

Audit SEO

Toutes les clés pour analyser et optimiser votre site web

Jean-Marc COURTIADÉ

En téléchargement

 webographie



Chapitre 1 : Introduction

A. Préambule.....	9
1. Il était une fois.....	9
2. Une question de vision	10
3. À qui s'adresse ce livre ?.....	11
4. L'audit SEO dans l'écosystème de services web	11
5. Champs d'intervention	12
B. Le framework d'audit	13
1. Définition du framework d'audit SEO.....	13
2. Des audits pour quels outputs	13
3. Postulats : une question de principe.....	14
a. Postulat n° 1 : UNE méthode, pas LA méthode.....	14
b. Postulat n° 2 : s'adapter est la clé.....	14
c. Postulat 3 : un parti pris nécessaire.....	15
4. Intérêt de ce framework	15
C. Différents types d'audits SEO	16
1. Trois domaines d'analyse	16
2. Outils et méthodes	18
a. Excel au cœur du système	18
b. Outils SEO	19
c. Outils bureautiques	21
d. Outils organisationnels	22
e. Autres outils techniques.....	22
D. Le rapport d'audit.....	22
1. Introduction.....	22
2. Pas un rapport d'audit mais des rapports d'audits.....	23
a. Types de rapports	23
b. Fichiers d'actions correctives	24
3. Recommandations générales pour la rédaction des rapports d'audit	24
E. Préparation d'un audit	25
1. Coût d'un audit.....	25
2. Autres logiciels	27
3. Délai	27
4. Freins à l'audit.....	28
5. Résumé du chapitre 1.....	28

Chapitre 2 : Préparation de l'audit

A. Les enjeux de l'audit technique.....	31
B. Avant de commencer	32
1. Principes	32
2. Plan d'action	33
3. À quel moment rédiger le rapport ?	34
C. Préparer l'audit technique.....	35
1. Recueil des informations préalables	35
2. Outils	37
D. Crawler le site.....	40
1. Pourquoi faire un crawl en début de projet ?.....	40
2. Éléments préalables au crawl	41
3. Préparer le crawl avec le client	43
4. Configurer Screaming Frog	43
a. Configuration du système	44
b. Choix du User-Agent	45
c. Limiter la vitesse de crawl	46
d. Robots.txt.....	47
e. Authentification.....	47
f. Analyser le contenu dupliqué.....	48
g. Utilisation de CDN.....	49
h. Configuration du crawler	50
i. Vérifier si les codes de tracking web analytics sont en place.....	61
j. API Access	62
k. Include	63
l. Exclude	63
5. Méthodologie de crawl	64
a. Utiliser une machine dédiée	64
b. Avertir le client	65
c. Faire des tests.....	65
d. Crawler les sites des concurrents.....	65
e. Que faire des crawls incomplets ?.....	65
f. Sauvegarder votre analyse.....	66
g. Crawler comme le Googlebot.....	66
E. État des lieux organique.....	68
1. Avant de commencer	68
2. Informations fournies par Google Analytics.....	69
a. Paramétrage de Google Analytics (Universal Analytics)	69
b. Évolution du trafic organique et comparaison avec le trafic global	70
c. Focus sur certains KPI.....	70
d. Évolution du trafic organique d'une année à l'autre	70

e. Analyse géographique et langues.....	71
3. Autorité du nom de domaine	71
4. Analyser l'état d'indexation du site	72
a. Vue d'ensemble de l'indexation.....	73
b. Analyse des données de Google Search console.....	73
5. Informations sur l'environnement.....	74

