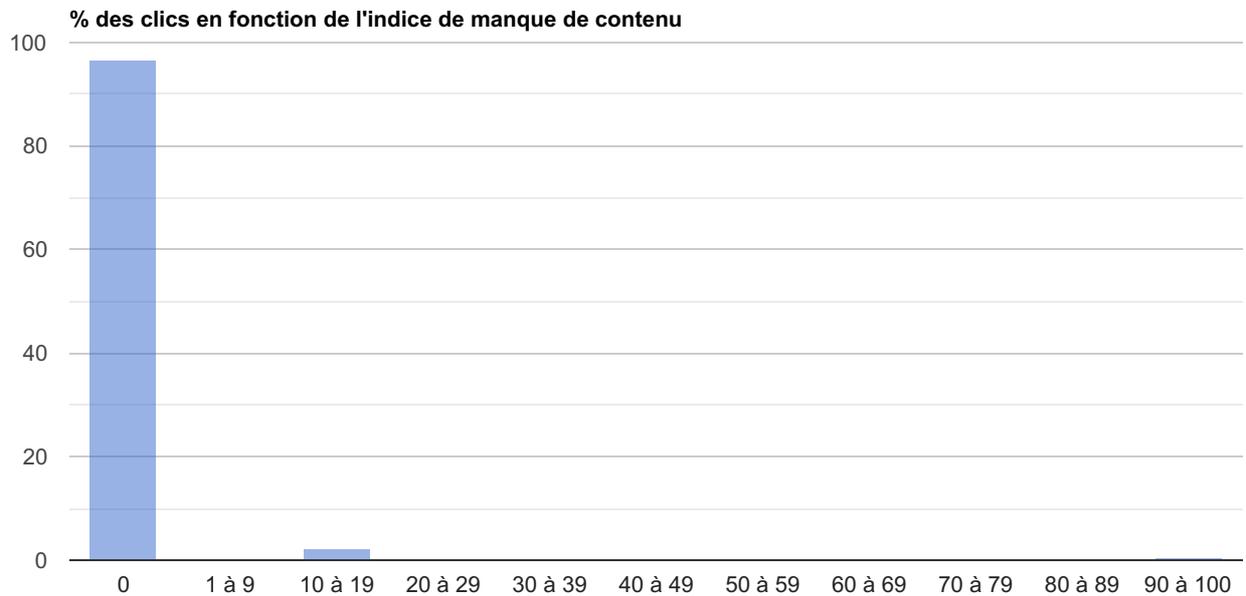
Pour l'ensemble des pages HTML indexables, la moyenne de l'indice de manque de contenu est de **2,8**.

L'histogramme ci-dessous montre la répartition des pages du site en fonction de leur indice de manque de contenu, ainsi que les statistiques sur le trafic généré :

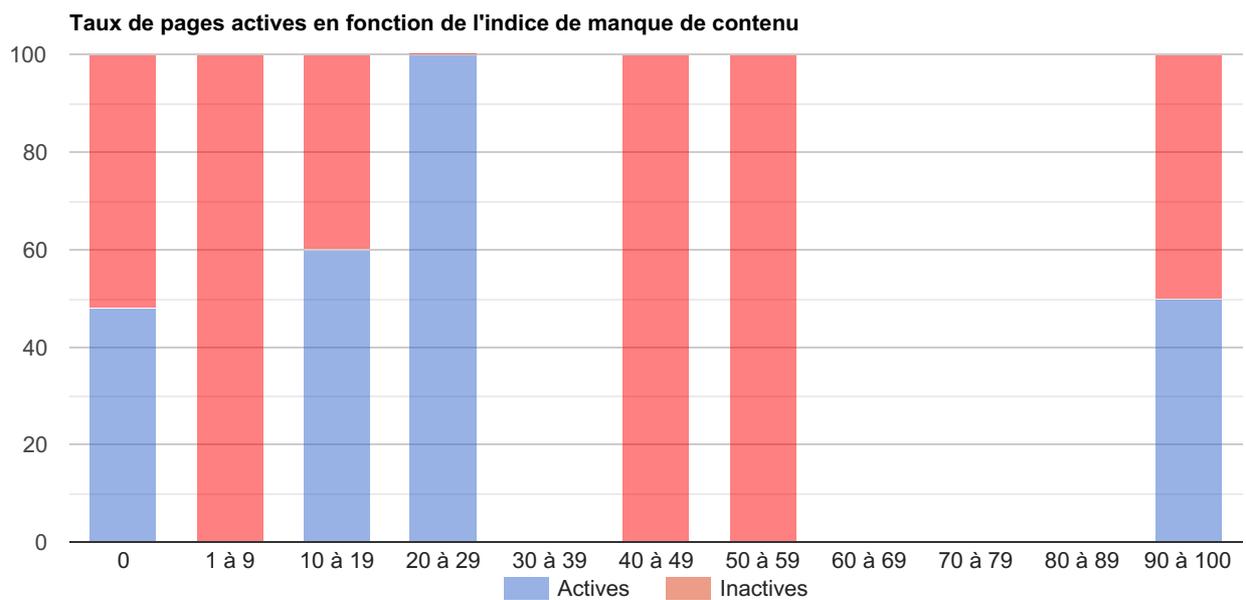


Pour compléter, voici une étude de la répartition du trafic généré en référencement naturel (= "organic") selon l'indice de manque de contenu.

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.



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



Le tableau ci-dessous donne la synthèse :

| | Erreurs | Avertissements | |
|--------------------------------------------|---------|----------------|-----|
| Nb de pages sans doute trop courtes | 16 | n/a | |
| Nb de pages indexables | | | 225 |
| Taux d'erreur | 7,11 % | n/a | |

Seules les pages manquant de contenu sont présentes dans l'annexe. Le fichier annexe parisitour_fr-2024-01-31_08-3_taille-texte.xlsx contient les données associées.

7.2 Balises H1-H6

Pour vous donner le plus de chances de réussir votre référencement, il est recommandé 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.) il est généralement conseillé uniquement une balise H1.

Voici les règles qu'il est recommandé de suivre (considérées sinon comme des erreurs) :

- au minimum, une page doit avoir au moins 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 qu'il est également recommandé 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, déterminé par 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 l'analyse :

| | Erreurs | Avertissements | |
|-----------------------------------------------------------|----------------|-----------------------|-----|
| Nb de pages sans balise H1 | 27 | 0 | |
| Nb de pages avec une balise H1 mais vide | 34 | n/a | |
| Nb de pages où la 1ère balise Hn n'est pas une H1 | n/a | 23 | |
| Nb de pages avec saut de niveau entre les Hn | n/a | 140 | |
| Nb de pages avec une balise Hn trop longue | n/a | 73 | |
| Nb de pages avec plusieurs balises H1 | n/a | 74 | |
| Nb de pages avec balise H2-H6 isolée | n/a | 36 | |
| Nb de pages avec le même libellé dans plusieurs Hn | n/a | 62 | |
| Nb de pages avec le même libellé dans H1 et Title | n/a | 37 | |
| Nb de pages indexables | | | 225 |
| Taux d'erreur | 27,11 % | 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 parisitour_fr-2024-01-31_08-1_balises-h1-h6.xlsx contient les données associées.

7.3 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 cela est considéré comme une erreur.

Voici les résultats de l'analyse :

| | Erreurs | Avertissements | |
|----------------------------------------------------------------------------------|----------------|-----------------------|-----|
| Nb de balises H1 utilisées plus d'une fois sur le site (hors balise vide) | 94 | n/a | |
| Nb de pages indexables par Google concernées (balise vide incluse) | 150 | n/a | |
| Nb de pages indexables | | | 225 |
| Taux d'erreur | 66,67 % | n/a | |

Annexe

Le fichier annexe parisitour_fr-2024-01-31_08-2_doublons-h1.xlsx contient les données associées.

