LU08c - LogAudit

Lernziele

- Die drei wichtigsten Kennzahlen für LogFiles nennen und erkläre können.

LogFiles analysieren

Das Zugriffsprotokoll (access_log) eines Webservers enthält viele interessante Informationen über die Nutzung einer Website. Oder, auf den Punkt gebracht, sie enthält: Wann welche Inhalte wie oft von wie vielen Nutzern abgerufen werden.

Die Nutzung einer Website kann durch Kennzahlen beschrieben werden, die ohne grossen Aufwand und mit Hilfe von Auswertungsprogrammen aus den LogFiles generiert werden können. Dies sind:

- Summe aller Aktiväten auf dem Server (Traffic)
- Anzahl der abgerufenen Seiten (Page Impressions)
- Anzahl aller übertragenen Dateien (Hits)
- Anzahl der Anwendersitzungen (Visits)

Website Traffic	Der Begriff *Website Traffic* ist ein Oberbegriff und bezeichnet die gesamte Nutzungsintensität einer Website. Sprich ist die Messung von Besuchern, Seitenaufrufen oder übertragenen Datenmengen und ist damit quasi die **Summe aller Aktivitäten auf dem Webserver**.
Page Impressions	*Page Impressions* (oft auch Page Views genannt) sind die Anzahl der HTML-Dokumente, die vom WebServer abgerufen wurden. Grafiken und andere Dateiformate bleiben unberücksichtigt. In einer LogFile-Analyse sollten somit immer mehr *Hits* als *Page Impressions* vorhanden sein. Der Zugriff auf eine HTML-Seite, in der 10 Grafiken enthalten sind, verursacht 11 Hits und eine Page Impression.
Hits	*Hits* sind die Anzahl aller Objekte, die vom WebServer abgerufen werden. Dabei ist es gleichgültig, ob es sich dabei um statische HTML-Dokumente oder dynamisch durch CGI-Skripte erzeugte Dokumente handelt, um Grafiken, Audio- oder Videodateien oder andere Dateitypen. Der Begriff *Hits* wurde in den Anfangsjahren des WWW eingeführt, ist aber eine wenig aussagekräftige Messgrösse: Wenn eine HTML-Seite viele grafische Elemente enthält, werden auch mehr Hits erzeugt, ohne dass dies etwas über das Benutzerverhalten aussagen würde.
Visits	*Visits* sind die Anzahl der Anwendersitzungen einer Website. Eine Anwendersitzung ist dabei ein zusammenhängender Nutzungszeitraum einer Website durch einen Clientrechner, beginnend mit dem erstmaligen Abruf einer Ressource auf dieser Website. Ein *Visit* wird als beendet betrachtet, wenn der Clientrechner einen festgelegten Zeitabschnitt lang keine Anforderung gestellt hat. Visits sind nicht gleichbedeutend mit der Anzahl der Benutzer einer Website. Die Berechnung der Anzahl der eindeutigen Benutzer ist nur eingeschränkt und unter einigen Annahmen möglich. Welche Probleme hier bestehen, wird nachfolgend erläutert.

Man kann es sich wie eine russische Matroschka vorstellen: Traffic > Visits > Page Views > Hits, und parallel dazu die Impressions als Zählung, wie oft ein Element im Sichtfeld landet.

Mit der Ermittlung dieser vier Werte sind die Möglichkeiten der LogFile-Analyse längst noch nicht erschöpft. Das Verhältnis aus Page Impressions und Visits – wie viele Seiten werden pro Anwendersitzung abgerufen – sagt viel über das Nutzungsverhalten der Besucher aus. Eine geringe Zahl an Page Impression pro Visit kann auf eine wenig attraktive Website hindeuten. Doch ein hoher Wert muss nicht unbedingt das Gegenteil bedeuten. Eine grosse Zahl an Page Impression pro Visit kann auch Ausdruck einer nicht intuitiven Benutzerführung und Navigation sein.

