

# Trafic de bots : définition, types et bonnes pratiques de prévention

## Sydney va

11 mars 2024 14 min de lecture Contributeurs : Shannon Willoby et Bartłomiej Barcik

### Qu'est-ce que le trafic de bots ?

Le trafic de robots est un trafic non humain vers des sites Web et des applications généré par des programmes logiciels automatisés, ou « robots », plutôt que par des utilisateurs humains.

Le trafic des robots n'est pas un trafic de valeur, mais il est courant de le voir. Les robots d'exploration des moteurs de recherche (également appelés « spiders ») peuvent visiter votre site régulièrement, par exemple.

Le trafic des robots n'entraîne généralement pas de conversions ni de revenus pour votre entreprise. Toutefois, si vous êtes une entreprise de commerce électronique, vous pouvez rencontrer des robots d'achat qui effectuent des achats pour le compte de créateurs humains. Cependant, cela concerne principalement les entreprises qui vendent des articles très demandés, comme des billets de concert ou des sorties limitées de baskets.

Certains robots visitent votre site pour explorer les pages à des fins d'indexation par les moteurs de recherche ou pour vérifier les performances du site. D'autres robots peuvent tenter de récupérer (extraire des données) de votre site ou de surcharger intentionnellement vos serveurs pour attaquer l'accessibilité de votre site.

### Bons robots contre mauvais robots : identifier les différences

Il existe des robots bénéfiques et nuisibles. Ci-dessous, nous expliquons en quoi ils diffèrent.

#### Bons robots

Les robots courants incluent, sans s'y limiter :

- **Les robots d'exploration des outils SEO** : les robots d'exploration, tels que [SemrushBot](#), explorent votre site pour vous aider à prendre des décisions éclairées, comme l'optimisation des balises méta et l'évaluation de l'indexabilité des pages. Ces robots sont utilisés à bon escient pour vous aider à respecter les meilleures pratiques SEO.
- **Robots de surveillance de site** : ces robots peuvent vérifier les pannes du système et surveiller les performances de votre site Web. Nous utilisons [SemrushBot](#) avec des outils tels que [Site Audit](#) et bien d'autres pour vous alerter des problèmes tels que les temps d'arrêt et les temps de réponse lents. Une surveillance continue permet de maintenir des performances et une disponibilité optimales du site pour vos visiteurs.
- **Robots d'exploration des moteurs de recherche** : les moteurs de recherche utilisent des robots, tels que [Googlebot](#), pour indexer et classer les pages de votre site Web. Sans ces robots [qui explorent votre site](#), vos pages ne seraient pas indexées et les internautes ne trouveraient pas votre entreprise dans les résultats de recherche.

#### Les mauvais robots

Vous ne verrez peut-être pas régulièrement de trafic ni de preuves de la présence de robots malveillants, mais vous devez toujours garder à l'esprit le risque d'être ciblé.

Les robots malveillants incluent, sans s'y limiter :

- **Scrapers** : les robots peuvent extraire et copier le contenu de votre site Web sans votre autorisation. La publication de ces informations ailleurs constitue un vol de propriété

intellectuelle et une violation du droit d'auteur. Si des personnes voient votre contenu dupliqué ailleurs sur le Web, l'intégrité de votre marque peut être compromise.

- **Les robots de spam** : les robots peuvent également créer et distribuer du contenu de spam , comme des e-mails de phishing, de faux comptes de réseaux sociaux et des messages de forum. Le spam peut tromper les utilisateurs et compromettre la sécurité en ligne en les incitant à révéler des informations sensibles.
- **Bots DDoS** : les bots DDoS (Distributed Denial-of-Service) visent à surcharger vos serveurs et à empêcher les utilisateurs d'accéder à votre site Web en envoyant un flot de faux trafic. Ces bots peuvent perturber la disponibilité de votre site, ce qui peut entraîner des temps d'arrêt et des pertes financières si les utilisateurs ne sont pas en mesure d'accéder ou d'acheter ce dont ils ont besoin

## Comment le trafic des robots affecte les sites Web et les analyses

Le trafic des robots peut fausser les analyses du site Web et générer des données inexactes en affectant les éléments suivants :

- **Pages vues** : le trafic des robots peut gonfler artificiellement le nombre de pages vues, donnant l'impression que les utilisateurs interagissent avec votre site Web plus qu'ils ne le font en réalité
- **Durée de la session** : les robots peuvent affecter la durée de la session, qui mesure la durée pendant laquelle les utilisateurs restent sur votre site. Les robots qui parcourent votre site Web rapidement ou lentement peuvent modifier la durée moyenne de la session, ce qui rend difficile l'évaluation de la véritable qualité de l'expérience utilisateur.
- **Localisation des utilisateurs** : le trafic des robots crée une fausse impression de la provenance des visiteurs de votre site en masquant leurs adresses IP ou en utilisant des proxys
- **Conversions** : les robots peuvent interférer avec vos objectifs de conversion, tels que les soumissions de formulaires, les achats ou les téléchargements, avec de fausses informations et adresses e-mail