Chapitre 3 : Réaliser l'audit technique

A. Introduction	77
B. Analyses techniques de base	77
C. Extraction des données de Screaming Frog	86
1. Brève analyse des données dans Screaming Frog	86
2. Export des URLs	88
3. Mise en forme des données avec Excel	89
D. Analyse du crawl et de l'indexation	90
1. Technologies du site.....	90
2. Faire le bilan du crawl.....	91
3. Analyse du fichier robots.txt	93
4. Statut des éléments crawlés.....	96
5. Trouver les pages non indexables (deuxième méthode).....	105
6. Analyser les sitemaps	114
7. Liens de pages avec attributs nofollow, ugc ou sponsored	119
8. Contenu dupliqué	120
9. Analyser les attributs canonical	125
10. Analyser la profondeur des pages	131
E. Analyse avec pages actives et inactives	137
1. Introduction.....	137
2. Télécharger les données de Google Analytics.....	138
3. Pages actives vs Pages inactives vs Pages fantômes.....	141
F. Pages inutiles à l'indexation	146
1. Définition	146
2. Méthode d'analyse	147
3. Paramétrer Excel pour analyser les pages inutiles	148
4. Principaux points d'analyse	149
5. Vue d'ensemble des pages inutiles à l'indexation	150
G. Analyse : performance du site	151
1. La vitesse : un facteur déterminant	151
2. Concepts clés.....	152
3. Analyse des performances avec Google Lighthouse.....	154
4. Analyse des performances avec Google PageSpeed Insights	157
5. Analyse des performances avec Dareboost	160

6. Analyses avec des logiciels tiers	162
7. Analyse avec Screaming Frog	164
H. Analyse : éléments on-page	165
1. Introduction.....	165
2. Un point sur le rapport d'audit	165
3. Balise title	166
4. Balise meta description	167
5. Balise meta keywords.....	167
6. Balise H1	168
7. Balises H2	169
8. Les URLs.....	170
9. Les images	171
10. Le nom de domaine	172
I. Analyse du contenu.....	173
1. Images.....	173
2. Pages maigres.....	175
J. Analyse multilingue.....	178
K. Autres domaines d'analyse	185

Chapitre 4 : Réaliser un audit sémantique et concurrentiel

A. Avant de commencer	189
1. Introduction.....	189
2. Périmètre d'intervention.....	190
3. Chaque chose à sa place et une place pour chaque chose.....	191
B. Prérequis	191
1. Informations préalables.....	191
2. Trouver les principaux concurrents.....	193
3. Crawler les sites du client et des concurrents	194
C. État des lieux organique.....	194
1. Fonctionnement de SEMrush	195
2. Vue d'ensemble	196
a. Principaux indicateurs.....	196
b. Analyse de l'Authority Score (AS)	197
c. Autres données générales.....	198
d. Données générales de recherche organiques	199
e. Données générales de concurrence.....	200
f. Données de Netlinking	201

D. Analyse de la concurrence	202
1. Données générales SEMrush	202
a. Courbes d'évolution	203
b. Sources du trafic	203
c. Parcours du trafic	204
d. Trafic par pays	204
2. Section Market Explorer	205
a. Matrice de positionnement	205
b. Dynamique domaine contre marché	206
E. Analyse des mots-clés	207
1. Catégorisation des mots-clés	207
a. Télécharger les données	207
b. Traiter les données	208
c. Création des tableaux croisés dynamiques	209
d. Analyse des concurrents	210
2. Intentions de recherche principales	211
a. Keyword Magic Tool	211
b. Organic Traffic Insights	212
3. Chevauchement de mots-clés	215
4. Keyword Gap Analysis	217
a. Avec SEMrush	217
b. Avec Excel	218
F. Pages actives et inactives	223
1. Introduction	223
2. Méthodologie	223
a. Crawl du site et export des données	223
b. Export des données de SEMrush	224
c. Analyse des pages d'entrée organique	224

Chapitre 5 : Réaliser un audit de netlinking

A. Introduction	227
1. La force du lien	227
a. Un peu d'histoire	227
b. Le PageRank	227
2. Enjeux du netlinking	228
a. Encore libre	228
b. TrustRank : qualité vs quantité	228
B. Prérequis	229
1. Informations préalables	229
2. Outils	230

