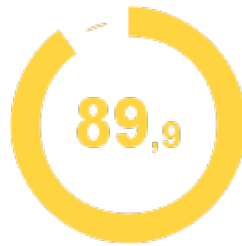




Audit Technique SEO : RM Tech

Site analysé : cdiscount.com

Vous avez encore du
potentiel SEO à
exploiter !



Meilleure note possible :
100

1 Introduction

1.1 Objectif de ce rapport

Ce document a pour objectif de vous présenter les résultats de notre audit technique (sous l'angle du référencement naturel) de votre site. Vous y trouverez une série de recommandations avec de nombreuses explications, ainsi que des annexes contenant la liste exhaustive de tous les éléments à améliorer sur votre site. Ce document a été généré avec notre outil d'audit technique SEO RM Tech (<https://fr.myrankingmetrics.com>).

L'audit de votre site a eu lieu le **10 février 2018** et a porté sur **10 000 URL**. Toutes les analyses qui en découlent doivent donc être considérées en fonction de l'état du site ce jour-là. Par ailleurs, les conditions de l'analyse sont rappelées ci-dessous :

- Site analysé : cdiscount.com
- URL de départ : <https://www.cdiscount.com/>
- Protocoles : aucune restriction (HTTP et HTTPS)
- Nombre maximal d'URL à explorer : 10 000
- Motif d'inclusion : aucun
- Motif d'exclusion : .jpg OR .jpeg OR .png OR .gif
- User-Agent pour le site : Mozilla/5.0 (Windows NT 10.0; WOW64; rv:48.0) Gecko/20100101 Firefox/48.0
- User-Agent pour le robots.txt : Googlebot
- Vitesse de l'exploration : 1.0 requête par seconde
- Ignorer le robots.txt : non
- Prendre en compte les liens externes ayant un attribut rel="nofollow" : non
- Limiter l'exploration au sous-domaine www.cdiscount.com : oui
- CMS déclarés : aucun

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

1.2 Intérêt de cette analyse

Ce rapport de synthèse ainsi que ses annexes vous permettent :

- D'avoir le détail des erreurs techniques encore présentes sur votre site ;
- D'avoir la liste détaillée de toutes vos URL indexables par Google (à comparer au nombre indiqué par Google dans Search Console) ;
- De vous aider à corriger les éventuels problèmes de masse noire ;
- De vous aider à établir un plan de migration / refonte de votre site en cas de changements d'URL ;
- De disposer d'une liste exhaustive de vos URL, avec pour chacune :
 - la balise title
 - la balise meta description
 - l'éventuelle URL canonique
 - l'éventuelle balise meta robots
- D'étudier les niveaux de profondeur de vos URL. Selon nous, c'est un point stratégique, surtout pour les gros sites. Notre rapport vous aide à comprendre quelles URL sont mises en avant dans le maillage interne (notion de hiérarchie). Rappelez-vous que dans l'idéal, vous devriez adapter votre maillage en fonction de la vie du site (effets saisons, produits phares, etc.) ;
- De vous aider à créer un ou plusieurs fichiers sitemaps en fonction de vos besoins.

1.3 Rapport d'audit et annexes contenant toutes les données

Ce rapport d'audit détaille tous les éléments analysés. Les données sont disponibles sous forme exhaustive dans des fichiers annexes au format Excel ainsi qu'au format TSV (Tab-separated values) que vous pouvez facilement ouvrir dans Excel ou tout autre tableur.

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

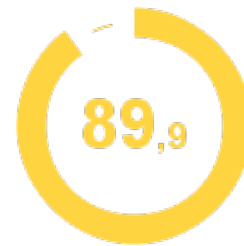
Annexes

Les annexes sont disponibles dans le fichier ZIP [annexes-cdiscount_com-2018-02-10.zip](#).

2 Synthèse des résultats

2.1 Score RM Tech

D'abord, le score RM Tech pour cet audit est de **89,9** sur 100. Il représente le niveau d'optimisation du référencement naturel en ce qui concerne les éléments analysés sur le site (techniques, contenus, liens internes, qualité...). Nous recommandons d'atteindre 95 (score en vert) ; entre 70 et 95 le score est en jaune et en-dessous de 70 il est en rouge.




2.2 QualityRisk

D'après notre algorithme QualityRisk d'évaluation de la qualité, le risque moyen est de **11,0** sur l'ensemble des pages indexables (le pire possible étant 100).









En outre, **5,62 % des pages indexables sont de trop faible qualité** (elles ont un QualityRisk ≥ 20).

Voyez les détails [ici](#).

Élément analysé	Optimisation	Erreurs constatées sur les pages	
		en pourcentage	en nombre de pages
QualityRisk (évaluation de la qualité)		5,62 % ⁽³⁾	499

2.3 Taux d'erreur

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

Élément analysé	Optimisation	Erreurs constatées sur les pages	
		en pourcentage	en nombre de pages
URL répondant en erreur		2,70 % ⁽¹⁾	270
Indexabilité des pages HTML		8,67 % ⁽²⁾	844
Temps de téléchargement		83,72 % ⁽²⁾	8 146
Profondeur et maillage interne		38,48 % ⁽³⁾	3 419
Balise title		88,01 % ⁽³⁾	7 821
Balise meta description		12,05 % ⁽³⁾	1 071
Contenu		11,21 % ⁽³⁾	996
Liens sortants externes		0,72 % ⁽⁴⁾	43

- (1) pourcentage calculé par rapport au nombre d'URL explorées, soit ici 10 000 URL
- (2) pourcentage calculé par rapport au nombre d'URL de pages HTML sans erreur (code 200), soit ici 9 730 URL
- (3) pourcentage calculé par rapport au nombre d'URL de pages HTML indexables, soit ici 8 886 URL
- (4) pourcentage calculé par rapport au nombre d'URL externes distinctes trouvées dans les liens sortants follow, soit ici 6 012 URL

Explications sur le code couleur :