7.4 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.

4 types de problèmes ont été identifiés :

- 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). Les réseaux sociaux suivants ont été exclus des analyses : LinkedIn, Twitter, Facebook, Pinterest, Viadeo, Myspace, Youtube, Dailymotion, Instagram, Vimeo, Flickr.

Les analyses ont porté sur l'ensemble de la page, en incluant le haut de page, les éventuelles marges et le pied de page.

Voici les résultats de l'analyse :

| | Erreurs | Avertissements | |
|----------------------------------------------------------------|----------------|-----------------------|-----|
| Nb de pages avec en tout moins de 5 liens | 0 | n/a | |
| Nb de pages avec en tout plus de 300 liens | n/a | 0 | |
| Nb de pages avec plus de 10 liens externes | n/a | 0 | |
| Nb de pages avec 1 ou plusieurs liens internes nofollow | n/a | 0 | |
| Nb total | 0 | 0 | |
| Nb de pages indexables | | | 225 |
| Taux d'erreur | 0 % | 0 % | |

Annexe

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

7.5 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 êtes 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).

L'ensemble des pages HTML indexables ont été analysées 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. Les cas de contenus presque dupliqués n'ont pas été étudiés (il s'agit des cas 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 l'analyse :

| | Erreurs | Avertissements | |
|----------------------------------------------------------------------|----------------|-----------------------|-----|
| Nb d'URL de pages HTML ayant un contenu strictement identique | 0 | n/a | |
| Nb de pages indexables | | | 225 |
| Taux d'erreur | 0 % | 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.

7.6 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 l'analyse :

| | Erreurs | Avertissements | |
|------------------------------------------|----------------|-----------------------|-----|
| Nb de pages avec un ID de session | 0 | n/a | |
| Nb de pages indexables | | | 225 |
| Taux d'erreur | 0 % | 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.

7.7 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 l'analyse :

| | Erreurs | Avertissements | |
|--------------------------------------------------|----------------|-----------------------|-----|
| Nb de pages avec une variable de tracking | 0 | n/a | |
| Nb de pages indexables | | | 225 |
| Taux d'erreur | 0 % | 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.

7.8 Conclusion sur le contenu des pages

Pour optimiser votre référencement, aucune page ne doit comporter une ou plusieurs des erreurs décrites précédemment. Ce rapport a listé les cas où une page HTML contient au moins une de ces erreurs.

Pour cette conclusion, seules les erreurs sont prises en compte (pas les avertissements).

Voici les résultats de l'analyse :

| | Erreurs | |
|------------------------------------------------------------------|----------------|-----|
| Nb de pages avec une ou plusieurs erreurs dans le contenu | 153 | |
| Nb de pages indexables | | 225 |
| Taux d'erreur | 68 % | |

8 Liens sortants externes

Dans cette rubrique, l'analyse porte sur tous les liens qui sortent du site (ils pointent donc vers un autre nom de domaine). Ils sont appelés « liens externes », qu'il ne faut pas confondre avec des backlinks (qui peuvent être eux aussi internes ou externes).

Les liens nofollow ont été ignorés, ainsi que les liens vers les réseaux sociaux suivants : LinkedIn, Twitter, Facebook, Pinterest, Viadeo, Myspace, Youtube, Dailymotion, Instagram, Vimeo, Flickr.

8.1 Nombre total de liens sortants externes

311 liens externes ont été identifiés, pointant vers 29 URL distinctes.

Annexe

Le fichier annexe `parisitour_fr-2024-01-31_09-1_liens-sortants-externes.xlsx` contient les données associées.

8.2 Erreurs sur les liens sortants externes

L'analyse concerne ici uniquement les cas où les liens sortants externes 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 ce cas de figure est considéré simplement comme un avertissement.

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

| Code HTTP | Nb d'URL | % des URL externes |
|-----------|----------|--------------------|
| 301 | 1 | 3,45 % |
| 500 | 4 | 13,79 % |

| | Erreurs | Avertissements | |
|----------------------------------------------------------|---------|----------------|----|
| Nombre total de liens sortants externes redirigés | n/a | 1 | |
| Nombre total de liens sortants externes en erreur | 4 | n/a | |
| Nombre total de liens sortants externes | | | 29 |
| Taux d'erreur | 13,79 % | 3,45 % | |

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.

9 Images

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

238 images (autorisées au crawl) ont été trouvées et analysées sur **237 pages** HTML.

Remarque : 1 image de 1x1 pixel a été trouvée mais ignorée dans cette analyse.

Annexe

Le fichier annexe parisitour_fr-2024-01-31_10_images.xlsx contient la liste des images.

Annexe

Le fichier annexe parisitour_fr-2024-01-31_10_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 parisitour_fr-2024-01-31_10_pages-et-images.xlsx contient la liste des couples Page / Image.

9.1 Images bloquées par le(s) fichier(s) robots.txt

En analysant les 237 pages HTML accessibles, **aucune image** n'est bloquée dans les fichiers robots.txt.

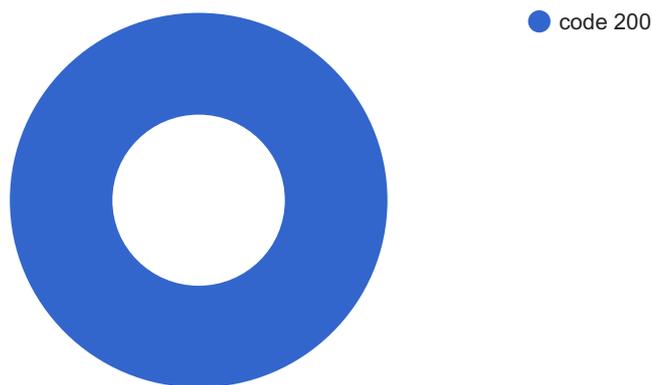
9.2 Codes HTTP des URL des images

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 | 0 | 238 | 238 | 100 % |
| Total erreurs | 0 | 0 | 0 | 0 % |

Voici ce que cela donne sous forme graphique :

Répartition des codes HTTP (% des images crawlées)



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

À partir d'ici, toutes les analyses qui suivent portent sur les images accessibles (code HTTP 200, le code final en cas de redirection).

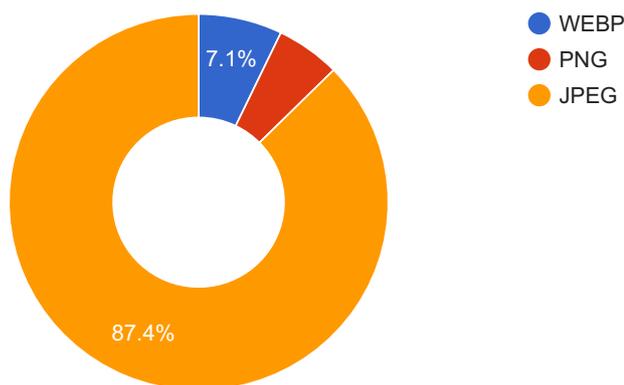
9.3 Types MIME (formats d'images)

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 | 208 | 87,39 % |
| image/webp | 17 | 7,14 % |
| image/png | 13 | 5,46 % |

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.