C. Analyses du netlinking	231
1. Préparation de l'analyse	231
a. Fonctionnement de Majestic	231
b. Méthodologie	231
c. Quel nom de domaine auditer ?	231
2. Analyses générales	232
a. Introduction	232
b. Citation Flow (CF) et Trust Flow (TF)	232
c. Analyse des profils de liens et de domaines référents	234
d. Nombre de backlinks, domaines référents, IP et Referring Subnets	235
e. Analyse du Link Graph	237
f. Autres informations récapitulatives	238
3. Analyse des Trust Flow par sujet	239
a. Fonctionnement	239
b. Analyse du site	239
c. Analyse des concurrents	242
4. Analyser les concurrents	242
a. Méthodologie	242
b. Outil Comparator	243
c. Outil Backlinks History	244
d. Flow Metrics History	244
e. Comparer les sujets	245
f. Clique Hunter	246
5. Analyse des ancres de liens	247
a. Nuage de mots-clés	247
b. Extraction des données et analyses	247
6. Remarques supplémentaires	248
a. Conseil n° 1	248
b. Conseil n° 2	248
c. Conseil n° 3	248
d. Conseil n° 4	249
D. Conclusion	249
 Remerciements	 251
Index	253

A. Introduction

Nous entrons dans le cœur de notre analyse, en résumé, jusqu'à présent vous avez :

- fait un point complet du périmètre du projet,
- décelé les points de blocage au crawl du site,
- paramétré Screaming Frog et réalisé un crawl du site, en plusieurs tentatives si besoin,
- analysé l'environnement organique du site,
- analysé le statut de son indexation dans Google Search Console.

Nous allons maintenant passer à l'analyse purement technique du site. Cette étude sera subdivisée en plusieurs parties que nous allons traiter.

Sachez cependant que ce plan peut être adapté aux besoins du client, notamment en ce qui concerne les sites e-commerce ou multilingues.

Pour rappel, en réalisant ces analyses, vous rédigerez également **le rapport d'audit** et de **cahier de préconisations techniques**. Commençons par faire un recueil des informations essentielles du site.

B. Analyses techniques de base

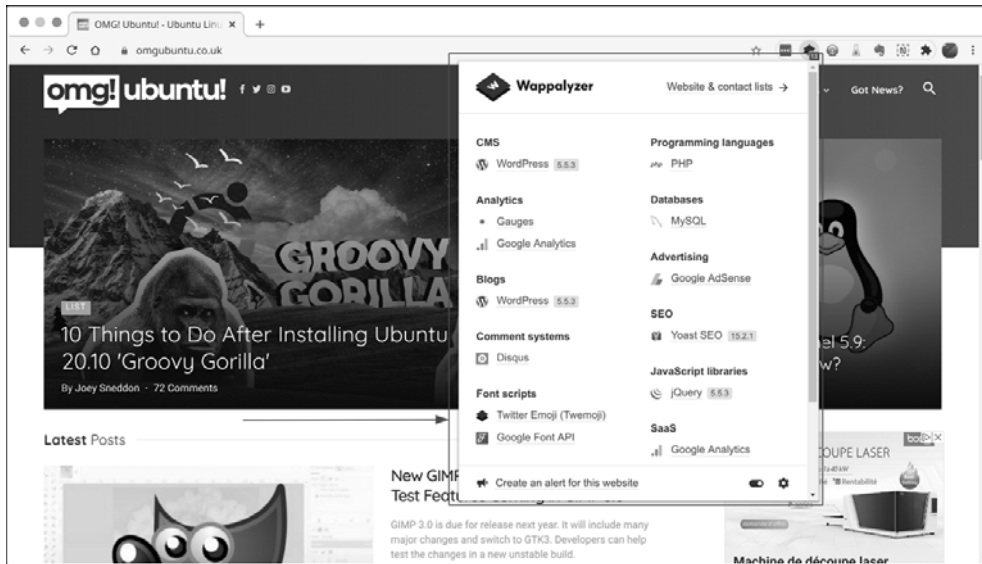
Après avoir passé en revue l'environnement organique du site, il s'agit maintenant de faire une analyse générale de ses **aspects techniques**. Bien sûr, la taille du site aura un impact sur vos recherches. Un site de petite taille (avec quelques centaines de pages) n'a pas les mêmes problématiques que le site d'un groupe international. Ce sera à vous d'adapter vos recherches dans ce sens.