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

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

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

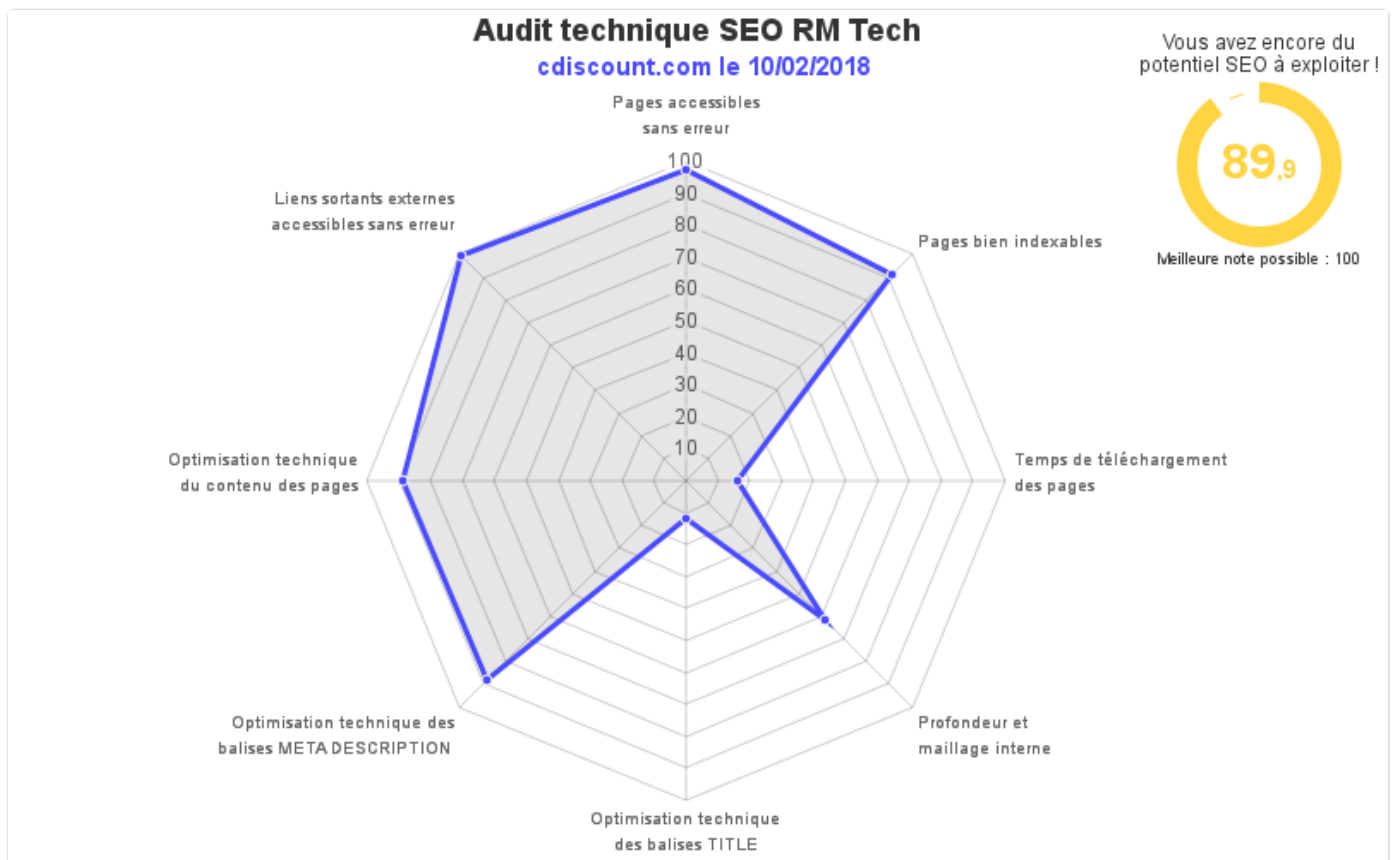
2.4 Performance

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

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

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

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



3 Crawl et indexabilité

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

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

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

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

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

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

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

Fichier robots.txt	Code HTTP	Nombre d'URL bloquées
https://www.cdiscount.com/robots.txt	200	128
http://www.cdiscount.com/robots.txt	200	0
Total		128

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

Annexe

Le fichier annexe cdiscount_com-2018-02-10-url-bloquees-robotstxt.xlsx contient la liste des URL bloquées au crawl.

Annexe

Le fichier annexe cdiscount_com-2018-02-10-robotstxt.zip contient l'ensemble des fichiers robots.txt récupérés par RM Tech.

3.2 Nombre total d'URL crawlées

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

Dans votre cas, nous nous sommes limités aux 10 000 premières URL rencontrées pendant notre crawl. Tout le reste de ce rapport concerne donc uniquement les pages que nous avons pu crawler.

Voici la répartition selon les protocoles HTTP et HTTPS :

Protocole	Nb d'URL	% des URL crawlées
HTTP	4	0,04 %
HTTPS	9 996	99,96 %

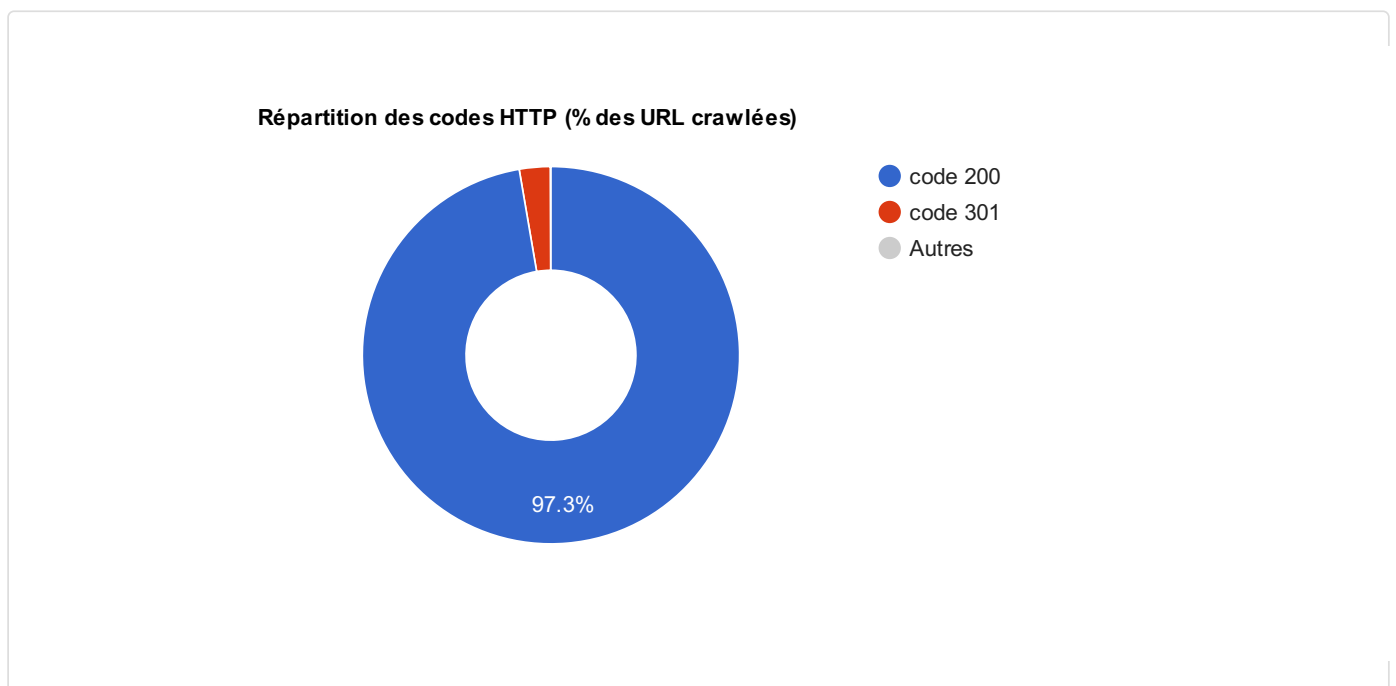
3.3 État des URL crawlées

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

Code HTTP	Nb d'URL en HTTP	Nb d'URL en HTTPS	Nb d'URL (HTTP et HTTPS)	% des URL crawlées
200	0	9 730	9 730	97,30 %
301	4	260	264	2,64 %
404	0	6	6	0,06 %
Total erreurs 3xx	4	260	264	2,64 %
Total erreurs 4xx	0	6	6	0,06 %
Total erreurs	4	266	270	2,70 %

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

Voici ce que cela donne sous forme graphique :



Annexe

Le fichier annexe cdiscount_com-2018-02-10-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 cdiscount_com-2018-02-10-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 cdiscount_com-2018-02-10-liens-vers-erreurs-3xx.xlsx contient les URL des pages contenant les liens en erreur.

