

Kaspar Szymanski

DIE BEDEUTUNG VON SERVER LOGS FÜR SEO

Die meisten Webseitenbetreiber sind sich der Bedeutung der Webserver Logs nicht bewusst. Sie werden oft gar nicht aufgezeichnet, geschweige denn analysiert. Gerade große Marken betreiben keine Server-Log-Analyse und verlieren dadurch unwiederbringlich wertvolle Daten. Diejenigen Organisationen hingegen, welche im Rahmen ihrer SEO-Bemühungen auch die Server-Log-Analyse betreiben, haben bei der Google-Suche oft die besseren Karten. Wenn Ihre Webseite aus 100.000 oder mehr Seiten besteht und Sie herausfinden möchten, weshalb Server Logs enorme Wachstumsmöglichkeiten bieten und wie Sie diese nutzen können, dann lesen Sie bitte weiter.

DER AUTOR



Kaspar Szymanski, SEO-Berater und ehemaliger Mitarbeiter des Google Search Quality Teams, ist einer der wenigen Suchexperten mit jahrelanger Erfahrung mit der Gestaltung und Umsetzung von Googles Richtlinien, Web-spam-Bekämpfung und einem hohen Bekanntheitsgrad als Sprecher auf SEO-Events. Seit der Gründung von SearchBrothers.de steht er Unternehmen beratend zur Seite und spezialisiert sich im erfolgreichen Umgang mit Google-Strafen und Suchmaschinen-Optimierung.



Abb. 1: Die Aufrufe durch relevante Suchmaschinen-Bots können nur durch die Filterung der Server-Log-Dateien ermittelt werden

Weshalb sind Server Logs wichtig

Immer wenn ein Bot eine URL anfragt, die auf einem Webserver gehostet wird, wird dort automatisch ein Log-Eintrag erstellt. Dieser enthält die Informationen, die während des Transfers ausgetauscht werden. Über einen längeren Zeitraum hinweg widerspiegeln die Server Logs die Historie von Anfragen und Antworten der Webseite. Die Informationen im Server Log beinhalten üblicherweise die IP-Adresse des Clients, Datum und Zeit der Anfrage, die URL der angefragten Seite, den HTTP-Antwort-Code, die Anzahl übertragener Bytes sowie Informationen zum Browser, dem Betriebssystem und der zuvor besuchten Seite. Server-Log-Dateien werden immer angelegt, also auch wenn normale Nutzer die Webseite aufrufen. Bei der Suchmaschinenoptimierung konzentrieren wir uns aber auf die Daten im Server Log, die durch Bots generiert wurden, denn wir müssen dabei die Datenschutzgesetze

(GDPR/CCPA/DSGVO) befolgen. Da für SEO-Zwecke keine Nutzerdaten verwendet werden, unterliegt die Analyse anonymisierter Webserver Logs keinen rechtlichen Restriktionen.

Ähnliche Erkenntnisse kann man bis zu einem gewissen Grad auch aus den Crawl-Statistiken der Google Search Console gewinnen. Diese sind aber in ihrem Umfang begrenzt und umfassen nur die letzten paar Monate. Ein klares Gesamtbild der langfristigen SEO-Trends kann nur durch Server-Log-Dateien gewonnen werden.

Was Server Logs beinhalten

Wenn ein Bot eine Seite auf dem Webserver anfragt, werden verschiedene Daten im Webserver Log gespeichert, darunter:

- » Die IP-Adresse des Clients
- » Die genaue Zeit der Anfrage, oft basierend auf der internen Serveruhr
- » Die angefragte URL
- » Die in der Anfrage verwendete HTTP-Methode

- » Der Statuscode der Serverantwort (z. B. 200, 301, 404, 500)
- » User-Agent-Informationen zum Bot, wie der Name des Suchmaschinen-Bots (z. B. Googlebot/2.1)

Ein typischer Log-Eintrag könnte so aussehen:

```
150.174.193.196 - - [15/Dec/2021:11:25:14 +0100]
„GET /index.html HTTP/1.0“
200 1050 „-“ „Googlebot/2.1
(+http://www.google.com/bot.html)“ „www.example.ai“
```

Schauen wir uns dieses Beispiel genauer an:

- » 150.174.193.196 ist die IP des Clients.
- » [15/Dec/2021:11:25:14 +0100] ist das Datum und die Zeit – inklusive Zeitzone – der Anfrage.
- » „GET /index.html HTTP/1.0“ beschreibt die verwendete HTTP-Methode (GET), die angeforderte Datei (index.html) sowie die verwendete Version des HTTP-Protokolls.
- » 200 ist der HTTP-Statuscode, den der Server zurückgeliefert hat.
- » 1050 ist die Größe der Serverantwort in Bytes.
- » „Googlebot/2.1 (+http://www.google.com/bot.html)“ beschreibt die Browserinformationen des Clients bzw. den Namen des Suchmaschinen-Bots.
- » „www.example.ai“ ist die zuvor besuchte URL.