9.4 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 |
|-------------------------------------------------------------------|-------------|-----------------------|
| Images avec code HTTP 200 (A) | 238 | |
| Dont images interdites d'indexation | 0 | 0 % |
| Dont images avec URL canonique différente de l'URL crawlée | 0 | 0 % |
| Nb total d'images non-indexables (B) | 0 | 0 % |
| Nombre d'images indexables (A - B) | 238 | 100 % |

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

9.5 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).

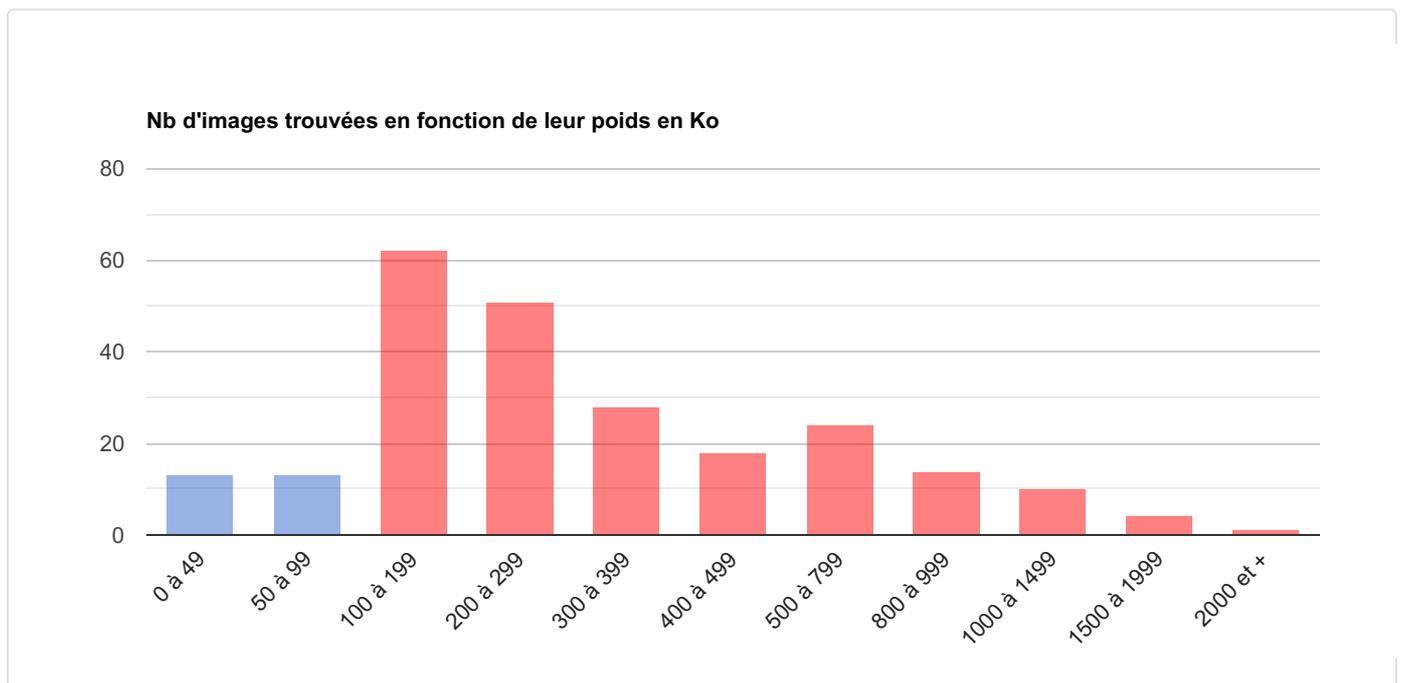
La moyenne du poids des images est de **378 Ko**.

212 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 | 13 | 5,46 % |
| 50 à 99 | 13 | 5,46 % |
| 100 à 199 | 62 | 26,05 % |
| 200 à 299 | 51 | 21,43 % |
| 300 à 399 | 28 | 11,76 % |
| 400 à 499 | 18 | 7,56 % |
| 500 à 799 | 24 | 10,08 % |
| 800 à 899 | 14 | 5,88 % |
| 1000 à 1499 | 10 | 4,20 % |
| 1500 à 1999 | 4 | 1,68 % |
| 2000 et + | 1 | 0,42 % |

Voici ce que cela donne sous forme graphique :



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

9.6 Temps de téléchargement des images

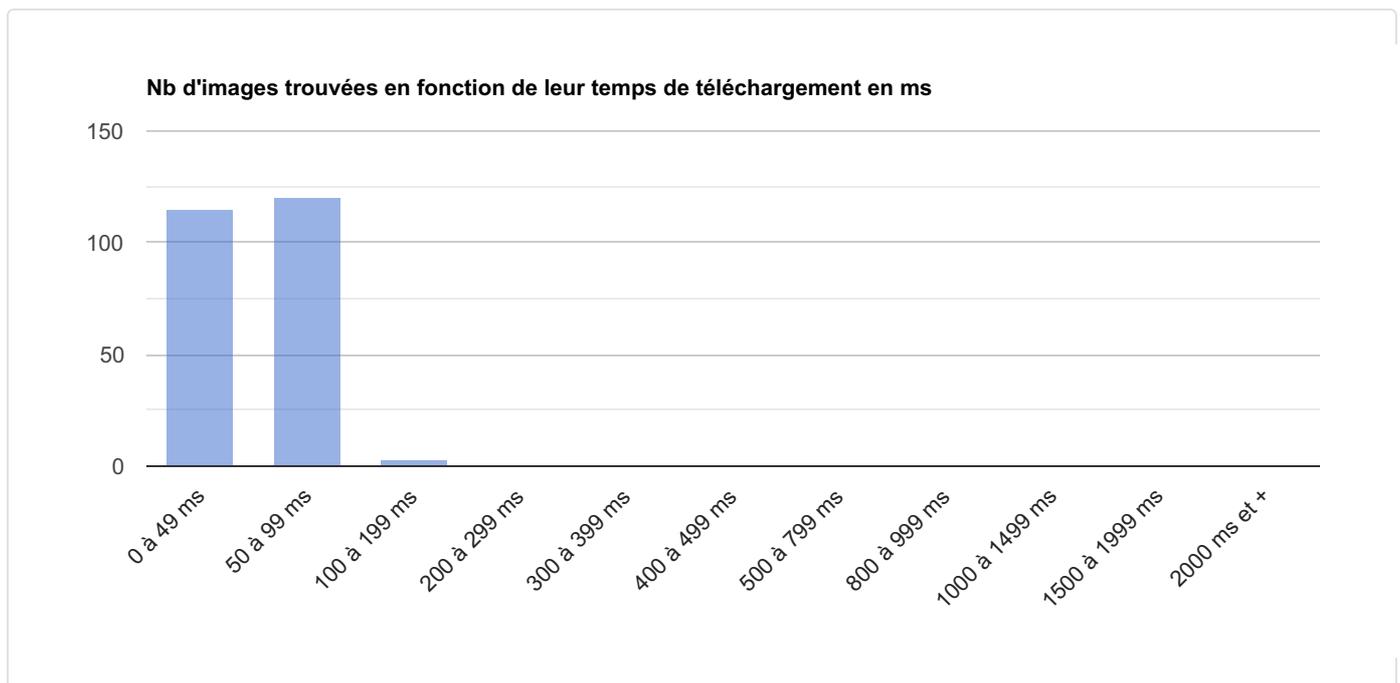
En complément de l'analyse du poids du fichier présentée ci-avant, le temps de téléchargement a été mesuré pour chaque image analysée en code 200.

La moyenne du temps de téléchargement des images est de **52 ms**.