3.4 Types MIME

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

Type MIME	Nb d'URL	% des URL crawlées en code 200
text/html	9 730	100 %

Annexe

Le fichier annexe cdiscount_com-2018-02-10-types-mime.xlsx contient les données associées.

3.5 Pages HTML crawlées sans erreur

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

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

Annexe

Le fichier annexe cdiscount_com-2018-02-10-pages-sans-erreur.xlsx contient les données associées.

3.6 Pages HTML indexables par les moteurs

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

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

Les pages HTML indexables sont toutes celles :

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

Voici le décompte pour votre site :

	Nb d'URL	% des URL HTML code 200
Pages HTML avec code HTTP 200 (A)	9 730	
Dont pages interdites d'indexation	149	1,53 %
Dont pages avec URL canonique différente de l'URL crawlée	713	7,33 %
Nb total de pages non-indexables (B)	844	8,67 %
Nombre de pages HTML indexables (A - B)	8 886	91,33 %

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

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

Dans votre cas, nous avons identifié 149 pages HTML interdites d'indexation représentant 1,53 % des pages HTML en code 200. Nous vous conseillons de vérifier si c'est normal d'interdire l'indexation de ces pages. Vous trouverez la liste des URL concernées dans les annexes.

Dans votre cas, nous avons identifié 713 pages HTML avec une URL canonique différente de l'URL crawlée, représentant 7,33 % des pages HTML en code 200. Nous vous conseillons de 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 cdiscount_com-2018-02-10-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 cdiscount_com-2018-02-10-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 cdiscount_com-2018-02-10-liens-vers-erreurs-balise-canonique.xlsx contient les URL des pages contenant les liens vers ces URL.

4 Temps de téléchargement

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

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

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

Voici les résultats de notre analyse :

	Nb d'URL	% des URL HTML code 200
Pages trop lentes (> 600 ms)	8 146	83,72 %
dont : pages vraiment trop lentes (> 1000 ms)	7 707	79,21 %
Nb de pages HTML code 200	9 730	
Taux d'erreur		83,72 %

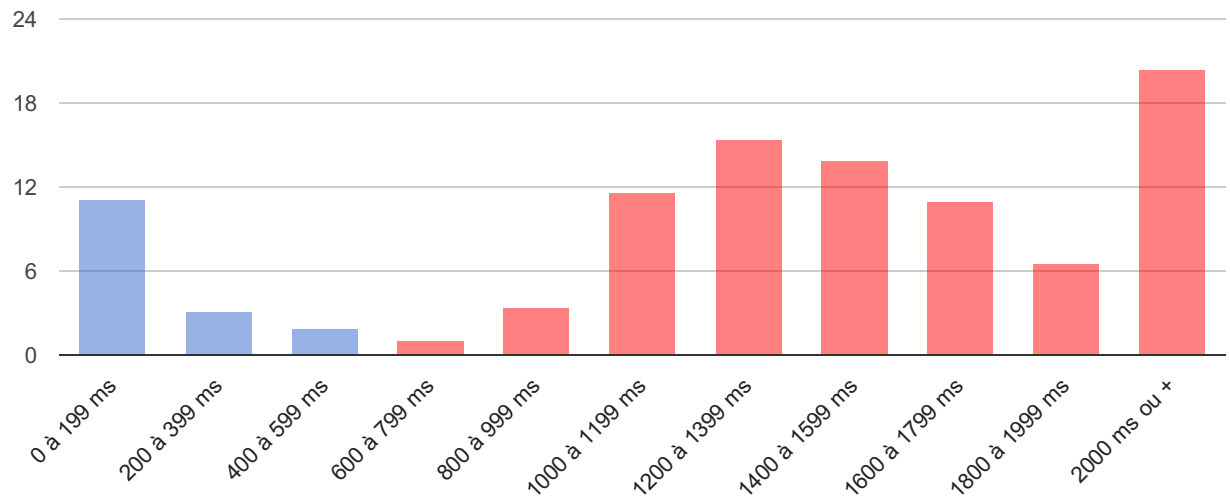
Remarque : il ne s'agit pas ici du temps de chargement complet de la page, puisque nous ignorons tous les fichiers externes (Javascript, CSS, images, etc.). Il s'agit du temps mis par le serveur pour répondre à la requête, générer la page HTML et la télécharger. C'est donc le même principe que le temps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

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

Temps de téléchargement (ms)	Nb de pages	Pourcentage des pages
de 0 ms à 199 ms	1 080	11,10 %
de 200 ms à 399 ms	309	3,18 %
de 400 ms à 599 ms	192	1,97 %
de 600 ms à 799 ms	110	1,13 %
de 800 ms à 999 ms	332	3,41 %
de 1000 ms à 1199 ms	1 137	11,69 %
de 1200 ms à 1399 ms	1 506	15,48 %
de 1400 ms à 1599 ms	1 360	13,98 %
de 1600 ms à 1799 ms	1 067	10,97 %
de 1800 ms à 1999 ms	647	6,65 %
2000 ms ou +	1 990	20,45 %

Voici ce que cela donne sous forme graphique :

Pourcentage des pages en fonction du temps de téléchargement



Annexe

Le fichier annexe cdiscount_com-2018-02-10-temps-telechargement.xlsx contient les données associées.

5 Profondeur des pages et maillage interne

5.1 Analyse de la profondeur des pages

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

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

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

Attention : dans votre cas, nous n'avons pas pu étudier toutes les URL de votre site, en raison de la limite appliquée sur le nombre d'URL à explorer. Cela signifie qu'en réalité, il existe d'autres URL plus profondes que celles mentionnées dans notre rapport. Nous vous recommandons de refaire plus tard une nouvelle analyse en augmentant le seuil du nombre d'URL à explorer : c'est la seule façon de vérifier sérieusement combien vous avez d'URL trop profondes (et lesquelles).