Wie Server Logs genutzt werden können

Bei der Suchmaschinenoptimierung gibt es drei Hauptgründe, weshalb Server Logs unschätzbare Erkenntnisse ermöglichen:

1. Sie helfen dabei, Anfragen unerwünschter Bots ohne SEO-Relevanz von Anfragen erwünschter Bots wie Googlebot, Bingbot oder Yandex-

- Bot zu unterscheiden.
2. Sie geben wertvolle Informationen über Crawl-Verhalten und die Priorisierung der Bots, wodurch das SEO-Team proaktiv das Budget für das Crawl-Management optimieren und anpassen kann.
 3. Sie erlauben die Überwachung der Serverantworten an die Suchmaschinen.

Bots, die lediglich vorgeben, ein Suchmaschinen-Bot zu sein, sind zwar ärgerlich, haben aber selten Einfluss auf eine Webseite. Mit spezialisierten Anbietern wie Cloudflare oder AWS Shield kann man den Zugriff durch unerwünschte Bots unterbinden. Bei der Analyse der Server Logs müssen die falschen Suchmaschinen-Bots nicht berücksichtigt werden. Der Server Log muss bei der Analyse ohnehin gefiltert werden, um herauszufinden, welche Teile der Webseite durch Suchmaschinen-Bots priorisiert werden. Je nach Markt liegt der Fokus auf Suchmaschinen-Bots von Google, Apple, Bing, Yandex oder anderen.

Gerade bei Webseiten mit tagesaktuellem Inhalt ist es für eine gute Nutzererfahrung entscheidend, wie oft die Seiten von den Suchmaschinen abgefragt werden. Anders ausgedrückt: Wenn die Inhaltsänderungen von den Suchmaschinen nicht schnell genug erkannt werden, können die Signale aus der Nutzererfahrung und die Platzierung in der organischen Suche nie ihr volles Potenzial entfalten.

Obwohl Google dazu neigt, alle verfügbaren Informationen zu crawlen und auch regelmäßig bekannte URLs abzufragen, sind die Ressourcen von Google nicht unbegrenzt. Dadurch sind große Webseiten mit Hunderttausenden von Zielseiten (Engl. Landingpages) vom Google-Algorithmus zur Zuweisung von Prioritäten abhängig, der die Zyklen zum Re-Crawling festlegt. Diese Zuweisung kann durch zuverlässige Betriebs-

Abb. 2: Eine möglichst große Überschneidung zwischen zu indexierenden Zielseiten und dem, was Google regelmäßig crawlt, ist ein positiver SEO KPI

zeit und reaktionsschnelle Webservices, die auf extrem schnelle Nutzererfahrung ausgelegt sind, positiv beeinflusst werden. Dies allein ist für SEO bereits sehr fördernd. Durch die Analyse der Server Logs über einen längeren Zeitraum wird es möglich, das Ausmaß der Überschneidungen zu ermitteln zwischen:

- » Der Anzahl zu durchsuchender Zielseiten
- » Der normalerweise kleineren Anzahl relevanter, optimierter und indexierter SEO-Zielseiten in der Sitemap
- » Dem, was Google schließlich regelmäßig im Crawling, der Indexierung und dem Ranking priorisiert

Diese Log-Analyse ist ein wesentlicher Bestandteil eines technischen SEO-Audits und die einzige Möglichkeit herauszufinden, wie viel vom Crawling-Budget verschwendet wird. Darüber hinaus zeigt die Analyse auf, ob indexierbare Filter, Platzhalter oder Seiten mit spärlichem Inhalt, offene Staging-Server oder andere überholte Teile der Webseite das Crawling und das Ranking in den Suchergebnissen negativ beeinflussen. In speziellen Fällen wie etwa

einer geplanten Migration entscheiden besonders die Erkenntnisse aus dem SEO-Audit mit der Server-Log-Analyse über Erfolg oder Misserfolg der Migration.

Die Analyse der Server Logs bringt großen Webseiten weitere wertvolle SEO-Hinweise. Sie beantwortet die Frage, **wie lange Google braucht, um die ganze Webseite erneut zu crawlen**. Wenn dies sehr lange dauert, also Monate oder noch länger, wäre es angebracht, das Crawling der primär zu indexierenden SEO-Zielseiten auszulösen. Andernfalls besteht die Gefahr, dass implementierte SEO-Verbesserungen selbst nach Monaten noch nicht von den Algorithmen der Suchmaschinen bemerkt wurden, was zu enttäuschenden Platzierungen in den Suchergebnissen führt.