Prenons un exemple. Les choses sont relativement plus simples pour le site d'une petite entreprise : peu d'URL, site monolingue, sous WordPress voire sous Joomla. Cela est différent d'un site dont le trafic a chuté après une migration et la fermeture de plusieurs sites satellites. Et cela est encore différent de l'analyse du site d'un groupe international, décliné en plusieurs langues (comme le chinois et le russe...) ou qui repose sur une base technique plus complexe (Drupal par exemple).

Un petit aparté : malgré la complexité de ces derniers types de sites, ce sont souvent ceux sur lesquels il est possible de mettre en place des actions correctives de grande ampleur, leurs propriétaires disposant de fonds conséquents.

Voici quelques pistes et outils qui vous donneront un aperçu des technologies employées sur un site.

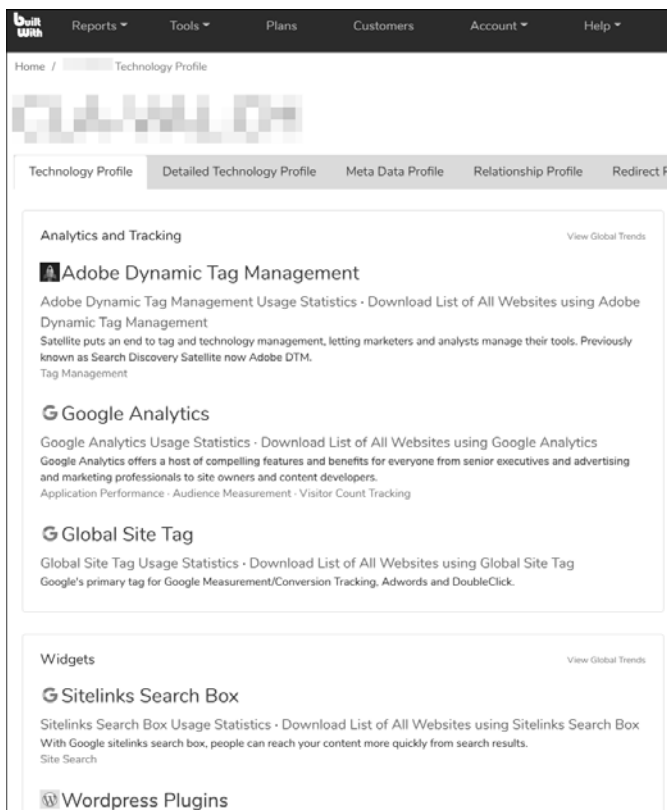
Wappalyzer : plug-in disponible sur Chrome et Firefox, il donne un aperçu de la base technique du site, notamment type de CMS, langage de programmation, type de serveur, solution analytics.



Un autre plug-in nommé **WhatRuns** est également disponible sur Chrome et Firefox (rendez-vous sur les sites des extensions des navigateurs).

- **BuiltWith.com** : ce site est une véritable mine d'or pour les apprentis chercheurs. Il présente les différentes technologies en place sur le site analysé et maintient l'historique des modifications effectuées : changements de CMS au cours des années, d'adresse IP, de code de suivi Google Analytics, etc. Il vous propose également les sites faisant une redirection depuis un nom de domaine autre. Il se peut d'ailleurs que le mandataire de l'audit ne les connaisse même pas.