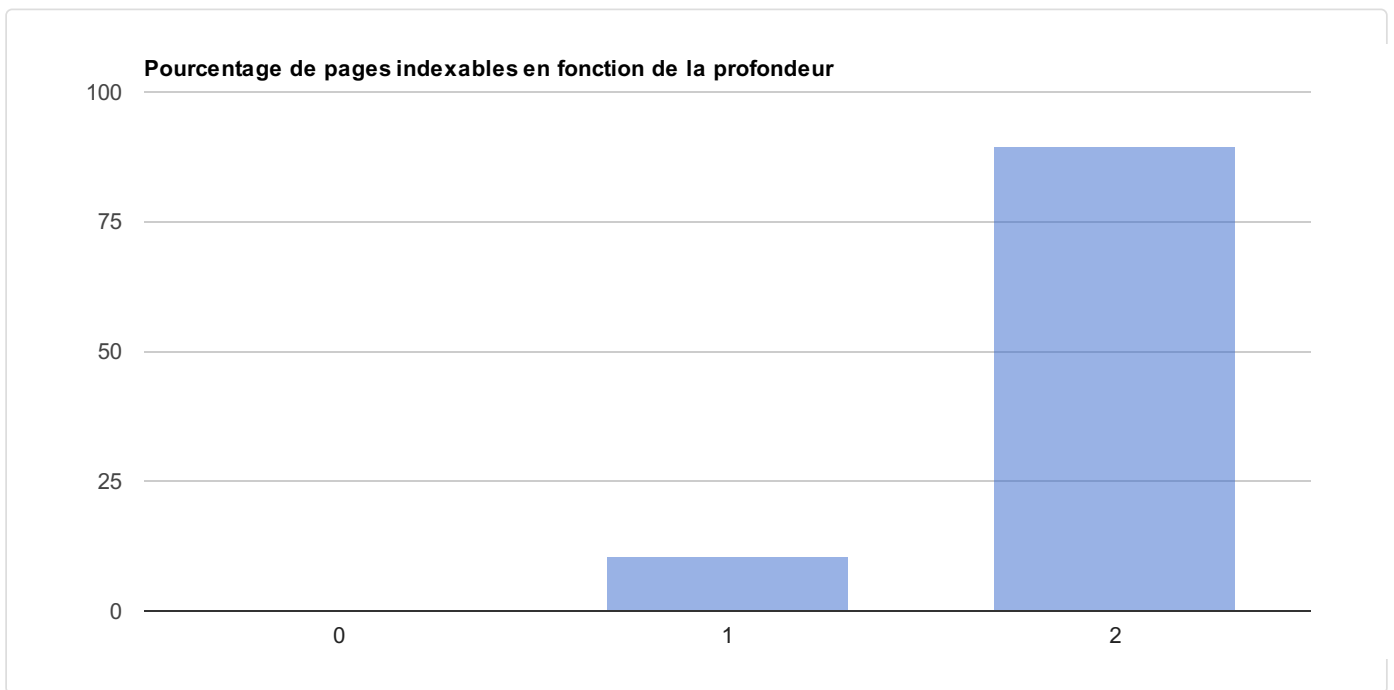
L'étude a porté sur les pages HTML indexables.

Voici les résultats selon les niveaux :

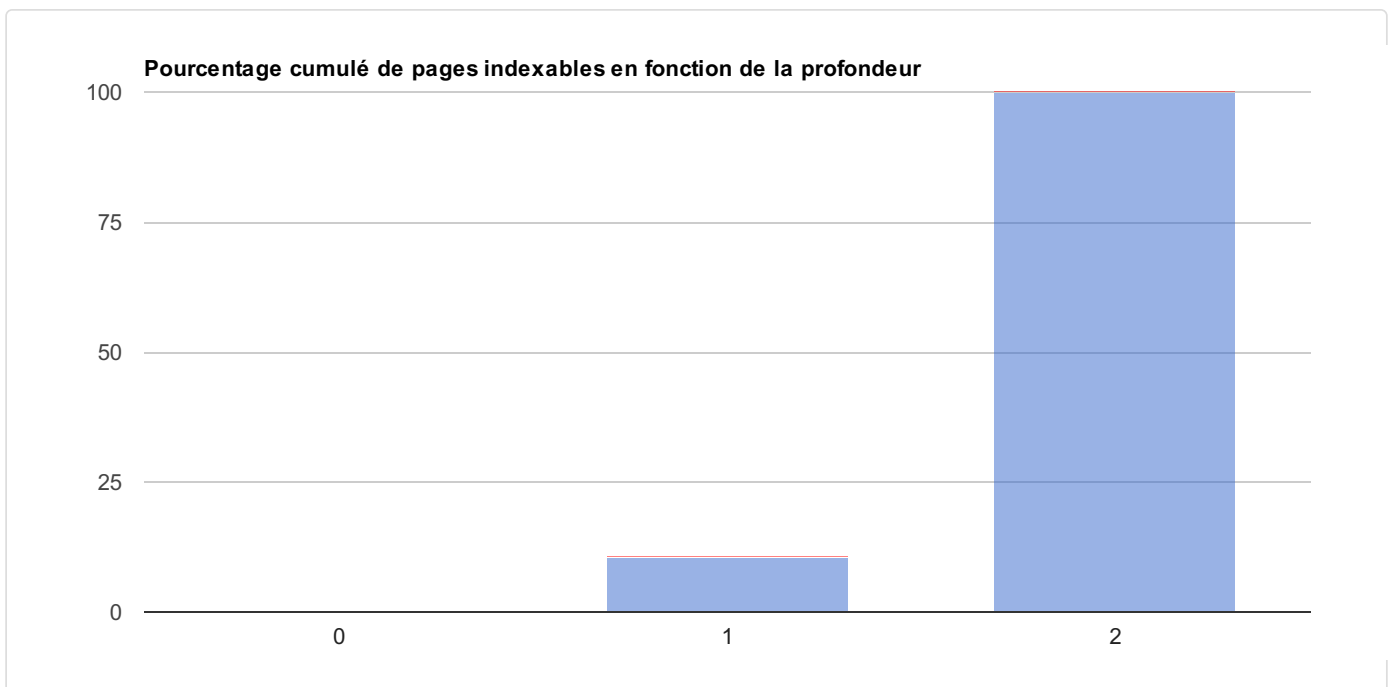
Profondeur	Nb de pages	Nb cumulé de pages	Pourcentage des pages	Pourcentage cumulé
0	1	1	0,01 %	0,01 %
1	935	936	10,52 %	10,53 %
2	7 950	8 886	89,47 %	100 %

Dans votre cas, aucune page n'est trop profonde.

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



Le graphique ci-dessous montre le pourcentage cumulé :



5.2 Analyse du maillage interne

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

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

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

Remarque importante : attention, comme pour la profondeur, nous ne conseillons pas d'augmenter artificiellement le

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

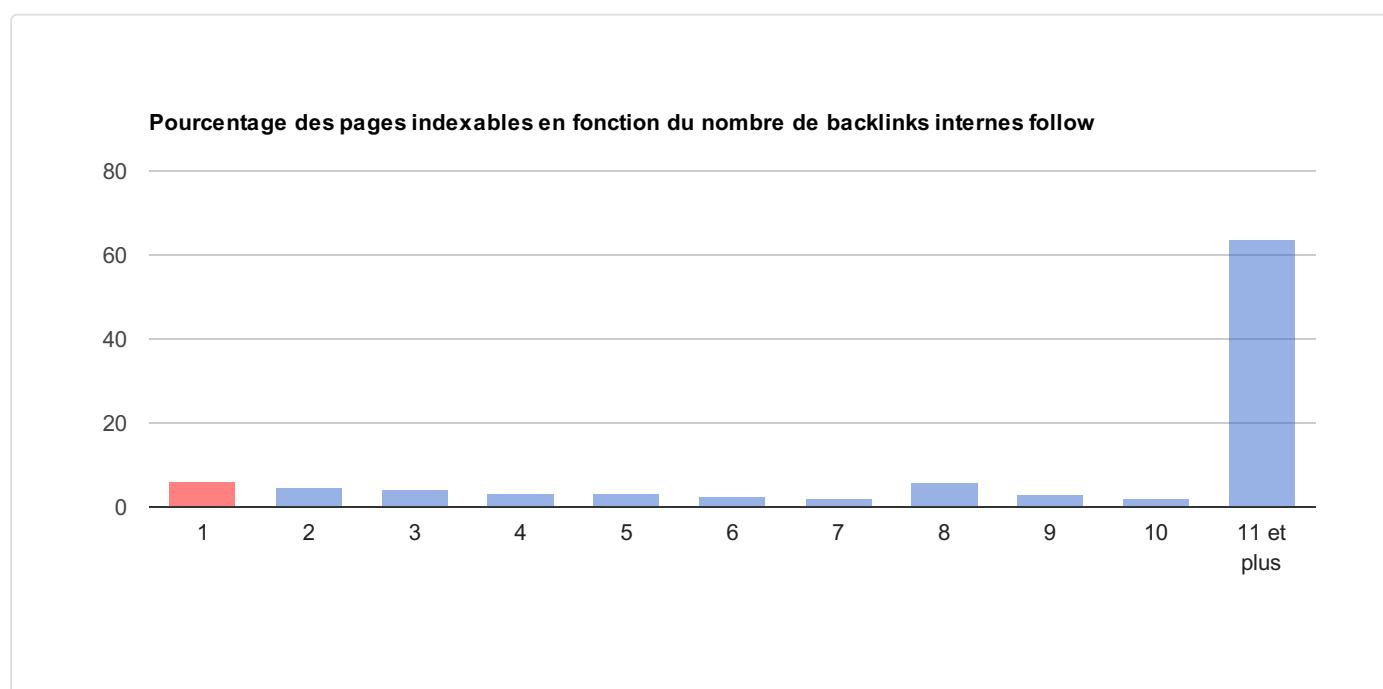
Nous avons étudié les pages qui reçoivent trop peu de backlinks internes ou qui ont trop peu de variété dans les ancres de liens internes (liens follow uniquement).

