



L'analyse de logs pour tous.

SEOLYZER

SEO
Good Vibes

Hey!



Olivier PAPON, dans le SEO depuis 2005.

Créateur de la solution **Seolyzer**



DECATHLON

ELLE

AIRFRANCE

Calvin Klein

JUST EAT



@Seolyzer_io

SEOLYZER



Quando quelqu'un
entend "logs".

D'où viennent les logs ?



Code source
= un hit



1 Police = un hit

1 Fichier CSS = un hit

1 Image = un hit

1 Fichier Javascript = un hit

Un hit = une demande envoyée au serveur



Astuce : dans votre navigateur, visualisez toutes les ressources chargées par une page avec F12 puis onglet "Network"

1 Fichier Javascript

```

1590479784 200 66.249.79.58 seolyzer.io GET /robots.txt HTTP/1.1 - Mozilla/5.0 (compatible; Googlebot/2.1;
+http://www.google.com/bot.html) 2009 244 443 seolyzer.io GET /robots.txt HTTP/1.1
1590479784 200 66.249.79.158 seolyzer.io GET /fr/ HTTP/1.1 - Mozilla/5.0 (Linux; Android 6.0.1; Nexus 5X Bui
ld/MMB29P) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.92 Mobile Safari/537.36 (compatible; Googlebot/2.1;
+http://www.google.com/bot.html) 20458 11943 443 seolyzer.io GET /fr/ HTTP/1.1
1590480263 200 66.249.79.159 seolyzer.io GET /fr/ HTTP/1.1 - Mozilla/5.0 (Linux; Android 6.0.1; Nexus 5X Bui
ld/MMB29P) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.92 Mobile Safari/537.36 (compatible; Googlebot/2.1;
+http://www.google.com/bot.html) 16737 11943 443 seolyzer.io GET /fr/ HTTP/1.1
1590480869 200 66.249.79.157 seolyzer.io GET /fr/ HTTP/1.1 - Mozilla/5.0 (Linux; Android 6.0.1; Nexus 5X Bui
ld/MMB29P) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.92 Mobile Safari/537.36 (compatible; Googlebot/2.1;
+http://www.google.com/bot.html) 11943 443 seolyzer.io GET /fr/ HTTP/1.1
1590483605 200 66.249.79.158 seolyzer.io GET /fr/ HTTP/1.1 - Mozilla/5.0 (compatible; Googlebot/2.1; +http://
www.google.com/bot.html) 19363 443 seolyzer.io GET /fr/ HTTP/1.1
1590496989 200 66.249.79.157 seolyzer.io GET /fr/ HTTP/1.1 - Mozilla/5.0 (Linux; Android 6.0.1; Nexus 5X Bui
ld/MMB29P) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.92 Mobile Safari/537.36 (compatible; Googlebot/2.1; +h
t://www.google.com/bot.html) 11943 443 seolyzer.io GET /fr/ HTTP/1.1
1590497792 200 66.249.79.158 seolyzer.io GET /fr/ HTTP/1.1 - Mozilla/5.0 (compatible; Googlebot/2.1; +http://ww
w.google.com/bot.html) 15366 10947 443 seolyzer.io GET /fr/ HTTP/1.1
1590500821 200 66.249.79.158 seolyzer.io GET /fr/live-google-mobile-first-index/ HTTP/1.1 - Mozilla/5.0 (Lin
ux; Android 6.0.1; Nexus 5X Build/MMB29P) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.92 Mobile Safari/537
.36 (compatible; Googlebot/2.1; +http://www.google.com/bot.html) 15366 10947 443 seolyzer.io GET /fr/live-go
ogle-mobile-first-index/ HTTP/1.1
1590505322 304 66.249.79.158 seolyzer.io GET /wp-content/uploads/2014/09/seo-log-analysis-real-time-300x181.png
HTTP/1.1 - Googlebot-Image/4.1 443 seolyzer.io GET /wp-content/uploads/2014/09/seo-log-analysis-r
eal-time-300x181.png HTTP/1.1

```



 **Astuce :** si une adresse IP commence par 66.249, elle appartient à priori à Google.

Que contiennent les logs ?