Site web : <https://builtwith.com/>



- **PublicWWW** : réservé à des recherches plus approfondies, ce site vous permet de réaliser des investigations sur des éléments non disponibles via les moteurs de recherche, mais disponibles librement dans leur code source :
- Codes de suivi Google Analytics (pour vérifier si un code UA a pu être utilisé sur différents sites au cours de son existence).
- Quels sites utilisent un thème WordPress particulier.
- Quels sites font mention d'une marque (ou de votre nom, si vous souhaitez analyser votre e-réputation, mais c'est un autre sujet).
- Ou encore la liste des designers ou développeurs web qui détestent Internet Explorer : <https://publicwww.com/js/%22stupid+ie%22+filetype%3Ajs/>

Site web : <https://publicwww.com/>

The screenshot shows the PublicWWW Source Code Search Engine interface. At the top, there's a navigation bar with links for 'PublicWWW', 'Examples', 'Clusters', and 'Pricing', along with 'Sign Up' and 'Log In' buttons. The main heading is 'Source Code Search Engine' with a subtitle 'Find any alphanumeric snippet, signature or keyword in the web pages HTML, JS and CSS code.' Below this is a search bar with the placeholder text 'Search for code snippet, signature or keyword...' and a 'Search' button. To the right of the search bar, it displays '536 159 029 web pages' and 'November 10, 2020'. On the left side, there's a section titled 'Ultimate solution for digital marketing and affiliate marketing research...' followed by a list of use cases. On the right side, there's a section titled 'Usage Examples' showing various code snippets from popular frameworks and services like AngularJS, Bootstrap, Add This, reCAPTCHA, Akamai, Algolia, HubSpot, and comScore.

Source Code Search Engine
Find any alphanumeric snippet, signature or keyword in the web pages HTML, JS and CSS code.

Search for code snippet, signature or keyword... **Search**

query syntax: RegEx, ccTLDs, etc.

Usage Examples

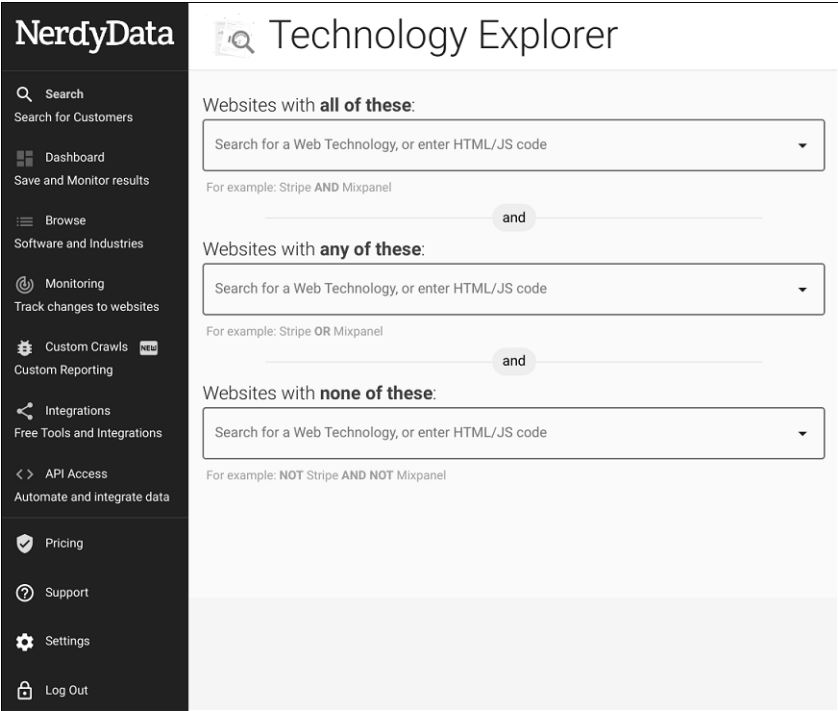
- AngularJS: "angular.min.js"
- Bootstrap: "bootstrap.min.js"
- Add This: "addthis_widget.js"
- reCAPTCHA: "recaptcha/api.js"
- Akamai: "X-Akamai-Transformed"
- Algolia: "AlgoliaSearch"
- HubSpot: "hubspot"
- comScore: "Begin comScore Tag"