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

Voici les résultats de notre analyse :

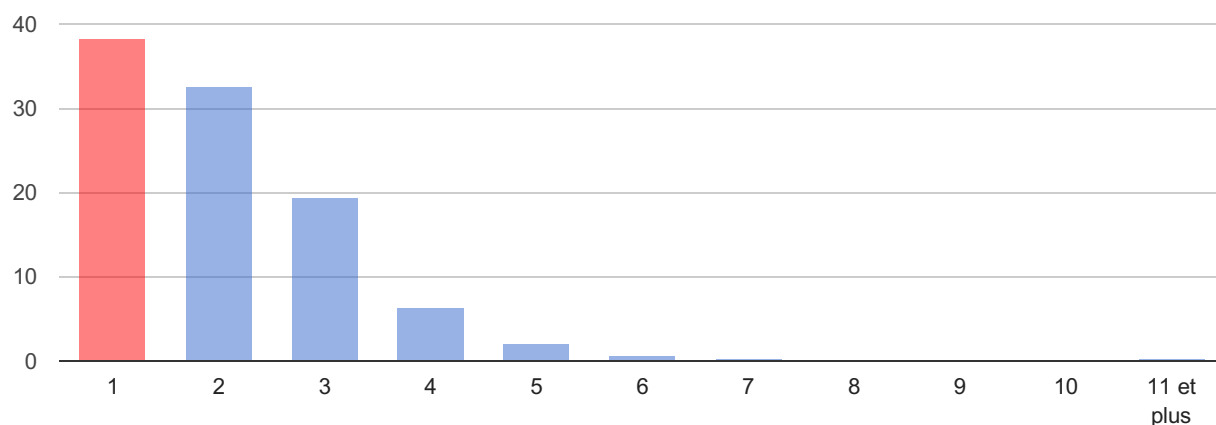
	Erreurs	Avertissements	
Nb de pages avec seulement 1 backlink interne	536	n/a	
Nb de pages avec seulement 2 backlinks internes	n/a	405	
Nb de pages avec toujours la même ancre (backlinks internes)	3 411	n/a	
Nb de pages avec seulement 2 ancres distinctes (backlinks internes)	n/a	2 910	
Nb total d'erreurs et d'avertissements	3 419	3 116	
Nb de pages indexables			8 886
Taux d'erreur	38,48 %	35,07 %	

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



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

Pourcentage des pages indexables en fonction du nombre d'ancres distinctes dans leurs backlinks internes follow



5.3 Conclusion sur la profondeur et le maillage interne

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

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

Voici les résultats de notre analyse :

	Erreurs	
Nb de pages trop profondes ou avec un maillage interne insuffisant	3 419	
Nb de pages indexables		8 886
Taux d'erreur	38,48 %	

Annexe

Le fichier annexe cdiscount_com-2018-02-10-profondeur-maillage.xlsx contient les données associées.

6 Balise title

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

6.1 Taille de la balise Title

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

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
Nb de pages sans balise title ou avec une balise title vide	0	n/a	
Nb de pages avec seulement 1 mot dans title	2	n/a	
Nb de pages avec seulement 2 mots dans title	1	n/a	
Nb de pages avec seulement 3 mots dans title	6	n/a	
Nb de pages avec entre 71 et 80 caractères dans title	n/a	708	
Nb de pages avec entre 81 et 100 caractères dans title	3 326	n/a	
Nb de pages avec + de 100 caractères dans title	4 451	n/a	
Nb total d'erreurs et d'avertissements	7 786	708	
Nb de pages indexables			8 886
Taux d'erreur	87,62 %	7,97 %	

Annexe

Le fichier annexe cdiscount_com-2018-02-10-balise-title.xlsx contient les données associées.

6.2 Doublons dans la balise title

Pour optimiser la balise title, il faut que chaque page ait une balise title unique, sinon nous considérons cela comme une erreur.

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
Nb de balises title utilisées plus d'une fois sur le site (hors balise vide)	32	n/a	
Nb de pages indexables par Google concernées (balise vide incluse)	81	n/a	
Nb de pages indexables			8 886
Taux d'erreur	0,91 %	n/a	

Annexe

Le fichier annexe cdiscount_com-2018-02-10-meme-title.xlsx contient les données associées.

6.3 Répétition de mots dans la balise Title

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

	Erreurs	Avertissements	
Nb de pages avec 2 fois un même mot dans title	n/a	5 734	
Nb de pages avec 3 fois un même mot dans title	731	n/a	
Nb de pages avec 4 fois un même mot dans title	80	n/a	
Nb de pages avec 5 fois ou plus un même mot dans title	17	n/a	
Nb total d'erreurs et d'avertissements	828	5 734	
Nb de pages indexables			8 886
Taux d'erreur	9,32 %	64,53 %	

Annexe

Le fichier annexe cdiscount_com-2018-02-10-meme-mot-dans-title.xlsx contient les données associées.

6.4 Votre marque dans la balise Title

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

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
Nb de pages qui contiennent la marque "cdiscout" dans title	n/a	8 854	
Nb de pages indexables			8 886
Taux d'erreur		99,64 %	

Annexe

Le fichier annexe cdiscount_com-2018-02-10-marque-dans-title.xlsx contient les données associées.

6.5 Conclusion sur la balise Title

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

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

Voici les résultats de notre analyse :

	Erreurs	
Nb de pages avec une ou plusieurs erreurs dans la balise Title	7 821	
Nb de pages indexables		8 886
Taux d'erreur	88,01 %	

7 Balise meta description

7.1 Taille de la balise meta description

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

Voici les résultats de notre analyse :

	Erreurs	
Nb de pages avec balise meta description vide ou absente	119	
Nb de pages avec une meta description de moins de 90 caractères	10	
Nb total	129	
Nb de pages indexables		8 886
Taux d'erreur	1,45 %	

Annexe

Le fichier annexe cdiscount_com-2018-02-10-meta-desc.xlsx contient les données associées.