Voici la répartition par tranches de 200 ms :

| Temps de téléchargement (ms) | Nb de pages | Pourcentage des pages |
|------------------------------|-------------|-----------------------|
| 0 à 49 ms | 115 | 48,32 % |
| 50 à 99 ms | 120 | 50,42 % |
| 100 à 199 ms | 3 | 1,26 % |
| 200 à 299 ms | 0 | 0 % |
| 300 à 399 ms | 0 | 0 % |
| 400 à 499 ms | 0 | 0 % |
| 500 à 799 ms | 0 | 0 % |
| 800 à 999 ms | 0 | 0 % |
| 1000 à 1499 ms | 0 | 0 % |
| 1500 à 1999 ms | 0 | 0 % |
| 2000 ms et + | 0 | 0 % |

Voici ce que cela donne sous forme graphique :



9.7 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).

Toutes les images ont été utilisées au moins une fois avec un attribut alt présent : c'est bien ! Il faut néanmoins s'assurer que pour des images stratégiques, cet attribut ne soit pas vide.

77 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.

9.8 Images sans dimensions en HTML

Comme Google le recommande officiellement ([source](#)) pour optimiser le score CLS (Content Layout Shift, critère officiel du positionnement Google depuis mai 2021), les images doivent avoir des dimensions indiquées dans le code HTML (aussi bien width que height).

Aucune image ne pose problème, toutes les images ont les 2 dimensions indiquées dans le code HTML.

9.9 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.

25 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.

9.10 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.

Toutes vos pages HTML en HTTPS contiennent uniquement des images HTTPS : c'est bien.

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

9.11 Bilan des images

Voici les résultats du bilan de notre analyse au niveau des images elles-mêmes (celles autorisées au crawl) :

| | Erreurs | |
|---------------------------------------|----------------|-----|
| Nb d'images inaccessibles | 0 | |
| Nb d'images non indexables | 0 | |
| Nb d'images trop lourdes | 212 | |
| Nb total d'images avec erreurs | 212 | |
| Nb d'images | | 238 |
| Taux d'erreur | 89,08 % | |

9.12 Bilan des pages comportant des images

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

| | Erreurs | Avertissements | |
|-------------------------------------------------------------------------------------------|----------------|-----------------------|-----|
| Nb de pages avec image(s) inaccessible(s) | 0 | n/a | |
| Nb de pages en HTTPS avec image(s) en HTTP | 0 | n/a | |
| Nb de pages avec image(s) sans attribut ALT | 0 | n/a | |
| Nb de pages avec + de 50 images | 0 | n/a | |
| Nb de pages avec image(s) trop lourde(s) | 237 | n/a | |
| Nb de pages avec + de 1 000 Ko d'images | 129 | n/a | |
| Nb de pages avec image(s) n'ayant pas les 2 dimensions indiquées dans le code HTML | 0 | n/a | |
| Nb de pages avec image(s) dont la taille est forcée dans le code HTML | n/a | 237 | |
| Nb total d'images avec erreurs et avertissements | 237 | 237 | |
| Nb de pages avec images | | | 237 |
| Taux d'erreur | 100 % | 100 % | |

10 Analyses avancées

Toutes les analyses qui suivent sont basées uniquement sur les pages indexables découvertes par RM Tech lors du crawl. Quand une page est disponible en HTML et AMP, ses performances prises en compte sont le regroupement des 2 formats.

10.1 QualityRisk

Google améliore sans arrêt son algorithme afin de favoriser les contenus de qualité et pénaliser les autres. C'est un objectif majeur démarré avec l'algorithme Panda en 2011 et qui se poursuit avec les Core Updates.

En se basant sur des aspects techniques mais aussi des éléments du contenu qui nuisent à l'expérience utilisateur, le risque est calculé pour chaque page, noté QR pour "Quality Risk". Plus QR est élevé (le maximum étant 100), plus vous devriez étudier les recommandations fournies en annexe pour améliorer la qualité de la page.

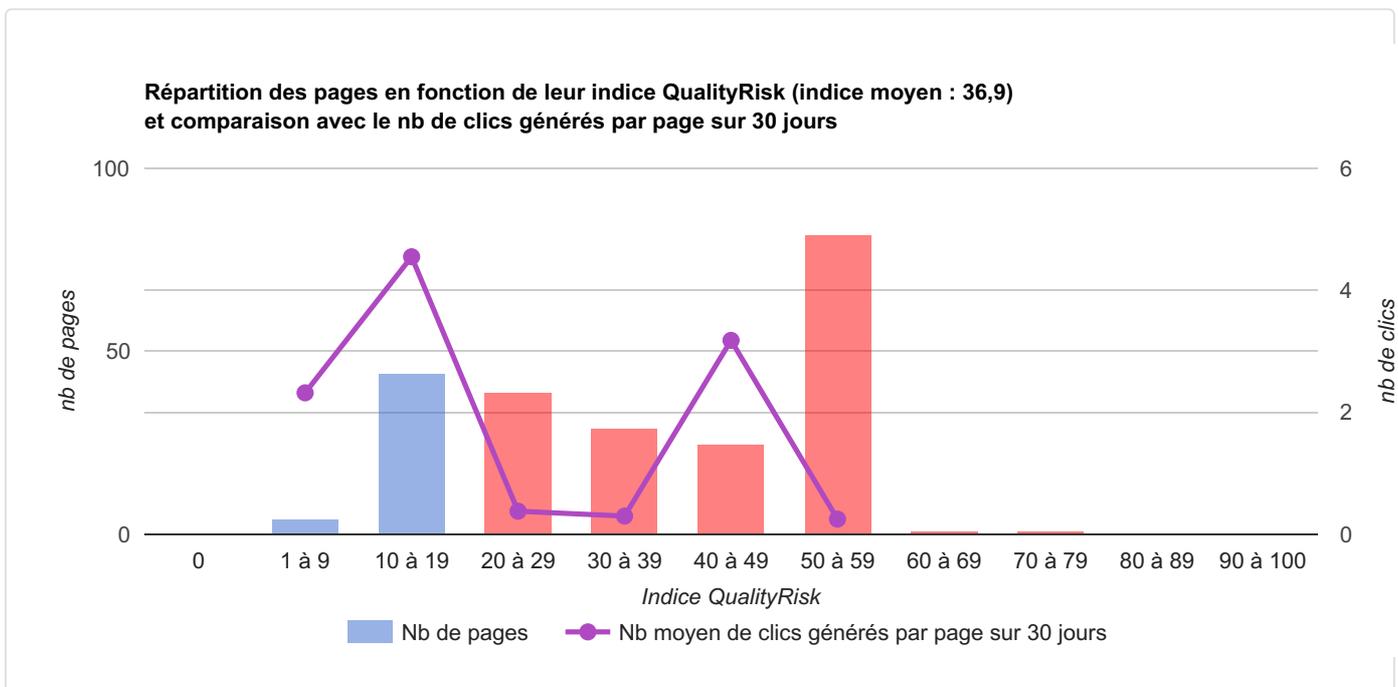
Pour réduire le QualityRisk d'une page, suivez tous les conseils déjà listés précédemment dans ce rapport d'audit. En voici le résumé :

- 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 liens entrants 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 (il est recommandé de ne pas dépasser 300 liens follow)
- 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 ! D'ailleurs, QualityRisk (et donc le score technique de tout l'audit) doit être vu comme une condition nécessaire mais pas suffisante. En effet, il y a pas mal d'autres éléments du SEO non analysés ici, notamment la qualité du contenu, son adéquation avec les intentions de recherche, mais aussi l'UX ou bien sûr les backlinks. Enfin, pour que vos optimisations aient un impact, il faut que Google repasse sur toutes les URL concernées (crawl, indexation, analyse de multiples signaux), ce qui peut prendre du temps.