Auch die Serverantworten sind für die Sichtbarkeit in der Google-Suche von entscheidender Bedeutung. Die Google Search Console erlaubt einen wichtigen Blick auf die kürzlichen Serverantworten. Diese Daten der Google Search Console sind einerseits



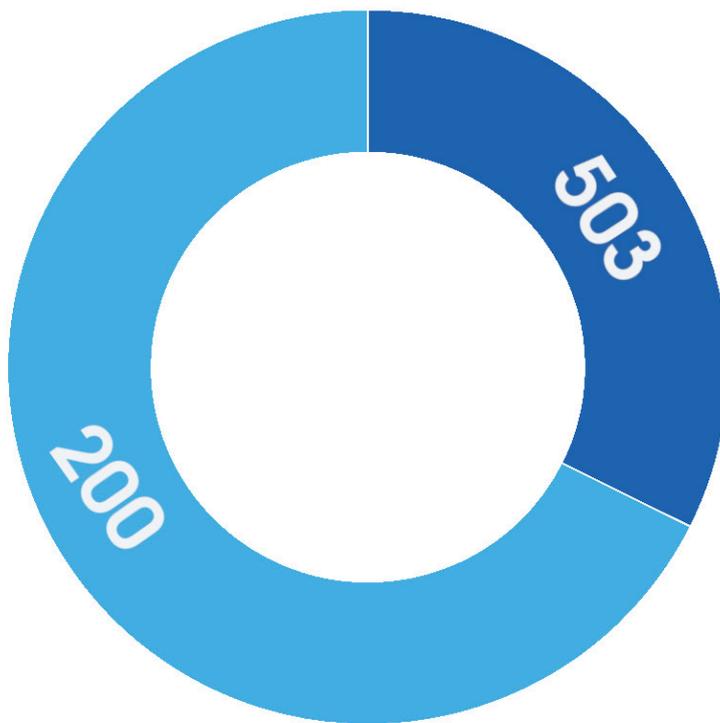


Abb. 3: Zu viele Non-200-OK-Serverantworten haben einen negativen Einfluss auf die Sichtbarkeit der Webseite in der organischen Suche

repräsentativ, aber eben auch nur eine begrenzte Auswahl. Man kann damit zwar die dringendsten Probleme erkennen, mittels einer Server-Log-Analyse können aber alle HTTP-Antworten analysiert werden, inklusive der mengenmäßig relevanten Non-200-OK-Antworten, welche die Platzierung in den Suchergebnissen gefährden können. Andere mögliche Antworten wie z. B. 503 Service unerreichbar wegen Serverwartung können, wenn sie zahlreich sind, auf ein Leistungsproblem hinweisen.

Wo soll man anfangen?

Trotz des offensichtlichen Potenzials, das der Server-Log-Analyse innewohnt, wird dieses von den meisten Webseitenbetreibern nicht genutzt. Entweder werden die Server Logs gar nicht erst aufgezeichnet oder sie werden überschrieben oder sind unvollständig. Eine überwältigende Mehrheit der Webseiten behält die Daten aus den Server Logs nicht lange genug, um sie sinnvoll verwerten zu können. Das ist wiederum umso vorteilhafter für diejenigen Websitebetreiber, die anders als die

Mitbewerber die Log-Dateien sammeln und für die Suchmaschinenoptimierung auswerten.

Bei der Planung der Datenakquise über Server Logs ist es zunächst wichtig zu wissen, welche Datenelemente zwingend im Server Log vorhanden sein müssen, damit die Daten zweckdienlich sind. Die folgende Auflistung soll dabei als Leitfaden dienen:

- » IP-Adresse des Clients, der die Anfrage gemacht hat
- » Informationen zum Browser (User Agent) oder Name des Bots
- » Schema der Anfrage
- » War es eine Anfrage über HTTP, HTTPS, WSS oder ein anderes Protokoll?
- » Hostname der Anfrage
- » An welche Domain oder Subdomain wurde die Anfrage gerichtet?
- » Anfragepfad (dies ist oft der Dateipfad auf dem Server in Form einer relativen URL)
- » Anfrageparameter, die Teil des Anfragepfads sein können
- » Zeitpunkt der Anfrage, inklusive Datum, Zeit und Zeitzone

- » Anfragemethode
- » Status-Code der HTTP-Antwort
- » Antwortzeiten

Der Hostname und der Typ der Anfrage werden bei der Aufzeichnung oft vergessen, wenn es sich beim Anfragepfad um eine relative URL handelt. Deshalb ist es wichtig, mit der IT-Abteilung zu klären und sicherzustellen, dass diese Datenelemente bei einer relativen URL ebenfalls in die Server Logs geschrieben werden. Eine Notlösung wäre, die gesamte angefragte URL als ein Datenelement zu speichern. Diese beinhaltet dann auch den Typ, den Hostnamen, den Pfad und die Parameter in einem String.