7.2 Doublons dans la balise meta description

Pour optimiser la balise meta description, il faut qu'elle soit unique.

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
Nb de balises meta description utilisées plus d'une fois (hors balise vide)	102	n/a	
Nb de pages concernées	942	n/a	
Nb de pages indexables			8 886
Taux d'erreur	10,60 %	n/a	

Annexe

Le fichier annexe cdiscount_com-2018-02-10-meme-meta-desc.xlsx contient les données associées.

7.3 Conclusion sur la balise meta description

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

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

Voici les résultats de notre analyse :

	Erreurs	
Nb de pages avec une ou plusieurs erreurs dans la balise meta description	1 071	
Nb de pages indexables		8 886
Taux d'erreur	12,05 %	

8.1 Balises H1-H6

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

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

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

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

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

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
Nb de pages sans balise H1	35	0	
Nb de pages avec une balise H1 mais vide	23	n/a	
Nb de pages où la 1ère balise Hn n'est pas une H1	n/a	35	
Nb de pages avec saut de niveau entre les Hn	n/a	3	
Nb de pages avec une balise Hn trop longue	n/a	2 895	
Nb de pages avec plusieurs balises H1	n/a	47	
Nb de pages avec 1 seule balise H2-H6	n/a	7 251	
Nb de pages avec le même libellé dans plusieurs Hn	n/a	65	
Nb de pages avec le même libellé dans H1 et Title	n/a	29	
Nb de pages indexables			8 886
Taux d'erreur	0,65 %	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 cdiscount_com-2018-02-10-balisesh.xlsx contient les données associées.

8.2 Doublons dans la balise H1

Pour optimiser la balise H1, il faut que chaque page ait une balise H1 unique, sinon nous considérons cela comme une erreur.

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
Nb de balises H1 utilisées plus d'une fois sur le site (hors balise vide)	97	n/a	
Nb de pages indexables par Google concernées (balise vide incluse)	936	n/a	
Nb de pages indexables			8 886
Taux d'erreur	10,53 %	n/a	

Annexe

Le fichier annexe cdiscount_com-2018-02-10-doublons-h1.xlsx contient les données associées.

8.3 Taille du texte (nb de mots)

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

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

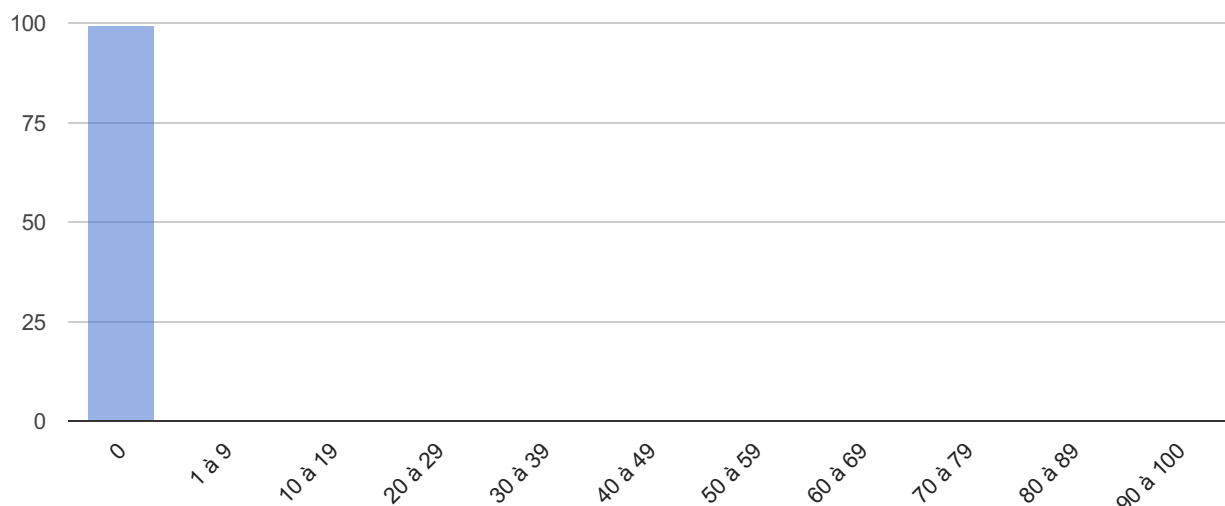
Une approche simpliste consiste à compter le nombre de mots dans l'ensemble de la page HTML (incluant l'entête, les menus, les barres latérales et le pied de page). Comme vous vous en doutez, cela ne fonctionne pas car ce qui compte, c'est la taille du contenu dans la zone principale de la page.

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

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

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

Pourcentage de pages selon leur indice de manque de contenu, de 0 (bien) à 100 (le pire)



Le tableau ci-dessous donne la synthèse :

	Erreurs	Avertissements	
Nb de pages sans doute trop courtes	35	n/a	
Nb de pages indexables			8 886
Taux d'erreur	0,39 %	n/a	

Annexe

Le fichier annexe cdiscount_com-2018-02-10-nb-mots-body.xlsx contient les données associées.

8.4 Mise en exergue (strong ou bold)

Pour vous donner le plus de chances de réussir votre référencement, c'est en général mieux si chaque page contient quelques mots mis en exergue (gras ou bold). C'est néanmoins un point mineur dans l'optimisation du référencement (c'est pourquoi nous considérons cela uniquement comme un avertissement). Utilisez la mise en exergue (balise strong) pour mettre en évidence certains mots ou expressions dans la page, avant tout pour capter l'attention de l'internaute.

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
Nb de pages qui ne contiennent aucun mot en gras (strong ou bold)	n/a	0	
Nb de pages indexables			8 886
Taux d'erreur	n/a	0 %	

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

8.5 Analyse du nombre de liens par page

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

Nous cherchons à identifier 4 types de problèmes :

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

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

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
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	3 467	
Nb de pages avec plus de 10 liens externes	n/a	0	
Nb de pages avec 1 ou plusieurs liens internes nofollow	n/a	8 886	
Nb total	0	8 886	
Nb de pages indexables			8 886
Taux d'erreur	0 %	100 %	

Le fichier annexe cdiscount_com-2018-02-10-nb-liens.xlsx contient les données associées.

8.6 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).

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

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
Nb d'URL de pages HTML ayant un contenu strictement identique	0	n/a	
Nb de pages indexables			8 886
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.

8.7 Identifiants de session dans les URL

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

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
Nb de pages avec un ID de session	0	n/a	
Nb de pages indexables			8 886
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.

8.8 Paramètres de tracking dans les URL

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

Voici les résultats de notre analyse :

	Erreurs	Avertissements	
Nb de pages avec une variable de tracking	0	n/a	
Nb de pages indexables			8 886
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.

8.9 Conclusion sur le contenu des pages

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

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

Voici les résultats de notre analyse :

	Erreurs	
Nb de pages avec une ou plusieurs erreurs dans le contenu	996	
Nb de pages indexables		8 886
Taux d'erreur	11,21 %	

9 Liens sortants externes

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

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

9.1 Nombre total de liens sortants externes

Nous avons identifié 21 098 liens externes pointant vers 6 012 URL distinctes.

Annexe

Le fichier annexe `cdiscount_com-2018-02-10-liste-liens-externes.xlsx` contient les données associées.