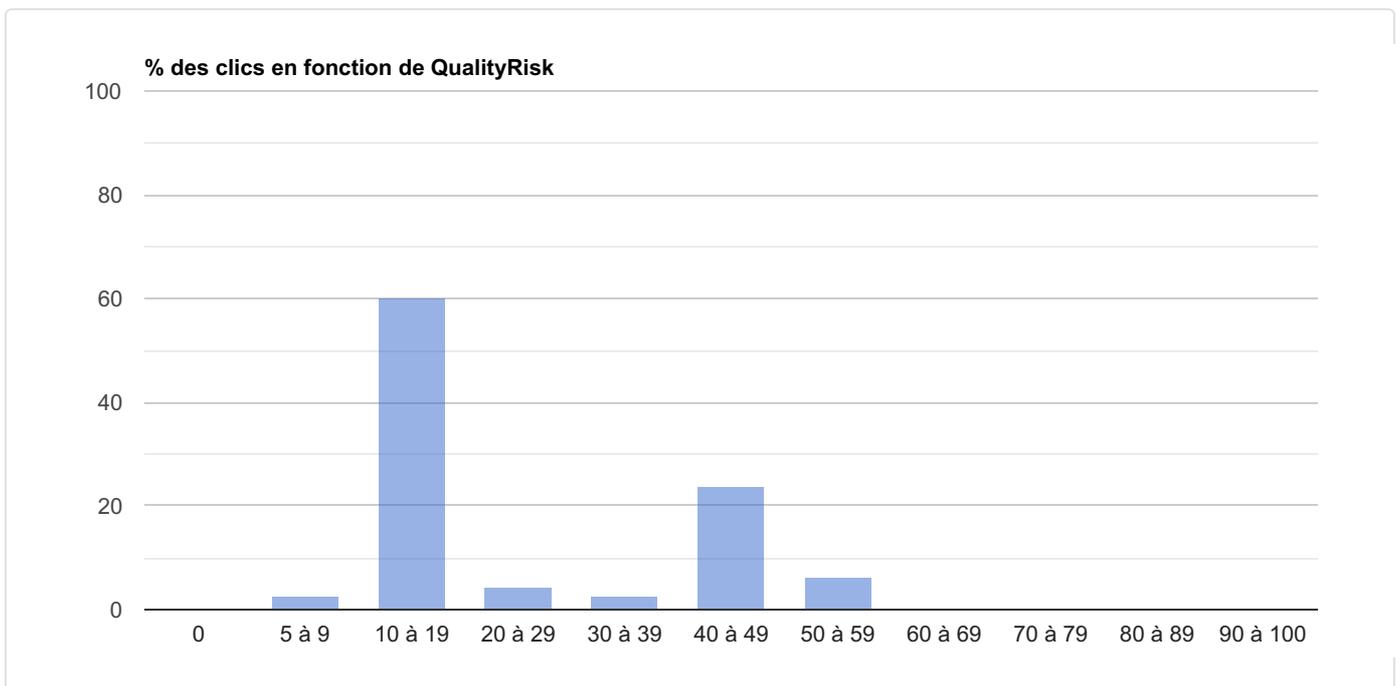
Pour l'ensemble des pages indexables de ce site, **le QualityRisk moyen est de 36,9** (le pire possible étant 100).

L'histogramme ci-dessous montre la répartition des pages du site en fonction de leur indice QualityRisk, ainsi que les statistiques sur le trafic généré :



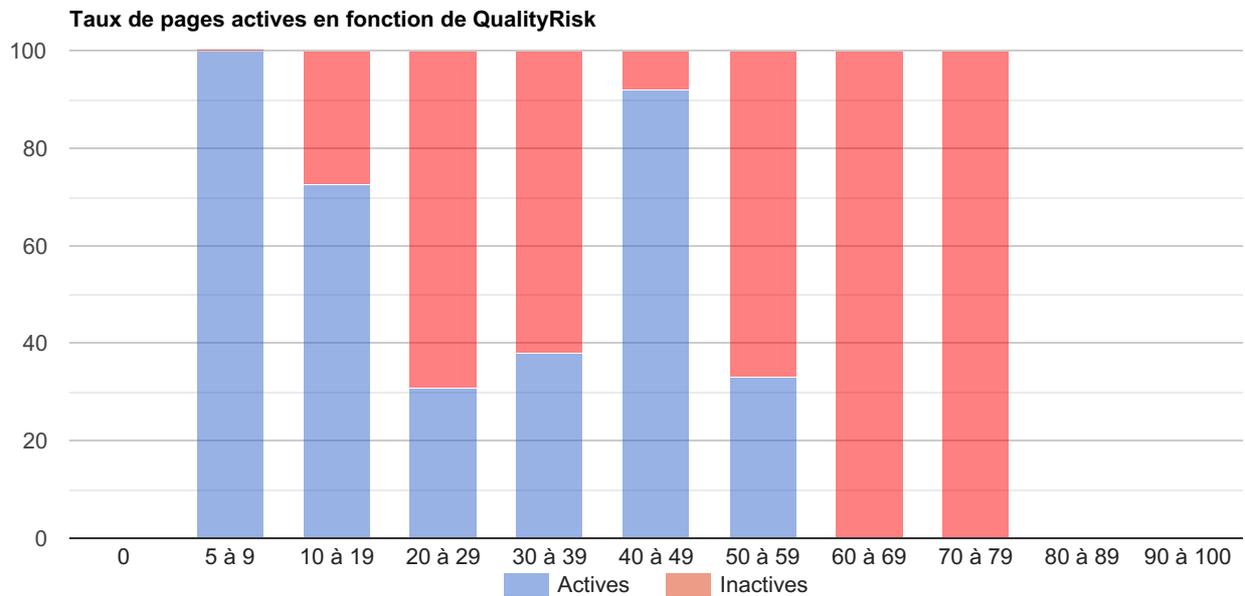
Pour compléter, voici l'étude de la répartition du trafic généré en référencement naturel (= "organic") selon QualityRisk.

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.

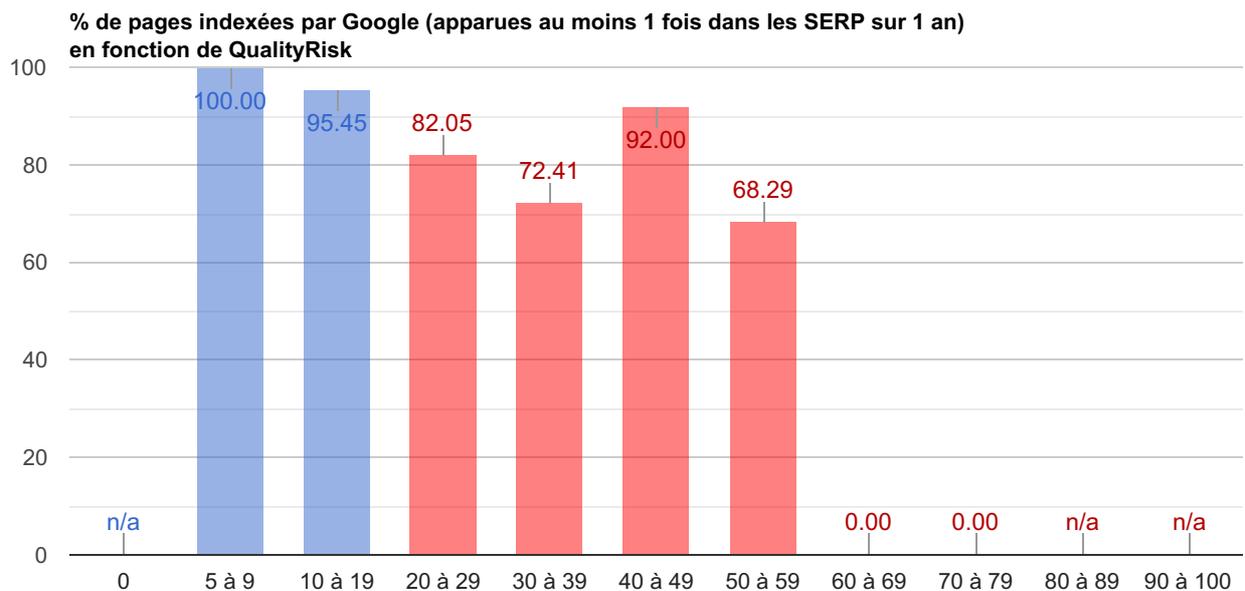


Le graphique qui suit indique le pourcentage de pages actives (en bleu) et de pages inactives (en rouge) selon QualityRisk.