Chaque ligne contient

- timestamp
- client-ip.
- url
- status code
- user-agent
- referrer
- weight (bytes)
- response time (optionnel)
- protocole ou port (optionnel)

Où récupérer les logs ?



Cela dépend de votre infrastructure / hébergement

CDN

ou Load Balancer



Live | Fichiers | API

Mutualisé



Live | Fichiers

Dédié



Live | Fichiers

Pourquoi utiliser les logs ?

“



Log files are so underrated, so much good information in them.

John Mueller. Webmaster Trends Analyst @Google.

SEOLYZER

SEO
Good Vibes

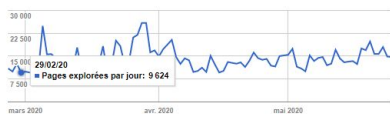
Google Search Console



Statistiques sur l'exploration

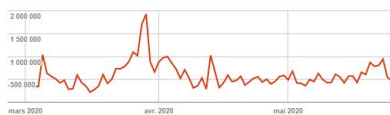
Activité Googlebot durant les 90 derniers jours

Pages explorées par jour



Élevé	Moyenne	Faible
25 919	14 038	7 557

Kilo-octets téléchargés par jour



Élevé	Moyenne	Faible
1 928 112	583 749	21 158

Temps de téléchargement d'une page (en millisecondes)



Élevé	Moyenne	Faible
587	460	270



Anciens outils et rapports

☒ Ciblage international

☒ Statistiques sur l'exploration

A chacun sa **vision**...



Propriétaire du site

Vision Business : catalogue...



Le SEO et son crawler

Vision OnPage technique



Googlebot

Vision historique et globale



**Crawler n'est pas
la vérité !**





Crawler

Simulation

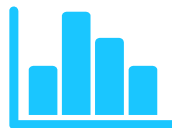


Instantané

A partir de la home page
Dépend de la configuration du crawler

Analyse de logs

Réalité



Historique

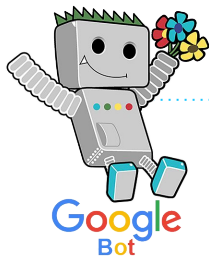


Live [monitoring](#)

Comprendre Googlebot



Site internet



! Coût



130 000 milliards de pages

Electricité, serveurs, bande
passante...

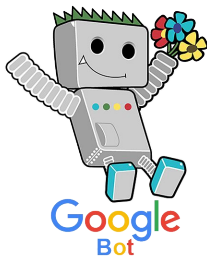
"Crawl budget"



Où est Googlebot ?



On résume !



Googlebot ♥ les sites où :