Falls Ihre Webseite weitere Dienste wie CDN (Content Delivery Networks) oder andere nutzt, sollten die dort gespeicherten Log-Dateien ebenfalls gesammelt und ausgewertet werden. Dazu müssen Sie mit dem Anbieter abklären, wie die Log-Dateien regelmäßig extrahiert und gespeichert werden können.

Hindernisse überwinden

Es sind meist zwei Gründe, die als vermeintlich unüberwindbare Probleme bei der Server-Log-Aufzeichnung bzw. -Analyse zur Sprache kommen: Kosten sowie auch rechtliche Bedenken. Auch wenn diese Faktoren sehr von den individuellen Umständen abhängen, wie z. B. dem vorhandenen Budget oder der anwendbaren Rechtsprechung, so muss dies nicht das Aus für das Vorhaben bedeuten. Eine langfristige Lösung könnte die Nutzung von Cloud-Speichern sein. Mit physischem Speichern kann man die Kosten noch mal erheblich senken. Bei Verkaufspreisen unter 600 US-Dollar für 20-TB-Festplatten können diese Kosten vernachlässigt werden. Die Kosten für Speicher-Hardware sinken schon seit Jahren und stellen daher kaum mehr eine Herausforderung für die Aufbewahrung von Server Logs dar.

Natürlich fallen noch zusätzliche Kosten an für die Software zur Log-Analyse oder den SEO-Audit-Anbieter, dieses Services. Diese Kosten müssen im Budget berücksichtigt werden, sind aber mit Hinblick auf die Vorteile der Server-Log-Analyse mehr als gerechtfertigt.

Dieser Artikel zeigt die bedeutenden Vorteile der Server-Log-Analyse für SEO auf, sollte aber indes nicht als rechtliche Empfehlung angesehen werden. Eine solche Empfehlung kann nur von einem qualifizierten Anwalt unter Berücksichtigung der rechtlichen Rahmenbedingungen und der geltenden Rechtsprechung erfolgen, denn in diesem Kontext können mehrere Gesetze und Regularien zur Anwendung kommen wie z. B. GDPR, CCPA oder DSGVO. Besonders beim Betrieb in der EU ist

Privatsphäre und Datenschutz eine wichtige Angelegenheit. Nutzerdaten sind allerdings bei der Server-Log-Analyse zu SEO-Zwecken nicht relevant. Dateneinträge, die aufgrund der IP nicht schlüssig verifiziert werden können, müssen ignoriert werden.

Konkret bedeutet das, dass alle Daten im Server Log, die nicht als Suchmaschinen-Bot identifiziert und bestätigt werden können, gemäß rechtlichen Empfehlungen nach einer gewissen Zeit entweder gelöscht oder anonymisiert werden müssen. Dieser Ansatz wird bei einigen der größten Webseitenbetreiber angewendet und hat sich bewährt.

Wann soll man beginnen?

Auf die große Frage, wann man mit dem Sammeln der Daten beginnen sollte, gibt es eine klare Antwort: Jetzt!

Die Daten aus dem Server Log können nur dann sinnvoll genutzt werden und zu umsetzbaren Ideen führen, wenn sie auch in ausreichender Menge vorhanden sind. Das kritische Maß der Server Logs liegt je nach Größe der Webseite und den Crawling-Prioritäten zwischen 6 und 36 Monaten.

Server Logs, die nicht aufgezeichnet wurden, können nachträglich nicht mehr beschafft werden. Die Bemühungen, die heute in die Aufbewahrung der Server Logs investiert werden, tragen voraussichtlich in einem Jahr die ersten Früchte. Daher muss das Sammeln der Daten in den Server Logs so früh wie möglich begonnen und so lange ununterbrochen fortgeführt werden, wie die Webseite betrieben wird und in den Suchergebnissen gute Ergebnisse erzielen soll. ¶

Digital-Marketing-Wissen bequem lernbar!

Digital-Marketing-Wissen bequem lernbar!
Unsere (Online-)Kurse! Für dich!

- geballtes Wissen zu den wichtigsten Themen des Digital Marketings als Kurs
- Expertenwissen
- 12 Monate verfügbar
- orts-ungebunden
- zeit-ungebunden
- jedes Kapitel jeder Zeit wiederholbar
- Wissensvermittlung mit vielen Tricks und Tipps
- sofort nach Zahlung verfügbar
- der Kurs ist auch in kleinen Kapiteln erhältlich

Kompletter Kurs auch als Professional
buchbar mit Prüfung und Zertifikat!

#ueberdentellerrandschauen #onlinemarketingwissen #weshareknowledge



121WATT

We share knowledge



Seminare
in staatlich anerkannter
MBA-Qualität



ISO 29990
Zertifiziertes
Qualitätsmanagementsystem
für Bildungsträger
www.tuev-sued.de/ms-zert

www.121watt.de