Le trafic des robots peut également avoir un impact négatif sur les performances de votre site Web et l'expérience utilisateur en :

- **Consommation des ressources du serveur** : les robots peuvent consommer de la bande passante et des ressources du serveur, en particulier s'ils sont malveillants ou volumineux. Cela peut ralentir les temps de chargement des pages, augmenter les coûts d'hébergement et même provoquer le blocage de votre site.
- **Endommager votre réputation et votre sécurité** : les robots peuvent nuire à la réputation et à la sécurité de votre site en volant ou en récupérant du contenu, des prix et des données. Une attaque (comme une attaque DDoS) peut vous coûter des revenus et la confiance de vos clients. Votre site étant potentiellement inaccessible, vos concurrents peuvent en tirer profit si les utilisateurs se tournent vers eux.

### Risques de sécurité associés aux robots malveillants

Tous les sites Web sont vulnérables aux attaques de robots, qui peuvent compromettre la sécurité, les performances et la réputation. Les attaques peuvent cibler tous les types de sites Web, quelle que soit leur taille ou leur popularité.

Le trafic des robots représente près de la moitié de tout le trafic Internet , et plus de 30 % du trafic automatisé est malveillant .

Les robots malveillants peuvent constituer des menaces de sécurité pour votre site Web car ils peuvent voler des données, envoyer du spam, détourner des comptes et perturber les services.

Les violations de données et les attaques DDoS sont deux menaces de sécurité courantes :

- **Violations de données** : des robots malveillants peuvent s'infiltrer sur votre site pour accéder à des informations sensibles telles que des données personnelles, des dossiers financiers et des informations de propriété intellectuelle. Les violations de données causées par ces robots peuvent entraîner des fraudes, le vol d'identité des personnes de votre entreprise ou des visiteurs de votre site, une atteinte à la réputation de votre marque, etc.
- **Attaques DDoS** : les robots malveillants peuvent également lancer des attaques DDoS qui ralentissent votre site ou le rendent indisponible pour les utilisateurs humains. Ces attaques peuvent entraîner une interruption de service, une perte de revenus et des utilisateurs insatisfaits.

## Comment détecter le trafic des robots

La détection du trafic des robots est importante pour la sécurité du site Web et des analyses précises.

### Identifier les robots avec des outils et des techniques

Il existe différents outils et techniques pour vous aider à détecter le trafic de robots sur votre site Web.

Certains des plus courants sont :

- **Analyse IP** : comparez les adresses IP des visiteurs de votre site avec les listes d'adresses IP de robots connus. Recherchez les adresses IP présentant des caractéristiques inhabituelles, telles que des taux de requêtes élevés, des durées de session courtes ou des anomalies géographiques.
- **Analyse du comportement** : surveillez le comportement des visiteurs et recherchez des signes indiquant une activité de robot, tels que des modèles répétitifs, une navigation inhabituelle sur le site et des temps de session courts

#### **Analyse du fichier journal**

Analysez les fichiers journaux de votre serveur Web. Les fichiers journaux enregistrent chaque requête adressée à votre site et fournissent des informations précieuses sur le trafic de votre site Web, telles que l'agent utilisateur, le référent, le code de réponse et l'heure de la requête.

Une analyse des fichiers journaux peut également vous aider à repérer les problèmes auxquels les robots d'exploration peuvent être confrontés sur votre site. [L'analyseur de fichiers journaux](#) de Semrush vous permet de mieux comprendre comment Google explore votre site Web.

Voici comment l'utiliser :

Accédez directement à l'analyseur de fichiers journaux ou connectez-vous à votre compte **Semrush**. Accédez à l'outil via la navigation de gauche sous « ON PAGE & TECH SEO ».

Avant d'utiliser l'outil, obtenez une copie du fichier journal de votre site à partir de votre serveur Web.

Le moyen le plus courant d'y accéder est d'utiliser un client FTP (File Transfer Protocol) comme [FileZilla](#) . Vous pouvez également demander une copie du fichier à votre équipe de développement ou informatique.

## Qu'est-ce qu'une analyse de fichier journal ? et comment la réaliser pour le référencement

Connor Lahey-10 février 2025-Contributeur : Sydney Go

### Que sont les fichiers journaux ?

Les fichiers journaux sont des documents qui enregistrent chaque requête adressée à votre serveur, qu'elle soit due à une personne interagissant avec votre site ou à un robot de moteur de recherche l'explorant (c'est-à-dire découvrant vos pages).