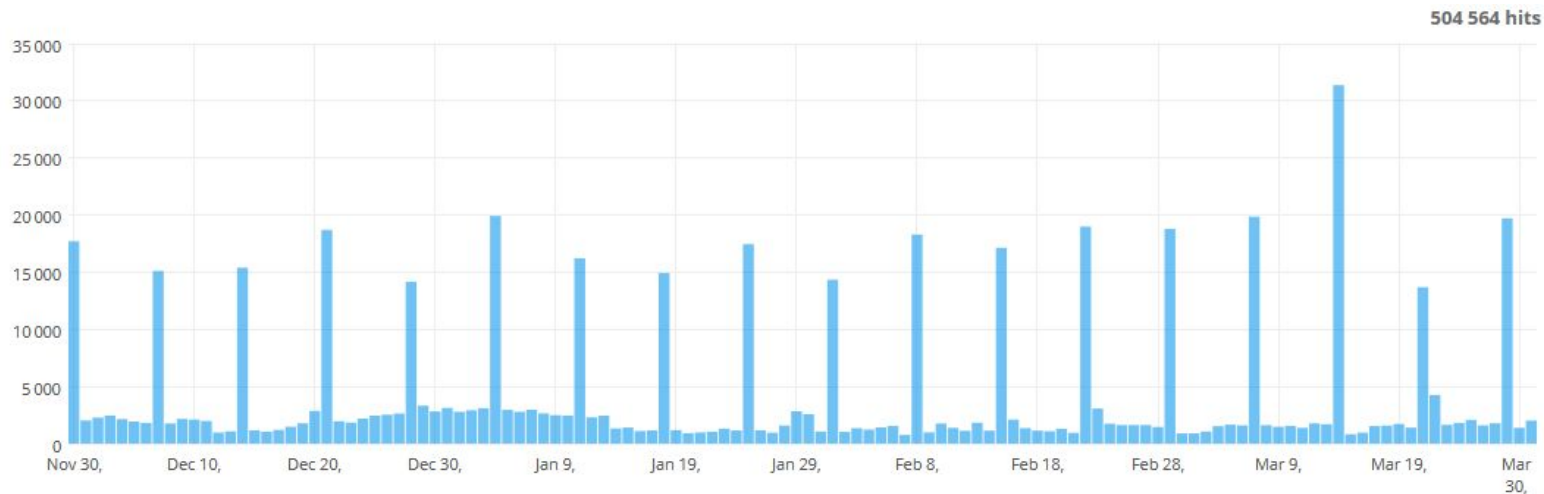
Il ne perd pas de temps

Les choses importantes sont
proches de l'entrée

Les rayons sont bien rangés

Les rayons ne sont pas vides

Googlebot, c'est toi ?



Googlebot, c'est toi ?



Astuce : vérifiez les Reverse DNS des IP pour identifier GB

<https://support.google.com/webmasters/answer/80553?hl=fr>

SEOLYZER



Comment manipuler les logs?



Vous oubliez !

Il y a des **outils free[mium]** pour ça.