9.2 Erreurs sur les liens sortants externes

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

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

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

Code HTTP	Nb d'URL	% des URL externes
301	11	0,18 %
302	12	0,20 %
403	1	0,02 %
500	36	0,60 %
503	4	0,07 %
504	2	0,03 %

	Erreurs	Avertissements	
Nombre total de liens sortants externes redirigés	n/a	23	
Nombre total de liens sortants externes en erreur	43	n/a	
Nombre total de liens sortants externes			6 012
Taux d'erreur	0,72 %	0,38 %	

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

10 Analyse de la qualité

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

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

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

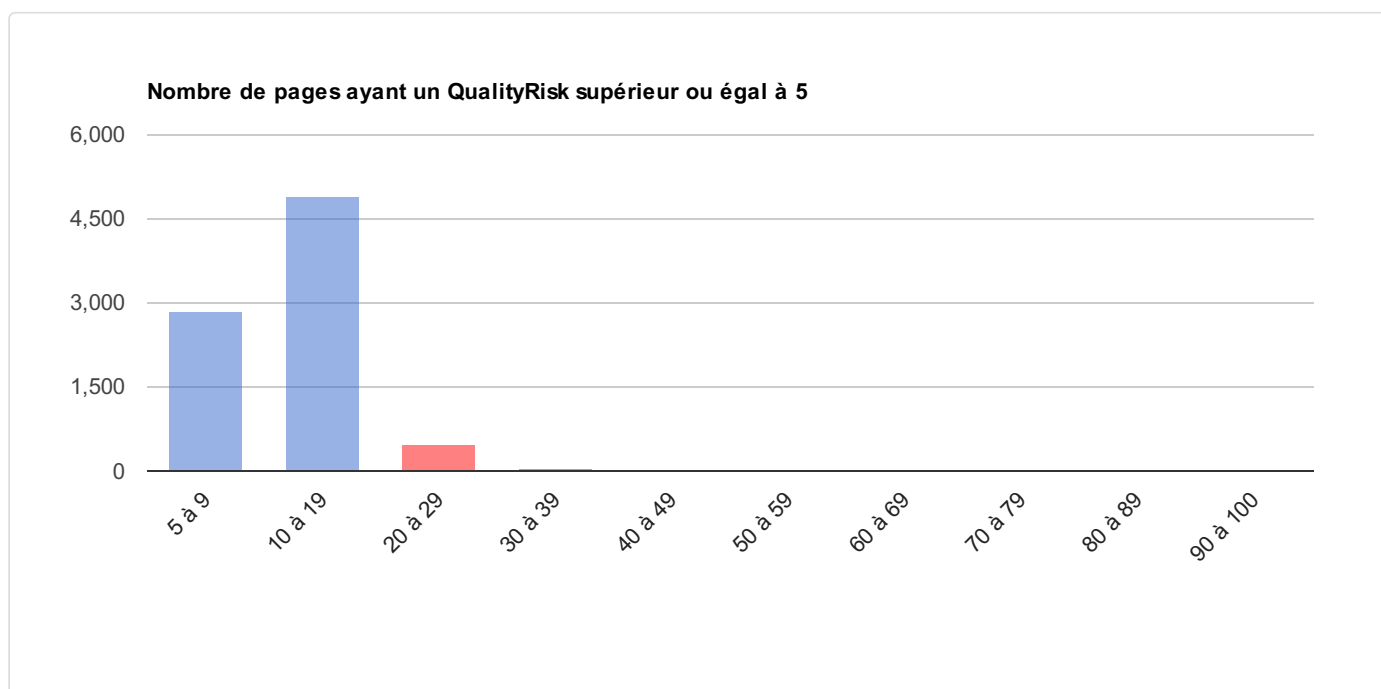
Pour réduire le QualityRisk d'une page, suivez tous les conseils déjà listés précédemment dans ce rapport d'audit. En voici le résumé :

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

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

Pour l'ensemble des pages indexables de ce site, le **QualityRisk moyen est de 11,0** (le pire possible étant 100).

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



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

	Erreurs	Avertissements	
Nb de pages avec un QualityRisk >= 20	499	n/a	
Nb de pages avec un QualityRisk entre 5 et 19	n/a	7 731	
Nb de pages indexables			8 886
Taux d'erreur	5,62 %	87 %	

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

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

Annexe

Dans l'annexe, nous listons uniquement les pages ayant un QualityRisk supérieur ou égal à 5. Le fichier annexe cdiscount_com-2018-02-10-qualite-des-pages.xlsx contient les données associées.

Annexe

Annexe complète pour les admins : cdiscount_com-2018-02-10-qualite-des-pages-admin.xlsx

11 Conclusion

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

Annexe

Le fichier annexe cdiscount_com-2018-02-10-conclusion.xlsx contient les données associées.

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

Annexe

Le fichier annexe cdiscount_com-2018-02-10-maillage-interne.xlsx contient les données associées.

Note : cette annexe est limitée à 1000000 lignes.

Vous trouverez ci-après une liste des problèmes techniques du référencement en cours sur votre site (étudié dans les conditions précisées dans l'introduction) à la date du 10 février 2018 et qu'il faut corriger au plus vite.

11.1 Problèmes de crawl et d'indexation

Il faut éviter de faire des liens internes vers des URL qui répondent en erreur. À la date de notre analyse, votre taux d'erreur était de 2,70 % (détails).

Il faut éviter de faire des liens internes vers des URL non indexables. À la date de notre analyse, votre taux d'erreur était de 8,67 % (détails).

11.2 Problèmes de temps de téléchargement

Nous vous conseillons de faire en sorte que chaque page soit téléchargée en 600 ms maximum. À la date de notre analyse, nous avons identifié que 83,72 % des URL HTML code 200 sont trop lentes, ce qui n'est pas optimal (détails).

11.3 Problèmes de profondeur des pages ou de maillage interne

Il faut que la majorité des URL soit à une profondeur inférieure ou égale à 4. À la date de notre analyse, nous n'avons trouvé aucune erreur de ce type (détails).

Il faut aussi que les pages reçoivent plusieurs liens entrants internes, avec des ancres distinctes. Notre analyse a montré que 38,48 % des pages n'ont pas des liens entrants internes assez efficaces.

11.4 Problèmes techniques concernant le contenu

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

Élément d'analyse	Taux d'erreur
balise title (détails)	88,01 %
balise meta description (détails)	12,05 %
contenu (détails)	11,21 %

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

Nb de pages ayant au moins une erreur technique	7 925
Nb de pages indexables	8 886
Taux d'erreur global	89,19 %

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

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

11.5 Problèmes sur les liens sortants externes

À la date de notre analyse, votre taux d'erreur était de 0,72 % (détails).

11.6 Les problèmes de masse noire

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

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

Étape 1 : Identification de vos contenus

Si votre site contient une boutique en ligne :

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

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

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

Si votre site contient un forum :

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

Concernant les autres pages de votre site :

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

Étape 2 : Calcul du nombre de pages à indexer

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

Étape 3 : Analyse

Comparez ce Nombre de pages à indexer au Nombre de pages indexables indiqué dans notre rapport (il était égal à 8 886 le jour de notre analyse) :

- si votre nombre de pages à indexer est égal à 8 886, alors vous n'avez aucun problème.
- si votre nombre de pages à indexer est inférieur à 8 886, alors vous avez sans doute des problèmes de masse noire et donc une sous-optimisation en termes de crawl et d'indexation. Google perd son temps (et ses ressources) à crawler et parfois à indexer des pages de votre site qui n'auraient pas dû l'être, au détriment du temps passé pour vos pages stratégiques.
- si votre nombre de pages à indexer est supérieur à 8 886, alors vous avez des problèmes d'indexation. Certains de vos contenus n'étant pas indexés, vous n'exploitez pas tout le potentiel de votre site dans Google. Cela peut être dû à de nombreux facteurs notamment des erreurs techniques, un mauvais maillage interne, une notoriété générale de votre site trop faible, des contenus dupliqués.