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.



Le graphique ci-dessous indique le pourcentage de pages indexées selon la valeur de QualityRisk. Ici, une page est dite indexée si elle est apparue au moins une fois dans les SERP de Google dans les 365 derniers jours (données Search Console).



Il est conseillé de ne pas avoir de pages avec un QualityRisk supérieur à 20. Le tableau ci-dessous donne la synthèse :

| | Erreurs | Avertissements | |
|------------------------------------------------------|----------------|-----------------------|-----|
| Nb de pages avec un QualityRisk >= 20 | 177 | n/a | |
| Nb de pages avec un QualityRisk entre 1 et 19 | n/a | 48 | |
| Nb de pages indexables | | | 225 |
| Taux d'erreur | 78,67 % | 21,33 % | |

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, sont listées uniquement les pages ayant un QualityRisk supérieur ou égal à 5. Le fichier annexe parisitour_fr-2024-01-31_11_analyse-qualite.xlsx contient les données associées.

10.2 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 (visibilité et trafic organique).

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.**

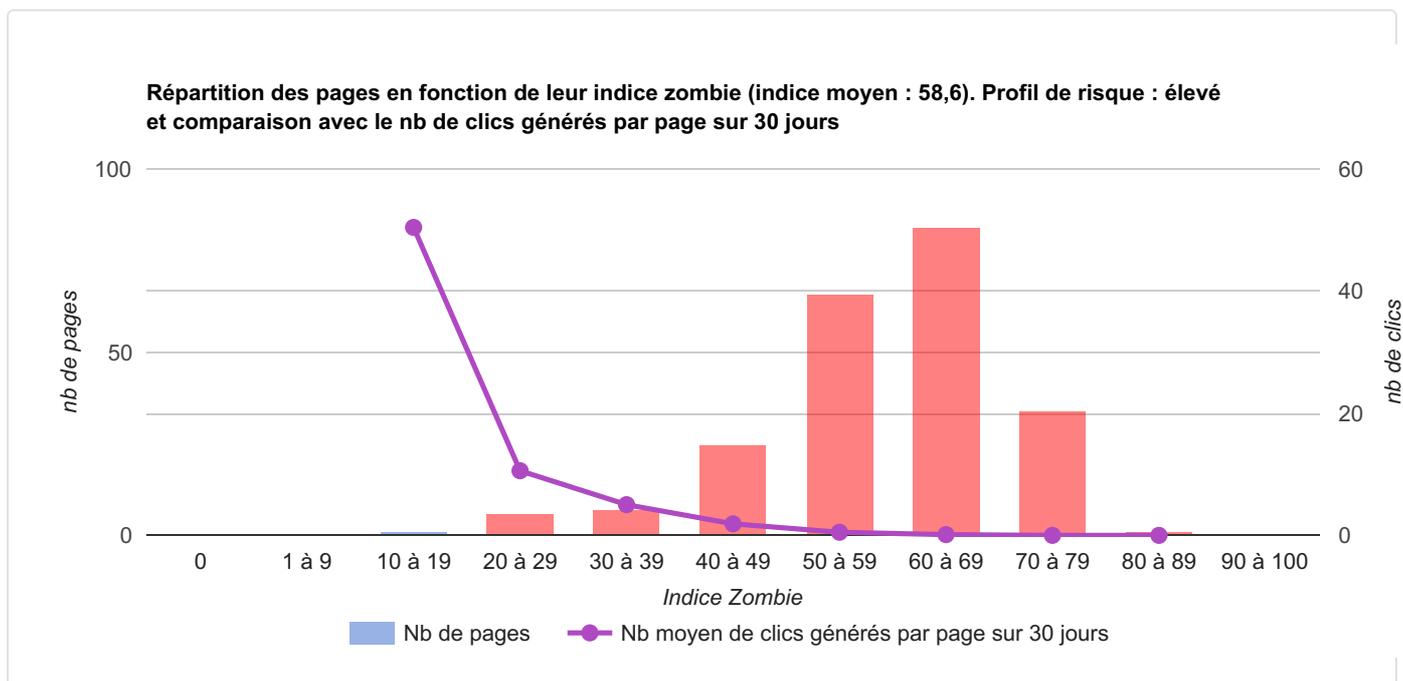
Pour l'analyse des pages zombies, les URL de pagination sont exclues, ainsi que l'URL de départ (souvent la page d'accueil) car elle génère du trafic sur le nom du site, ce qui n'est pas comparable. Parmi les 225 pages indexables trouvées lors du crawl, 1 est exclue. L'analyse des pages zombies porte donc sur 224 URL.

Au global, l'indice zombie moyen est de **58,6**. Le profil de risque est : **élevé**.

L'histogramme ci-dessous montre la répartition des pages du site en fonction de leur indice Zombie, ainsi que les statistiques sur le trafic généré.

Voilà la question à vous poser :

Est-ce là où le site contient le plus de pages (en termes d'indice Zombie) qu'elles génèrent le plus de trafic ?



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

| | Erreurs | Avertissements | |
|--------------------------------------------------------|----------------|-----------------------|-----|
| Nb de pages avec un indice Zombie >= 20 | 223 | n/a | |
| Nb de pages avec un indice Zombie entre 1 et 19 | n/a | 1 | |
| Nb de pages avec un indice Zombie de 0 | n/a | n/a | 0 |
| Nb de pages analysées | | | 224 |
| Taux d'erreur | 99,55 % | 0,45 % | |

Que faire des pages zombies ?

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

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, il vaut mieux 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.
- Il arrive que certaines pages jouant un rôle particulier aient un indice zombie important mais que vous pouvez ignorer (il n'y a rien à faire). C'est par exemple le cas des mentions légales, CGU/CGV, pages de contact...

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.