Ultimate solution for digital marketing and affiliate marketing research, PublicWWW allow you to perform searches this way, something that is not possible with other regular search engines:

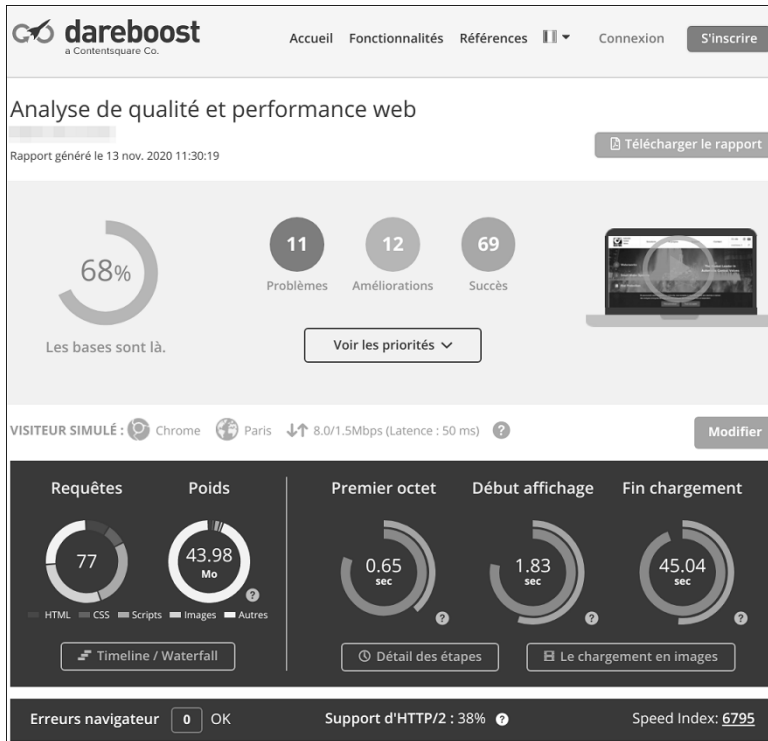
- Any HTML, JavaScript, CSS and plain text in web page source code
- References to StackOverflow questions in HTML, CSS and JS files
- Web designers and developers who hate IE
- Sites with the same analytics id: "UA-19778070-"
- Sites using the following version of nginx: "Server: nginx/1.4.7"
- Advertising networks users: "adserver.adtech.de"
- Sites using same adsense account: "pub-9533414948433288"
- Wordpress with theme: "/wp-content/themes/twenty sixteen/"
- Find related websites through the unique HTML codes they share, i.e. widgets & publisher IDs
- Identify sites using certain images or badges
- Find out who else is using your theme
- Identify sites that mention you
- References to use a library or a platform
- Find code examples on the internet
- Figure out who is using what JS widgets on their sites.

- **NerdyData** : à l'image de PublicWWW, NerdyData comme son nom l'indique, vous permet également de retrouver toutes sortes d'informations sur un site ou des technologies employées. Une extension pour Chrome est disponible, mais son intérêt est limité en comparaison avec Wappalyzer ou WhatRuns par exemple. En revanche, l'intérêt de cette solution repose sur son site web qui vous permet de faire des recherches approfondies sur des codes HTML (un code HTML précis se retrouve-t-il sur plusieurs sites de votre client ?) ou sur des secteurs entiers.