Wird die Zeit bei einer Analyse mitberücksichtigt, führt das zu Rückschlüssen auf die durchschnittliche Dauer eines Besuches. Ausserdem sind Aussagen über das Nutzungsverhalten an bestimmten Tageszeiten oder Wochentagen möglich. Auch Aussagen über die überweisende Website (Referrer), also die Website, von der der Besucher über einen Link oder eine Suchanfrage gekommen ist, lassen sich treffen. War dies eine Suchmaschine, können in der Regel sogar die Suchbegriffe ausgewertet werden, die der Benutzer dort eingegeben hat. Nicht zuletzt zeigt eine LogFile-Analyse auch Fehler in einer Website an. Der protokollierte Antwortcode 404 weist auf ein fehlendes Dokument hin. Ein häufiges Auftreten des Antwortcodes 500 deutet eventuell Ressourcen-Probleme des WebServers an. Auch versuchte und unerlaubte Zugriffe auf zugangsgeschützte Dokumente oder Verzeichnisse können ausgewertet werden.

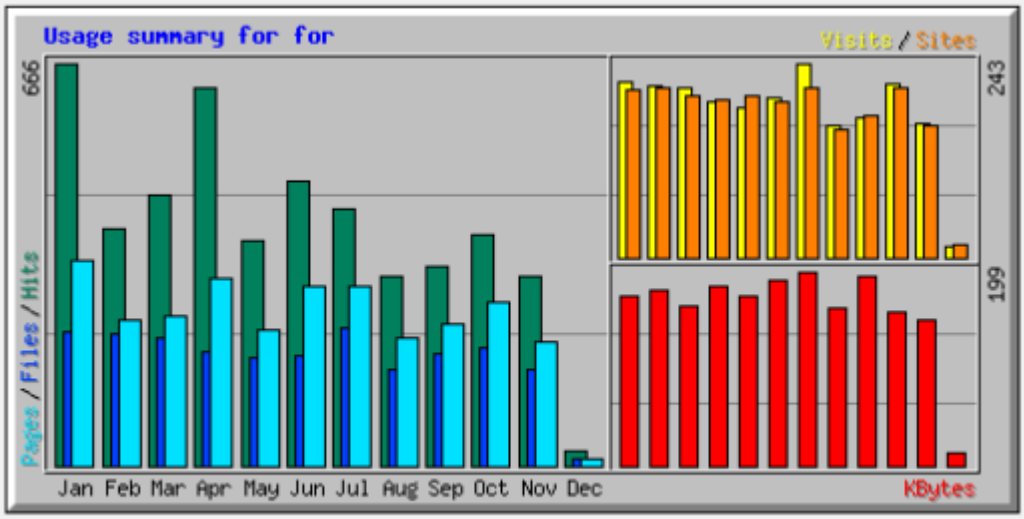
LogFiles mit webalizer auswerten

Für die Auswertung der WebServer LogFiles steht eine Vielzahl von Anwendungen zur Verfügung. Im folgenden Beispiel verwenden wir *webalizer* in dem Prinzip zu Erläutern.

Das Tool wird auf ein LogFile losgelassen, wertet dieses aus und speichert das Ergebnis in Form eines Reports ab. Der Report besteht aus Grafiken und Statistiken und wird in Form von HTML-Seiten generiert, die sich dann über den Web-Browser betrachten lassen. Der Zugriff auf die WebSite Statistik sollte dabei immer mit einem Passwort geschützt werden, um nicht ein Sicherheitsloch zu eröffnen.

Übersicht

Das nachfolgene Schaubild zeigt eine Visualisierung der Logfiles des Jahres 2017 einer Webseite durch Tools wie *Webalizer*.