Pendant notre formation SEO, nous expliquons longuement comment éviter et corriger tous ces problèmes, au cours des 2 modules suivants :

- module A1 : [Les meilleures techniques d'optimisation du référencement naturel](#)
- module B : [Stratégies de liens, duplicate content, Google Panda et Penguin, refonte](#)

12 Pour aller plus loin

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

12.1 Quand refaire un audit technique SEO ?

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

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

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

12.2 Formations

Organisme de formation déclaré depuis 2006, nous avons formé plus de 4000 entreprises à de nombreuses composantes du webmarketing. Voici notre catalogue de formations :

- Référencement
 - Les meilleures techniques d'optimisation du référencement naturel (module A1, 2 jours)
 - Booster son référencement par l'éditorial (module A2, 1 jour)
 - Stratégies de liens, duplicate content, Google Panda et Penguin, refonte (module B, 1 jour)
 - Optimiser le référencement de son site PrestaShop (module C, 1 jour)
- Google Analytics
 - Découvrir Google Analytics (module N, 1 jour)
 - Google Analytics pour le référencement naturel (module L, 1 jour)
 - Google Analytics pour profil marketing (module G, 2 jours)
 - Google Analytics pour profil technique (module D, 2 jours)
- Google AdWords
 - Optimiser ses campagnes AdWords (module K, 1 jour)
- Réseaux sociaux
 - E-reputation, Facebook et Twitter (module E, 1 jour)
 - Google+, données structurées (module P, 1 jour)

12.3 Accompagnement personnalisé

Si vous avez besoin d'analyses spécifiques basées sur nos outils que ce soit dans le cadre d'une refonte de site web, d'analyses concurrentielles ou autres n'hésitez pas à nous contacter par email (contact@ranking-metrics.fr) ou par téléphone au 01 64 61 03 22.

Vous pouvez également :

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

13.1 Fonctionnement de nos outils

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

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

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

13.2 Signification des codes HTTP

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

Les principales familles de codes :

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

Les codes les plus courants :

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

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

13.3 Signification des types MIME

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

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

13.4 Signification des termes SEO

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

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

- **Redirection (301 ou 302)** : on parle de redirection d'une URL A vers une URL B quand, en cherchant à accéder à A, le serveur nous renvoie vers B. En référencement, il faut privilégier la redirection permanente (code HTTP 301) dans la majorité des cas. Il existe de nombreux moyens de gérer les redirections. En savoir plus : [aide Google](#), [dossiers WebRankInfo](#).
- **Sitemap** : un fichier sitemap liste des URL que l'on souhaite faire crawler par les moteurs de recherche (en général pour qu'ils indexent les URL concernées). Ce n'est pas obligatoire mais ça peut aider à accélérer le crawl ou dans le suivi du référencement. En savoir plus : [aide Google](#), [dossiers WebRankInfo](#).
- **Lien nofollow** : si une page A fait un lien nofollow vers une page B, alors les moteurs de recherche feront comme si ce lien n'existe pas. En d'autres termes, la page B ne bénéficie d'aucun avantage (ou inconvénient) de recevoir un lien depuis A. Google impose que les liens commerciaux soient en nofollow, par le biais d'un attribut rel=nofollow sur la balise A (liens vendus, liens affiliés). En savoir plus : [aide Google](#), [dossiers WebRankInfo](#).

Table des matières

Audit Technique SEO : RM Tech	1
1 Introduction	2
1.1 Objectif de ce rapport	2
1.2 Intérêt de cette analyse	2
1.3 Rapport d'audit et annexes contenant toutes les données	2
Annexes	3
2 Synthèse des résultats	4
2.1 Score RM Tech	4
2.2 QualityRisk	4
2.3 Taux d'erreur	4
2.4 Performance	5
3 Crawl et indexabilité	6
3.1 Bilan des URL bloquées par le(s) fichier(s) robots.txt	6
Annexe	6
Annexe	6
3.2 Nombre total d'URL crawlées	6
3.3 État des URL crawlées	7
Annexe	8
Annexe	8
Annexe	8
3.4 Types MIME	8
Annexe	8
3.5 Pages HTML crawlées sans erreur	8
Annexe	8
3.6 Pages HTML indexables par les moteurs	8
Annexe	9
Annexe	9
Annexe	9
4 Temps de téléchargement	10
Annexe	11
5 Profondeur des pages et maillage interne	12
5.1 Analyse de la profondeur des pages	12
5.2 Analyse du maillage interne	13
5.3 Conclusion sur la profondeur et le maillage interne	15
Annexe	15
6 Balise title	16
6.1 Taille de la balise Title	16
Annexe	16
6.2 Doublons dans la balise title	16
Annexe	17
6.3 Répétition de mots dans la balise Title	17
Annexe	17
6.4 Votre marque dans la balise Title	17
Annexe	17
6.5 Conclusion sur la balise Title	17
7 Balise meta description	19
7.1 Taille de la balise meta description	19
Annexe	19
7.2 Doublons dans la balise meta description	19
Annexe	19
7.3 Conclusion sur la balise meta description	19
8 Contenu des pages	21
8.1 Balises H1-H6	21
Annexe	22
8.2 Doublons dans la balise H1	22
Annexe	22
8.3 Taille du texte (nb de mots)	22
Annexe	23
8.4 Mise en exergue (strong ou bold)	23

Annexe	24
8.5 Analyse du nombre de liens par page	24
Annexe	24
8.6 Analyse des contenus dupliqués internes	24
Annexe	25
8.7 Identifiants de session dans les URL	25
Annexe	25
8.8 Paramètres de tracking dans les URL	25
Annexe	26
8.9 Conclusion sur le contenu des pages	26
9 Liens sortants externes	27
9.1 Nombre total de liens sortants externes	27
Annexe	27
9.2 Erreurs sur les liens sortants externes	27
10 Analyse de la qualité	28
Annexe	29
Annexe	29
11 Conclusion	30
Annexe	30
Annexe	30
11.1 Problèmes de crawl et d'indexation	30
11.2 Problèmes de temps de téléchargement	30
11.3 Problèmes de profondeur des pages ou de maillage interne	30
11.4 Problèmes techniques concernant le contenu	30
11.5 Problèmes sur les liens sortants externes	31
11.6 Les problèmes de masse noire	31
Étape 1 : Identification de vos contenus	31
Étape 2 : Calcul du nombre de pages à indexer	32
Étape 3 : Analyse	32
12 Pour aller plus loin	33
12.1 Quand refaire un audit technique SEO ?	33
12.2 Formations	33
12.3 Accompagnement personnalisé	33
13 Explications complémentaires	34
13.1 Fonctionnement de nos outils	34
13.2 Signification des codes HTTP	34
13.3 Signification des types MIME	35
13.4 Signification des termes SEO	35
Table des matières	37