

Logfile-Analyse

aus Wikipedia, der freien Enzyklopädie <http://de.wikipedia.org/wiki/Logfile-Analyse>

Bei der **Logdateianalyse** wird die [Logdatei](#) eines Zeitraumes nach bestimmten Kriterien untersucht.

Je nach Art und Umfang der Aufzeichnung der Logdatei kann man daraus verschiedene Schlüsse ziehen.

Analyse von Webserver-Logdateien

Die interpretierten Statistiken ermöglichen es, den Aufbau und die Struktur der Internetseite zu optimieren. Sie sind Grundlage für [Benutzerfreundlichkeits](#)analysen oder liefern Aussagen über den Erfolg einer Marketingaktion. Ein Beispiel für eine Logfile-Analyse bietet die [Wikipedia Serverstatistik](#).

Einige mögliche Fragestellungen

- Wie lauten [IP-Adresse](#) und [Hostname](#) des Nutzers?
- Welchen [Browser](#) hat er genutzt?
- Auf welcher Seite stand der [Link](#), mit dem der Nutzer auf die Seite gekommen ist?
- Welche [Suchmaschine](#) und welche Suchwörter hat er genutzt?
- Wie lange blieb er auf der Webseite?
- Wie viele Seiten ruft er dabei auf?
- Auf welcher Seite hat er die Webseite verlassen?
- Welche [Ergänzungsmodule](#) hat er installiert?
- Welches [Betriebssystem](#) nutzt er?
- Welche Internetseiten hat Mitarbeiter Mustermann während der Arbeitszeit besucht? (hier muss in der Regel der [Betriebsrat](#) einbezogen werden)

Probleme bei der Analyse

Das Hauptproblem der Logfile-Analyse bei Webserver-Logfiles, ist die Tatsache, dass [HTTP](#) ein [zustandsloses Protokoll](#) ist. Das heißt, jede Anfrage eines [Clients](#) nach einer Webseite (oder jeder einzelnen darin vorkommenden Grafik usw.), ist für den Webserver eine eigenständige Aktion. Klickt sich der Benutzer durch eine Website, hat der Webserver keinerlei Kenntnis, dass der Benutzer gerade schon eine Seite abgerufen hat.

Um trotzdem ein zustandsbehaftetes HTTP zu ermöglichen, wird bei dynamisch erzeugten Webseiten gelegentlich beim ersten Aufruf des Benutzers eine so genannte [Session-ID](#) vergeben, die der Client dann bei den folgenden Anfragen immer mitsendet. Dies kann über einen [Cookie](#) oder einen an jeden [URI](#) zusätzlich angehängten Parameter erfolgen, wobei ein Cookie jedoch nicht im Logfile sichtbar ist und eine gesonderte Programmierung für die Logfile-Analyse benötigt.

Eine andere im HTTP-Protokoll vorhandene Möglichkeit einen Benutzer zu identifizieren, ist die Verwendung der IP-Adresse. Diese kann jedoch für viele unterschiedliche Benutzer dieselbe sein, wenn diese einen [Proxyserver](#), [Network Address Translation](#) oder Ähnliches verwenden. Sie sind daher nur mit größter Vorsicht zu verwenden, da eine IP-Adresse nicht mit einem Benutzer gleichzusetzen ist.

Oftmals hat der Betreiber einer Webseite jedoch keinen Zugriff auf das Logfile des Webservers, sodass häufig versucht wird mittels [Web-Bugs](#) dennoch eine statistische Auswertung zu ermöglichen. Hierzu werden kleine unsichtbare (1×1 Pixel, transparent) Bilder in die Webseite eingebunden, die auf einem Webserver abgelegt werden, dessen Logfile man auswerten kann.

Erweiterte Informationen, wie beispielsweise die [Bildschirmauflösung](#) oder eine Liste von installierten [Webbrowser-Plugins](#) werden auch gerne gewünscht, sind jedoch nicht in einem Logfile enthalten. Diese Informationen werden dann üblicherweise mittels einer clientseitigen Scriptsprache ermittelt, und ebenfalls mittels Web-Bug separat geloggt.

Programme zur Logfile-Analyse einer Website

- Open Source
 - Analog <http://www.analog.cx/>
 - AWStats <http://awstats.sourceforge.net/>
 - Webalizer <http://www.webalizer.org/>
 - LogReport <http://logreport.org/>
 - Visitors <http://www.hpimg.org/visitors>
- kostenlose Programme
 - Web Entry Miner WEM <http://www.ib.hu-berlin.de/~mayr/wem/>
 - W3 Statistics <http://free.w3statistics.com/>
- Kommerzielle Programme
 - APAGO® - User Tracking für große Websites <http://www.apago.de/>
 - W3 Statistics <http://www.w3statistics.com/>
 - ShopStat <http://www.shopstat.com/>
 - WebTrends <http://www.netiq.com/webtrends>
 - NetTracker <http://www.nettracker.de> und <http://www.sane.com/>
 - mycon Technologies [SalesAct](#) Echtzeit-Kundenanalyse und Webstatistiken
 - netUpdater LIVE | Echtzeit-User-Tracking-System | <http://www.netupdater.de>
 - Econda Monitor <http://www.econda.de/>
 - Funnel Web http://www.quest.com/funnel_web/analyzer/
 - [LogFile Analyse Pro](#)
 - Nedstat/Sitestat <http://www.nedstat.de/>
 - WebSpy <http://www.protectcom.de/webspy/>
- Produkt-Verzeichnisse
 - Deutschsprachig auf Web-Analytics.org <http://www.web-analytics.org/index.php/web-analyse-anbieter/>
 - Englischsprachig bei Terry Lund <http://www.terrylund.com/vendorListing.shtml>
 - <http://www.webanalyticsassociation.org/>