10.3 Présence dans les SERP Google et trafic généré

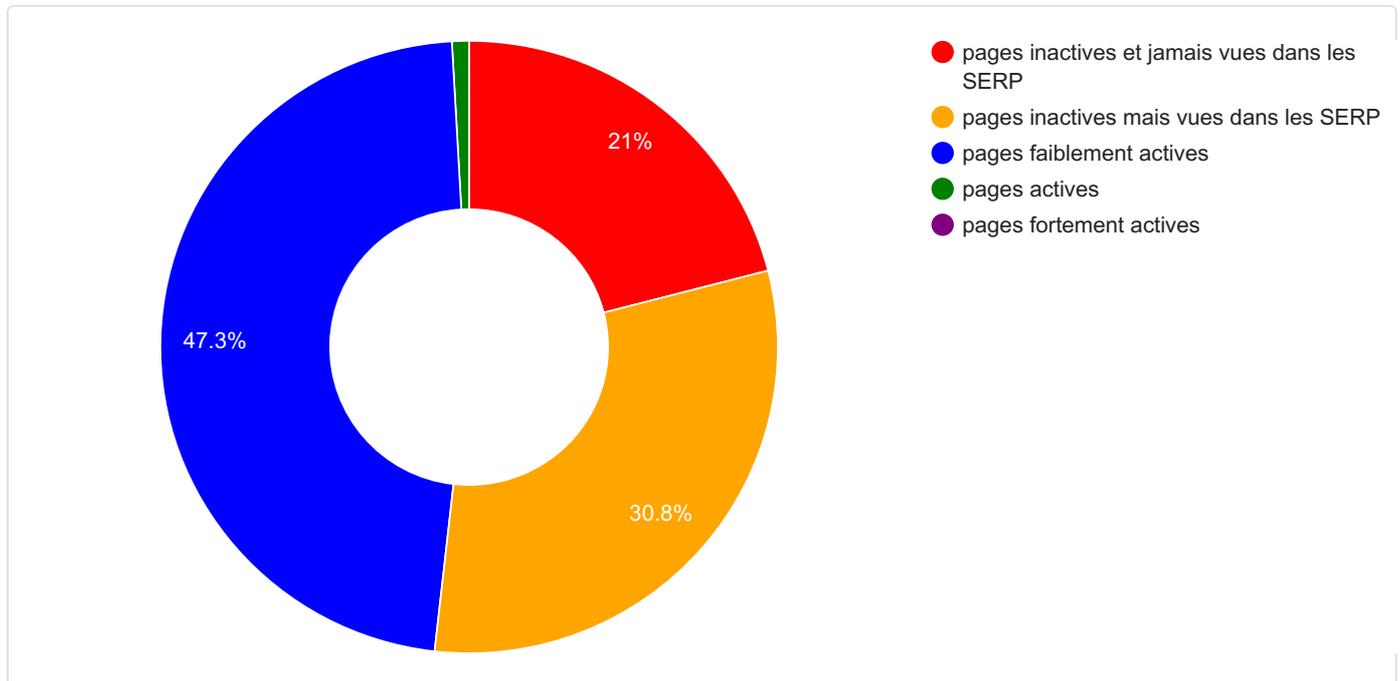
Sur la période étudiée, selon Google Search Console, l'URL de départ a généré **1 130 clics**. **2 923 clics** ont été générés par les 225 autres pages indexables découvertes par cet audit. Voici le bilan :

- Nb de pages inactives et jamais vues dans les SERP : **47**
- Nb de pages inactives mais vues dans les SERP : **69**
- Nb de pages faiblement actives (1 à 364 clics) : **106**
- Nb de pages actives (365 à 3649 clics) : **2**
- Nb de pages fortement actives (3650 clics et +) : **0**

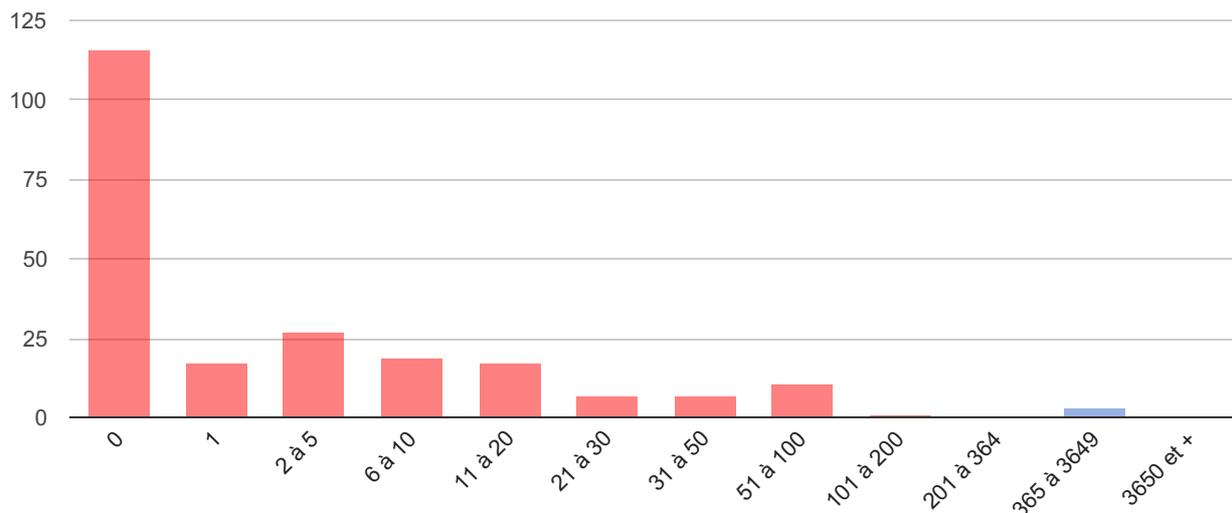
En résumé, il y a donc :

- 108 pages actives, soit 48,2 %
- 116 pages inactives, soit 51,8 %

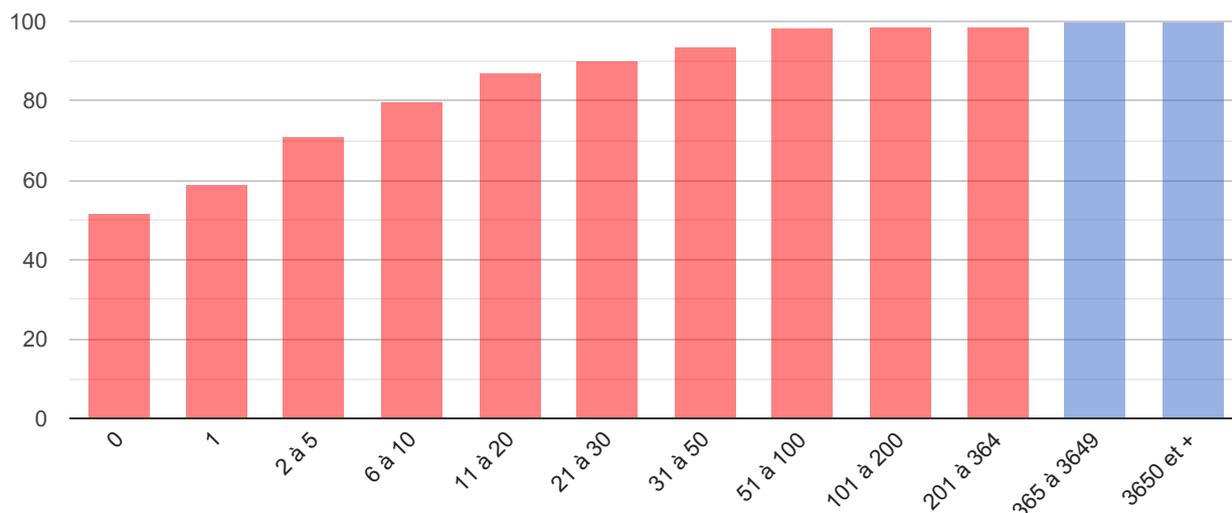
Voici ce que cela donne sous forme graphique :



Nombre de pages selon le nombre de clics générés sur 365 jours



Pourcentage cumulé de pages indexables selon le nombre de clics générés sur 365 jours



Comme vous pouvez le constater sur le graphique ci-dessus, **98,7 %** de vos pages génèrent (chacune) moins d'un clic 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).

10.4 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 l'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 SEO, 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, seules sont sélectionnées 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.