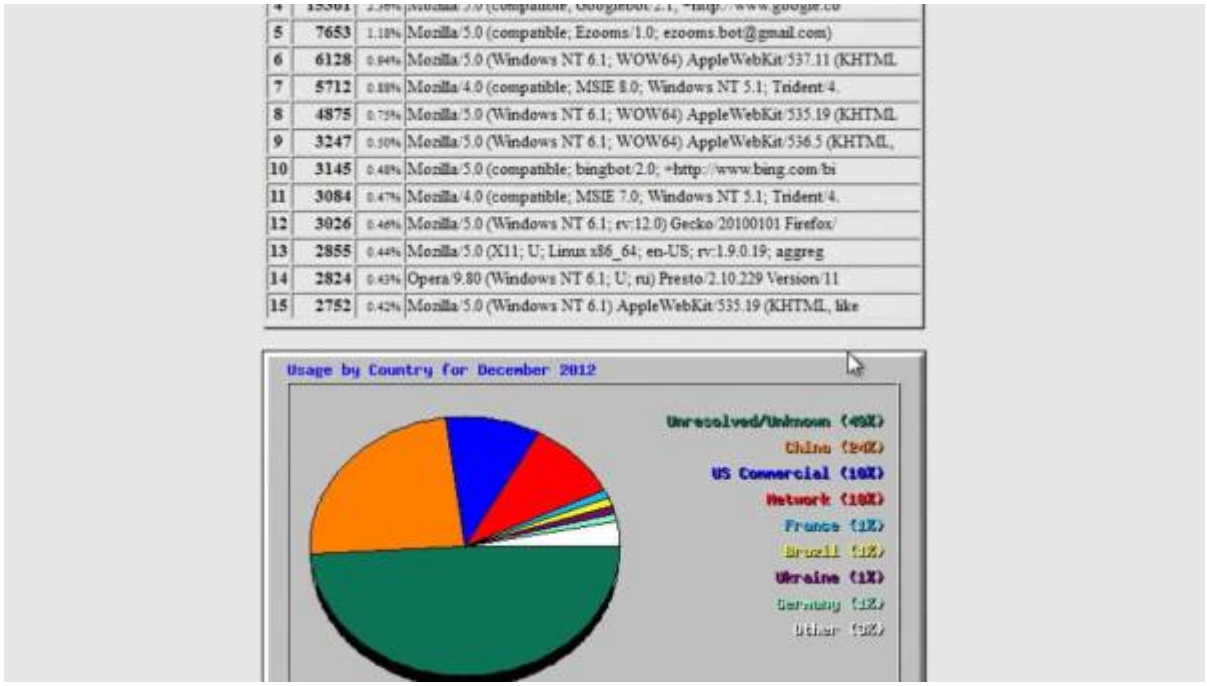


Summary by Month

Month	Daily Avg				Monthly Totals					
	Hits	Files	Pages	Visits	Sites	KBytes	Visits	Pages	Files	Hits
Dec 2017	6	3	3	3	15	13	13	13	12	24
Nov 2017	10	5	6	5	164	149	168	205	159	314
Oct 2017	12	6	9	7	213	157	216	272	194	383
Sep 2017	10	6	7	5	177	193	176	234	185	329
Aug 2017	10	5	6	5	159	162	164	213	158	313
Jul 2017	13	7	9	7	213	199	243	297	229	426
Jun 2017	15	6	9	6	195	190	199	296	183	472
May 2017	12	5	7	6	201	174	187	226	180	373
Apr 2017	20	6	10	6	198	183	194	309	189	624
Mar 2017	14	6	8	6	201	164	213	248	212	447
Feb 2017	14	7	8	7	211	179	215	241	219	392
Jan 2017	21	7	10	7	209	174	220	339	221	666
Totals						1937	2208	2893	2141	4763

Detailed Statistics

Nach einer Übersichtsgrafik liefert *webalizer* weitere Detailauswertungen zur konfigurierten WebSite. In der nachfolgenden Abbildung sind die Kennzahlen auf Monatsebeneersichtlich.



5	7653	1.18%	Mozilla/5.0 (compatible; Ezooms/1.0; ezooms.bot@gmail.com)
6	6128	0.94%	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.11 (KHTML
7	5712	0.88%	Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.
8	4875	0.75%	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/535.19 (KHTML
9	3247	0.50%	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML,
10	3145	0.48%	Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bi
11	3084	0.47%	Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1; Trident/4.
12	3026	0.46%	Mozilla/5.0 (Windows NT 6.1; rv:12.0) Gecko/20100101 Firefox/
13	2855	0.44%	Mozilla/5.0 (X11; U; Linux x86_64; en-US; rv:1.9.0.19; aggreg
14	2824	0.43%	Opera/9.80 (Windows NT 6.1; U; ru) Presto/2.10.229 Version/11
15	2752	0.42%	Mozilla/5.0 (Windows NT 6.1) AppleWebKit/535.19 (KHTML, like

From:
https://wiki.bzz.ch/ - **BZZ - Modulwiki**

Permanent link:
https://wiki.bzz.ch/modul/m183/learningunits/lu08/03?rev=1757323030

Last update: **2025/09/08 11:17**