Open Source

Watussi



Code modifiable



Freemium

Screeningfrog
1000 lignes

SEOLYZER
10000 pages

Fonctionnalités avancées



SEOLYZER



Pour une **analyse** pertinente du comportement de **Googlebot**



Catégorisez vos pages

Langue

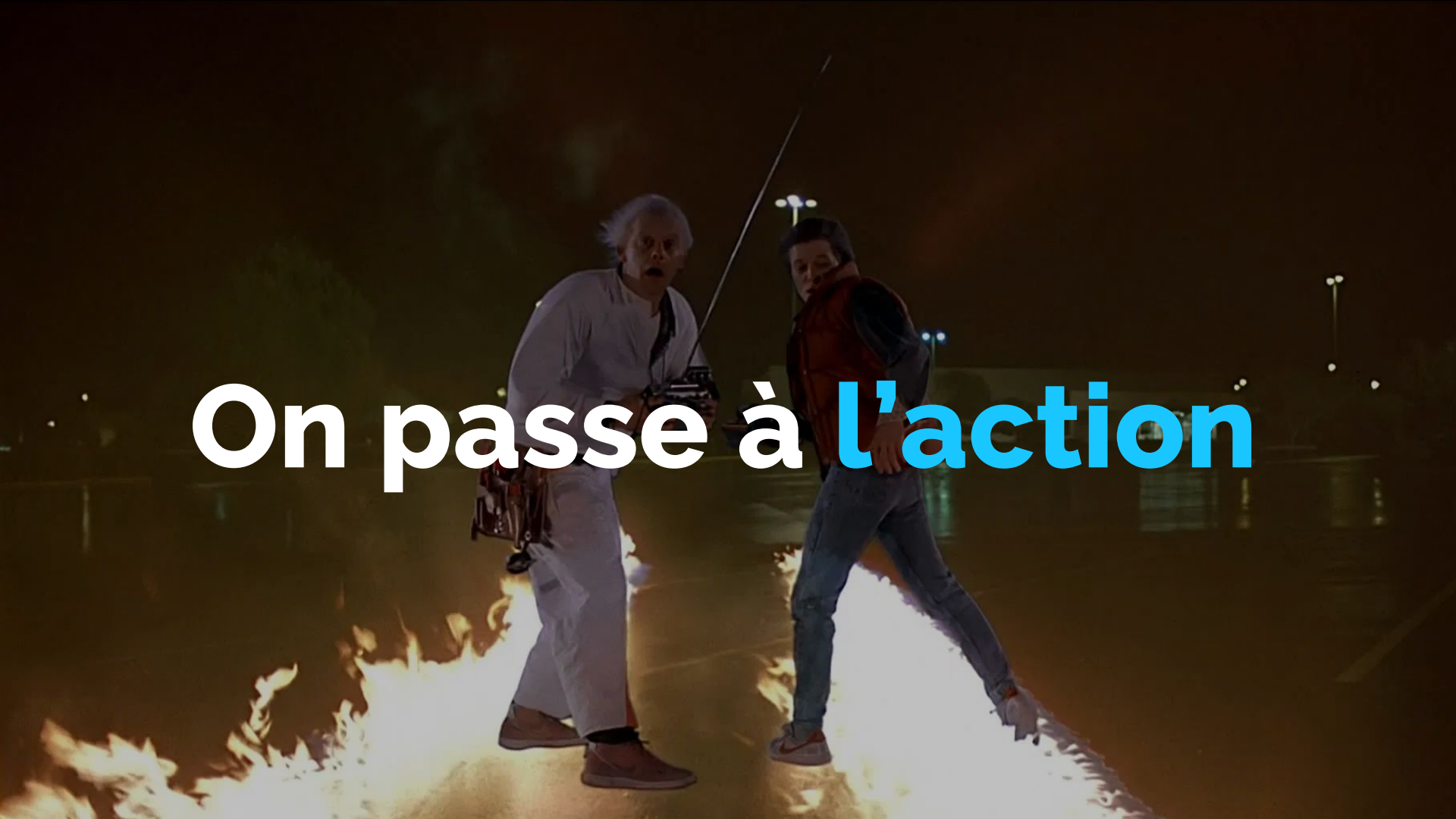
Produit

Editorial

Template

...

On peut catégoriser par l'**URL** ou par **élément scrapé** (Regex, Xpath, CSS)

A still from the movie 'Back to the Future' showing Doc Brown and Marty McFly running through a street at night. The ground is covered in flames, and Doc Brown is holding a briefcase. The text 'On passe à l'action' is overlaid in the center.

On passe à l'action

A man with curly hair is smiling while sitting in the driver's seat of a car. He is holding a small squirrel in his hands, and the squirrel is screaming with its mouth wide open. The background shows the interior of the car and a window looking out onto a dark, rainy street.

On [résumé|répète]

Avec l'analyse de logs vous...

Monitoriez

1

- Volume de crawl
- Codes HTTP
- Performance par page
- Performance moyenne
- Poids des pages
- Nouvelles pages (temps de prise en compte)
- Pages les plus crawlées
- Visites SEO
- Robots.txt...

Contexte

- Routine quotidienne
- Refonte de site
- Migration
- Lot de redirection
- Upgrade de CMS...

Améliorez

2

- Erreurs 4XX (Liens perdus)
- Erreurs 5XX
- Pages lentes
- Pièges à robots
 - Recherche interne crawlable
 - Pagination infinie
 - Facettes
 - Scroll infini
 - Paramètres indésirables....

Pensez “business”

3

Googlebot a t-il la [bonne vision](#) de mon site ?

Et pour aller **plus loin...**

4

Croisement de données

Crawl de Googlebot VS **Crawler SEO** [Pages orphelines]

Volume de crawl Googlebot VS **Profondeur des pages**

Volume de crawl Googlebot VS **Chiffre d'affaire**

Volume de crawl Googlebot VS **Positions dans Google**

Volume de crawl Googlebot VS **Liens externes**

Volume de crawl Googlebot VS **Contenu**

Si vous avez la data, la seule limite est votre imagination.

SEOLYZER





Merci !

Des questions?

 @Seolyzer_io olivier@seolyzer.io

SEOLYZER