Les fichiers journaux peuvent afficher des détails importants sur :

- 
- L'adresse IP faisant la demande
- Quel robot a exploré votre site (comme Googlebot ou DuckDuckBot)
- Le type de ressource à laquelle on accède (comme une page ou une image)

Voici à quoi peut ressembler un fichier journal

:

```
66.249.66.192 - - [21/Jun/2018:02:08:49 +0200] "GET /sitemap_index.xml HTTP/1.1" 301 -
 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
66.249.66.192 - - [21/Jun/2018:02:08:49 +0200] "GET /sitemap.xml HTTP/1.1" 200 1637 "-"
 "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
66.249.66.221 - - [21/Jun/2018:02:20:18 +0200] "GET /amp/seo/optimisation-on-site/
analyse-de-logs.html HTTP/1.1" 200 9293 "-" "Mozilla/5.0 (Linux; Android 6.0.1; Nexus 5X
Build/MMB29P) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2272.96 Mobile Safari/
537.36 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
66.249.66.223 - - [21/Jun/2018:02:20:24 +0200] "GET /amp/skin/frontend/amp/custom/fonts/
opensans-bold-webfont.woff2 HTTP/1.1" 200 19676 "-" "Mozilla/5.0 (Linux; Android 6.0.1;
Nexus 5X Build/MMB29P) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2272.96 Mobile
Safari/537.36 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
66.249.66.221 - - [21/Jun/2018:02:20:30 +0200] "GET /amp/skin/frontend/amp/custom/fonts/
opensans-bold-webfont.woff HTTP/1.1" 304 - "-" "Mozilla/5.0 (Linux; Android 6.0.1; Nexus
5X Build/MMB29P) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2272.96 Mobile
Safari/537.36 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
66.249.66.221 - - [21/Jun/2018:02:30:52 +0200] "GET /media/css_secure/
dc99e6e3eebcaa34d571d04b205401d0.css HTTP/1.1" 200 1571 "https://www.410-gone.fr/e-
commerce/magento/developpeur-magento/expert-magento/certifications/frontend-
developer.html" "Mozilla/5.0 AppleWebKit/537.36 (KHTML, like Gecko; compatible;
Googlebot/2.1; +http://www.google.com/bot.html) Safari/537.36"
66.249.66.223 - - [21/Jun/2018:03:45:46 +0200] "GET /architecture-web/apache.html HTTP/
1.1" 200 5383 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/
bot.html)"
66.249.66.221 - - [21/Jun/2018:04:02:46 +0200] "GET /amp/seo/optimisation-on-site/robots-
txt.html HTTP/1.1" 200 7454 "-" "Mozilla/5.0 (Linux; Android 6.0.1; Nexus 5X Build/
MMB29P) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2272.96 Mobile Safari/537.36
(compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
66.249.66.223 - - [21/Jun/2018:04:02:48 +0200] "GET /amp/skin/frontend/amp/custom/fonts/
opensans-bold-webfont.woff2 HTTP/1.1" 200 19676 "-" "Mozilla/5.0 (Linux; Android 6.0.1;
Nexus 5X Build/MMB29P) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2272.96 Mobile
Safari/537.36 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
```

Les serveurs stockent généralement les fichiers journaux pendant une durée limitée, en fonction de vos paramètres, des exigences réglementaires en vigueur et des besoins de votre entreprise.

## Qu'est-ce que l'analyse des fichiers journaux ?

L'analyse des fichiers journaux est le processus de téléchargement et d'audit des fichiers journaux de votre site pour identifier de manière proactive les bogues, les problèmes d'exploration et d'autres problèmes [techniques de référencement](#).

L'analyse des fichiers journaux peut montrer comment Google et d'autres moteurs de recherche interagissent avec un site. Elle peut également révéler [des erreurs d'exploration](#) qui affectent la visibilité dans les résultats de recherche.

L'identification de tout problème avec vos fichiers journaux peut vous aider à démarrer le processus de résolution de ces problèmes.

## À quoi sert l'analyse des fichiers journaux dans le référencement ?

L'analyse des fichiers journaux est utilisée pour collecter des données que vous pouvez utiliser pour améliorer [la capacité d'exploration](#) de votre site et, en fin de compte, vos performances SEO.

C'est parce que cela vous montre exactement comment les robots des moteurs de recherche comme Googlebot explorent votre site.