Artikel/Tutorials über Logfileanalyse

- Interessanter Artikel über den Aufbau von Apache Logfiles:
<http://www.devmag.net/servertechnik/logfiles.htm>
- [Magisterarbeit zum Thema Logfile-Analyse: "Entwicklung und Test einer logfilebasierten Metrik zur Analyse von Website Entries am Beispiel einer akademischen Universitäts-Website"](#)
- Logfile-Auswertung mit Access: <http://www.aspheute.com/artikel/20010502.htm>
- Artikel über die Vor- und Nachteile der Logfileanalyse im Vergleich zu anderen Analysemethoden:
[Welche Möglichkeiten der Datenerhebung gibt es?](#)
- [LogFile Analyse Tipps für die Suchmaschinenoptimierung](#)
- Kimball, R. / Merz, R.: The Data Webhouse Toolkit, New York et al. 2000.
- Massand, B. / Spiliopolou, M.: Web Usage Analysis and User Profiling, Berlin et al. 2000.
- Lutzky, C. / Teichmann, M.-H.: Logfiles in der Marktforschung: Gestaltungsoptionen für Analysezwecke, Jahrbuch der Absatz- und Verbrauchsforschung, 48. Jg., 2002, S. 295 - 317.
- Heindl, E.: Logfiles richtig nutzen. Webstatistiken erstellen und auswerten, Galileo Press 2003, Bonn

Logfile

aus Wikipedia, der freien Enzyklopädie <http://de.wikipedia.org/wiki/Logfile>

Eine **Logdatei** (engl. *log file*) beinhaltet das automatisch erstellte Protokoll aller oder bestimmter Aktionen von einem oder mehreren Nutzern an einem Rechner, ohne dass diese davon etwas mitbekommen oder ihre Arbeit beeinflusst wird.

Wesentlich ist das System-Logbuch (unter Linux meist in /var/log/messages, *aber auch Windows NT und Nachfolger schreiben entsprechende Einträge in ein oder mehrere Ereignisprotokolle*). Darin werden u. a. die Anmeldungen am System protokolliert, aber auch weitere wichtige Informationen abgelegt. Praktisch alle Hintergrundprogramme (so auch das Betriebssystem selbst - der Kernel) z. B. der [E-Mail-Server](#), der [Proxyserver](#) u. a. m. schreiben in eine Logdatei (sie haben ja keine Konsole, auf der die Informationen ausgegeben werden können).

Logdateien werden auch bei [Webservern](#) erstellt, können aber auch außerhalb des Internets bei Untersuchungen der [Benutzerfreundlichkeit](#) von Programmen oder allgemeinem Nutzerverhalten in einem System genutzt werden.

Kann man selber auf einem Server mit der eigenen Webpräsenz keine Logfiles abrufen, dann ermöglichen verschiedene Webdienste (z.B. [Netstat](#) oder [W3Research](#)) das Setzen eines kleinen Zählers, auf der eigenen Webpräsenz. Dieser Zähler wird immer wieder beim Abrufen der Seite vom Anbieter-Server geladen, wo darauf hin eine Auswertung der Zugriffe in Form von Tabellen und Grafiken erstellt wird.

Die Logdatei wird bei der [Logfile-Analyse](#) untersucht.

Beispiel einer typischen Webserver-Logdatei

```
183.121.143.32 - - [18/Mar/2003:08:04:22 +0200] "GET /images/logo.jpg HTTP/1.1" 200 512 "http://www.wikipedia.org/"
"Mozilla/5.0 (X11; U; Linux i686; de-DE; rv:1.7.5)
183.121.143.32 - - [18/Mar/2003:08:05:03 +0200] "GET /images/bild.png HTTP/1.1" 200 805 "http://www.google.org/".
```

Teilbereiche eines Eintrages

- Wer? - 183.121.143.32 - anfordernde Host-Adresse
- Wann? - [18/Mar/2003:08:04:22 +0200] - Zeitstempel (Datum, Uhrzeit, Zeitverschiebung)
- Was? - "GET /images/logo.jpg HTTP/1.1" - Anforderung, eines Bildes, Übertragungsprotokoll
- Ok? - 200 - Statusnummer (200=Erfolgreiche Anfrage)
- Wieviel - 512 - Menge der gesendeten Daten (Byte)
- Woher? - "<http://www.wikipedia.org/>" - Von welcher Internetseite wird angefordert
- Womit? - "Mozilla/5.0 (X11; U; Linux i686; de-DE; rv:1.7.6)" - Mit welchem Browser/Betriebssystem/Oberfläche

Literaturhinweise zur Auswertung

- Kimball, R. / Merz, R.: The Data Webhouse Toolkit, New York et al. 2000.
- Massand, B. / Spiliopolou, M.: Web Usage Analysis and User Profiling, Berlin et al. 2000.
- Lutzky, C. / Teichmann, M.-H.: Logfiles in der Marktforschung: Gestaltungsoptionen für Analysezwecke, Jahrbuch der Absatz- und Verbrauchsforschung, 48. Jg., 2002, S. 295 - 317.