Site : <https://www.nerdydata.com/>

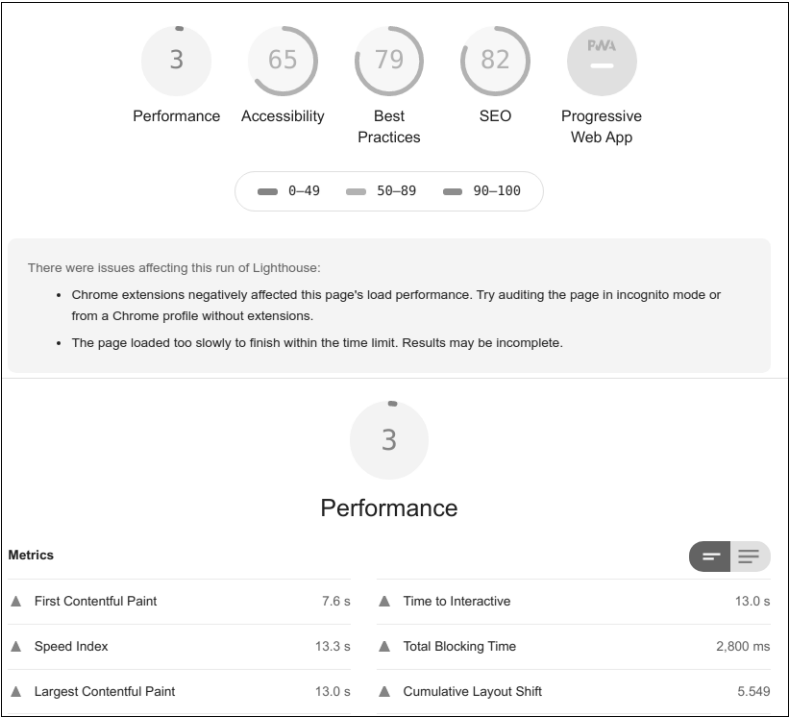


- **Dareboost** : créé initialement par une société française, Dareboost vous permet de réaliser une analyse détaillée de très nombreux points techniques et vous propose des solutions clé en main concernant des points de SEO, sécurité, UX, etc. Avec son système d'évaluation, vous pouvez vérifier quelles sont les performances d'un site. Dareboost offre d'autres services, notamment de monitoring, mais ils n'entrent pas dans le cadre de cet audit.



- **Google Lighthouse** : ce logiciel est très connu et aussi très visible, puisqu'il est intégré à Google Chrome (dans les outils de développeurs). Il est disponible comme plug-in pour Chrome et Firefox et accessible via le site **Google PageSpeed Insights dit PSI** : <https://developers.google.com/speed/pagespeed/insights/>

Il peut être aussi utilisé comme module sur un serveur Node.js. Notons qu'il s'intègre à Screaming Frog. Il fournit des indications sur les performances d'un site. Elles ne sont pas à prendre au pied de la lettre, mais peuvent être intéressantes pour un aperçu « high level ». Les données peuvent être exportées au format JSON pour être exploitées plus tard grâce au Lighthouse Report Viewer : <https://googlechrome.github.io/lighthouse/viewer/>



Pour un rendu plus concret des résultats, choisissez les options suivantes :

- **Appareil (Device)** : « **Mobile** », Google Lighthouse va ainsi simuler une connexion plus lente (la majeure partie des connexions au web se faisant via appareils mobiles).
- **Vider le cache (Clear Storage)** pour débiter les analyses avec un cache navigateur vide, comme lors d'une première connexion au site.
- **Simulate throttling** : pour simuler le changement de la page testée sur un smartphone moyen de gamme (Motorola 4G).

! Ces informations sont disponibles depuis la version 80 de Google Chrome, d'autres options étaient auparavant disponibles, mais peu fiables.

Attention : le problème avec Google Lighthouse/PSI est que l'on ne connaît pas l'emplacement du serveur de test ! Et vous obtiendrez des résultats bien différents selon que vous testiez le site web à Sydney, Saint-Brieuc ou à São Paulo.

La section dédiée à l'analyse des performances de sites web vous donnera de plus amples renseignements.