Par exemple, l'analyse des fichiers journaux permet de :

- Découvrez quelles pages les robots des moteurs de recherche explorent le plus et le moins
- Découvrez si les robots d'exploration peuvent accéder à vos pages les plus importantes
- Vérifiez s'il existe des pages de faible valeur qui gaspillent votre [budget d'exploration](#) (c'est-à-dire le temps et les ressources que les moteurs de recherche consacreront à l'exploration avant de passer à autre chose)
- Détectez les problèmes techniques tels que les erreurs [de code d'état HTTP](#) (comme « erreur 404 page non trouvée ») et [les redirections](#) rompues qui empêchent les moteurs de recherche d'accéder à votre contenu
- Découvrez les URL avec [une vitesse de page](#) lente , ce qui peut avoir un impact négatif sur vos performances dans les classements de recherche
- Identifiez [les pages orphelines](#) (c'est-à-dire les pages sans liens internes pointant vers elles) que les moteurs de recherche peuvent manquer
- Suivez les pics ou les baisses de fréquence d'exploration qui peuvent signaler d'autres problèmes techniques

## Comment analyser les fichiers journaux

Maintenant que vous connaissez certains des avantages de l'analyse des fichiers journaux pour le référencement, voyons **comment** le faire.

Vous aurez besoin de :

- Les fichiers journaux du serveur de votre site Web
- Accès à un analyseur de fichiers journaux

**voici quelques logiciels d'analyse des fichiers journaux de site Web:**

- **ManageEngine EventLog Analyzer:** Cet outil permet de collecter, d'analyser et de visualiser les journaux de différentes sources. Il offre des alertes en temps réel, des rapports détaillés et une fonction de corrélation des journaux pour détecter les menaces et résoudre les problèmes de sécurité.

- **Logmatic.io**: Cette plateforme d'analyse de journaux basée sur le cloud offre une solution complète pour la gestion des journaux, l'analyse des données et la visualisation des informations. Elle permet de surveiller les performances des applications, de détecter les anomalies et de résoudre les problèmes rapidement.
- **Splunk Enterprise**: Ce logiciel puissant permet de collecter, d'indexer et d'analyser de grandes quantités de données de journaux provenant de différentes sources. Il offre des fonctionnalités avancées de recherche, de visualisation et de création de rapports pour aider à comprendre les événements et les tendances.
- **Graylog**: Cet outil open source de gestion des journaux offre une plateforme centralisée pour collecter, stocker et analyser les journaux de différentes sources. Il permet de rechercher rapidement des informations spécifiques, de créer des tableaux de bord personnalisés et de générer des alertes pour les événements importants.
- **Datadog**: Cette plateforme de surveillance et d'analyse offre une solution complète pour la gestion des journaux, la surveillance des performances des applications et l'analyse des données. Elle permet de détecter les problèmes, de visualiser les tendances et de collaborer avec les équipes pour résoudre les incidents.

Il existe de nombreux autres logiciels d'analyse des fichiers journaux de PC, chacun ayant ses propres fonctionnalités et avantages. Le choix du logiciel dépendra de vos besoins spécifiques en matière de collecte, d'analyse et de visualisation des journaux. N'hésitez pas à me poser d'autres questions.

## **logiciel d'analyse des fichiers journaux de semrush**

**, Semrush propose un outil d'analyse des fichiers journaux, appelé Log File Analyzer.**

**Cet outil permet d'analyser les fichiers journaux du serveur web pour comprendre comment les moteurs de recherche explorent et indexent votre site web.**

**Voici quelques-unes des fonctionnalités clés de Log File Analyzer :**

- **Analyse des données d'exploration** : Identifiez les pages explorées par les robots des moteurs de recherche, la fréquence de leur exploration et les erreurs rencontrées.
- **Détection des erreurs** : Repérez les erreurs techniques, structurelles ou de navigation qui peuvent entraver l'exploration et l'indexation de votre site web.
- **Optimisation du budget d'exploration** : Identifiez les pages inutiles ou peu importantes que les moteurs de recherche explorent, afin d'optimiser leur budget d'exploration et de concentrer leurs efforts sur les pages les plus importantes.
- **Amélioration du référencement** : Utilisez les informations obtenues grâce à l'analyse des fichiers journaux pour améliorer la structure de votre site web, corriger les erreurs et optimiser votre contenu, afin d'améliorer votre classement dans les résultats de recherche.

**Pour utiliser Log File Analyzer, vous devez télécharger les fichiers journaux de votre serveur web et les importer dans l'outil Semrush.**

**Une fois les fichiers journaux analysés, vous pourrez consulter les rapports et les graphiques générés par l'outil pour comprendre comment les moteurs de recherche explorent votre site web et identifier les points à améliorer.**

**Voici quelques ressources utiles pour en savoir plus sur Log File Analyzer :**

- **Page d'aide de Semrush sur Log File Analyzer** : <https://www.semrush.com/log-file-analyzer/>

- **Article de blog Semrush sur l'analyse des fichiers journaux :**  
<https://ralfvanveen.com/fr/seo/analyse-des-fichiers-journaux-pour-le-referencement/>

**Annexe : bibliographie**

- <https://www.semrush.com/blog/bot-traffic/>