Annexe

Annexe Pages qui pourraient passer en page 1 : parisitour_fr-2024-01-31_search-console-potentiel-premiere-page.xlsx

11 Conclusion

Vous trouverez dans les annexes ci-dessous toutes les données mesurées et calculées par RM Tech sur les pages indexables trouvées pendant l'audit.

Annexe

Le fichier annexe parisitour_fr-2024-01-31_conclusion.xlsx contient les données associées.

L'annexe ci-dessous liste toutes les URL crawlées (indexables ou pas) 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).

Annexe

Le fichier annexe parisitour_fr-2024-01-31_conclusion-maillage-interne.xlsx contient les données associées.

12 Explications complémentaires

12.1 Fonctionnement de l'audit

Pour chaque audit, une analyse exhaustive du site est effectuée, page par page, en partant d'une page de départ sélectionnée, comme Google peut le faire.

Ce fonctionnement reproduit de manière assez fidèle celui de Googlebot (le robot de Google). Par exemple :

- Respect du fichier robots.txt de la même façon que Google (avec donc quelques particularités par rapport au standard)
- Prise en compte :
 - 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)
- Ne sont pas pris en compte :
 - les cookies
 - le javascript (et ajax ou équivalent)
 - le CSS
- Ne sont pas suivis :
 - les formulaires
 - les redirections meta
 - les liens nofollow
- Si un lien autorisé au crawl pointe vers une URL qui n'est pas du HTML (par exemple un PDF mais aussi une image), cette URL sera crawlée. Par contre, l'essentiel de l'analyse se faisant sur les pages HTML, ces URL seront exclues de la plupart des éléments de l'audit.
- Si l'option d'analyse des images est activée, les images étudiées sont celles trouvées dans l'attribut src de la balise , quand cet attribut indique une URL (donc pas sous la forme data:xxx). Les autres attributs de la balise sont ignorés, tout comme les images définies exclusivement par les CSS.
- Pour calculer le nombre de liens entrants internes d'une page (backlinks) :
 - sont recherchées les pages (indexables) qui ont une balise <a> qui amène au final vers la page qu'on étudie (de façon directe, sans redirection)
 - 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 n'est pas 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>

12.2 Signification des codes HTTP

Voici les principaux codes HTTP :

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

12.3 Signification des types MIME

Voici les principaux types MIME, c'est-à-dire les formats de données des documents situés sur un site Internet :

- 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

12.4 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** (ex- Webmaster Tools) : 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](#). 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).
- **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](#).
- **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](#).

- **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](#).
- **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](#).
- **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](#).
- **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](#).
- **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](#).

Table des matières

| | |
|---------------------------------------------------------------|-----------|
| Audit SEO | 1 |
| Annexes | 2 |
| 1 Synthèse des résultats | 3 |
| 1.1 Scores SEO | 3 |
| 1.2 Présence dans les SERP Google et trafic généré | 4 |
| 1.3 Taux d'erreurs SEO | 5 |
| 1.4 Notion d'entonnoir des URL | 6 |
| 1.5 Images | 7 |
| 2 Crawl et indexabilité | 9 |
| 2.1 Bilan des URL bloquées par le(s) fichier(s) robots.txt | 9 |
| Annexe | 9 |
| Annexe | 9 |
| 2.2 Nombre total d'URL crawlées | 9 |
| 2.3 État des URL crawlées | 10 |
| Annexe | 11 |
| Annexe | 11 |
| Annexe | 11 |
| 2.4 URL trouvables uniquement par des liens internes nofollow | 11 |
| 2.5 Types MIME | 11 |
| Annexe | 11 |
| 2.6 Pages HTML crawlées sans erreur | 11 |
| Annexe | 12 |
| 2.7 Pages HTML indexables par les moteurs | 12 |
| Annexe | 12 |
| Annexe | 12 |
| Annexe | 13 |
| 3 Temps de téléchargement | 14 |
| Annexe | 15 |
| 4 Profondeur des pages et maillage interne | 16 |
| 4.1 Analyse de la profondeur des pages | 16 |
| 4.2 Analyse du maillage interne | 18 |
| 4.3 Conclusion sur la profondeur et le maillage interne | 21 |
| Annexe | 21 |
| Annexe | 21 |
| Annexe | 21 |
| 5 Balise title | 22 |
| 5.1 Taille de la balise Title | 22 |
| Annexe | 22 |
| 5.2 Doublons dans la balise title | 22 |
| Annexe | 23 |
| 5.3 Répétition de mots dans la balise Title | 23 |
| Annexe | 23 |
| 5.4 Conclusion sur la balise Title | 23 |
| 6 Balise meta description | 25 |
| 6.1 Taille de la balise meta description | 25 |
| Annexe | 25 |
| 6.2 Doublons dans la balise meta description | 25 |
| Annexe | 25 |
| 6.3 Conclusion sur la balise meta description | 25 |
| 7 Contenu des pages | 27 |
| 7.1 Taille du texte (zone principale de la page) | 27 |
| Annexe | 29 |
| 7.2 Balises H1-H6 | 29 |
| Annexe | 30 |
| 7.3 Doublons dans la balise H1 | 30 |
| Annexe | 30 |
| 7.4 Analyse du nombre de liens par page | 30 |
| Annexe | 31 |

| | |
|---------------------------------------------------------------------------------|-----------|
| 7.5 Analyse des contenus dupliqués internes | 31 |
| Annexe | 32 |
| 7.6 Identifiants de session dans les URL | 32 |
| Annexe | 32 |
| 7.7 Paramètres de tracking dans les URL | 32 |
| Annexe | 33 |
| 7.8 Conclusion sur le contenu des pages | 33 |
| 8 Liens sortants externes | 34 |
| 8.1 Nombre total de liens sortants externes | 34 |
| Annexe | 34 |
| 8.2 Erreurs sur les liens sortants externes | 34 |
| 9 Images | 35 |
| Annexe | 35 |
| Annexe | 35 |
| Annexe | 35 |
| 9.1 Images bloquées par le(s) fichier(s) robots.txt | 35 |
| 9.2 Codes HTTP des URL des images | 35 |
| 9.3 Types MIME (formats d'images) | 36 |
| 9.4 Images indexables | 37 |
| 9.5 Images trop lourdes | 37 |
| 9.6 Temps de téléchargement des images | 38 |
| 9.7 Attribut ALT des images | 39 |
| 9.8 Images sans dimensions en HTML | 39 |
| 9.9 Images avec dimensions forcées en HTML | 40 |
| 9.10 Images HTTP sur pages HTTPS | 40 |
| 9.11 Bilan des images | 40 |
| 9.12 Bilan des pages comportant des images | 40 |
| 10 Analyses avancées | 42 |
| 10.1 QualityRisk | 42 |
| Annexe | 45 |
| 10.2 Pages zombies | 45 |
| Que faire des pages zombies ? | 46 |
| 1. Traitez les cas particuliers | 46 |
| 2. Améliorez la page | 46 |
| 3. Regroupez des pages | 46 |
| 4. Désindexez la page | 46 |
| 5. Supprimez (et désindexez) la page | 47 |
| 10.3 Présence dans les SERP Google et trafic généré | 47 |
| 10.4 Pages du site pouvant passer de la page 2 à la page 1 des résultats Google | 48 |
| Annexe | 49 |
| 11 Conclusion | 50 |
| Annexe | 50 |
| Annexe | 50 |
| 12 Explications complémentaires | 51 |
| 12.1 Fonctionnement de l'audit | 51 |
| 12.2 Signification des codes HTTP | 51 |
| 12.3 Signification des types MIME | 52 |
| 12.4 Signification des termes SEO | 52 |
| Table des matières | 54 |