Webalizer - Statistiksoftware für den Webserver

Es gibt eine ganze Reihe von Gründen, warum eine Webserver-Statistik benötigt wird. Für jeden Anbieter von WWW-Seiten ist es wichtig, zu überprüfen, welchen Anklang das eigene Angebot bei den weltweiten Internet-Nutzern findet. Auf dem Webserver der Humboldt-Universität kommt seit kurzem die Statistik-Software Webalizer (*http://www.webalizer.org/*) für die Analyse der Server-Logdaten zum Einsatz.

Die vom Webalizer erzeugte Statistik besteht aus zwei Teilen, der Jahresübersicht und den einzelnen Monatsstatistiken. Die Jahresübersicht gibt Zugriffshäufigkeiten unter verschiedenen Aspekten wieder, deren Bedeutung in Tabelle 1 kurz erläutert wird.

Neben den verschiedenen Anzahlen werden auch die dabei übertragenen Datenmengen in kBytes (1kByte=1024Bytes) wiedergegeben.

Bei der Interpretation der Zahlen ist Vorsicht geboten. Die Hits sind natürlich beeindruckend, verzerren aber das Bild von den eigentlichen Nutzerzahlen, da wirklich jede Anfrage, auch jeder Zugriff auf Bilder und Icons, gezählt wird. Schon eher brauchbar ist die Pages-Angabe, da dort nur Anfragen, die HTML-Seiten betreffen, gezählt werden. Jedoch gibt das auch noch keine Auskunft darüber, wie viele Nutzer auf die Seiten zugegriffen haben. Dafür eignen sich aber die Sites- und Visits-Angaben.

Die Monatsstatistik besteht aus vielen Einzelstatistiken (siehe Tabelle 2). Neben den allgemeinen Zugriffshäufigkeiten werden hier auch die am häufigsten besuchten Seiten und die zugreifenden Browser aufgelistet.

Hits	Anzahl aller Anfragen an den Webserver
Files	Anzahl aller Anfragen an, bei denen etwas an den Klienten zurückgeschickt wurde (also HTML-Seiten, Grafiken etc.)
Pages	Anzahl aller abgefragten HTML-Seiten (.htm-, .html-, aber auch .phtml- und .php3-Seiten ohne Bestandteile wie Grafiken etc.)
Sites	Anzahl der verschiedenen Internetadressen, von denen zugegriffen wurde
Visits	Anzahl der Besuche von einer Site, unter dem Begriff „Besuch“ werden hier alle Anfragen, die innerhalb eines bestimmten Zeitraums (z. B. innerhalb von 30 Minuten) von einer Site zugegriffen haben, zusammengefasst

Tab. 1: Die Elemente der Jahresübersicht

Monthly	Monatsübersicht, wie sie auch schon in der Jahresübersicht zu finden ist, nur etwas ausführlicher, z. B. auch eine Übersicht über die Antwort-Codes des Webservers
Dayly	Übersicht für die einzelnen Tage des jeweiligen Monats
Hourly	Durchschnitts- und Total-Werte für die einzelnen Stunden
Top URL's	Statistik, die zum einen eine Top-Liste von Seiten mit den größten Anfragehäufigkeiten, zum anderen eine Top-Liste von Seiten mit dem höchsten Datenaufkommen ausgibt
Entry Pages	Statistik, die zeigt, welche Seiten als Haupt-Einstiegsseiten benutzt werden, d.h. es wird danach sortiert, welche Seiten am häufigsten als Referrer (HTML-Seite, von der der Nutzer auf die nächste Seite gekommen ist) auftauchen
Exit Pages	Statistik, die zeigt, welche Seiten als Haupt-Ausstiegsseiten dienen, d.h. die Seiten, die am seltensten oder gar nicht als Referrer auftauchen (Gegenstück zur Entry Pages-Statistik)
Site	Statistik, die alle und vor allem vollständige IP-Adressen der zugreifenden Computer wiedergibt (Die Site-Statistik sollte deshalb in der Regel nicht öffentlich zugänglich gemacht werden!)
Referrers	Statistik, die alle externen Web-Adressen enthält, von denen ein Link auf die eigene Website verweist, sowie die Häufigkeit, mit der dieser Link benutzt wurde
Search	Statistik, die die am häufigsten angeklickten Suchbegriffe (Textanker) enthält
User Agents	Übersicht über alle zugreifenden WWW-Browser
Countries	Statistik, die Auskunft darüber gibt, aus welchen Ländern am häufigsten zugegriffen wurde

Tab. 2: Die Elemente der Monatsübersicht

Viele der einzelnen Statistiken werden auch als Diagramm dargestellt, was einen guten Überblick über zeitliche Verläufe oder Häufigkeiten ermöglicht.

Auf dem WWW- und dem WWW2-Server gibt es auf Anfrage (webadm@rz.hu-berlin.de) die Möglichkeit, für bestimmte Server-Bereiche (Verzeichnisse) eigene Statistiken zu erstellen (i. d. R. ab dem Zeitpunkt der Einrichtung der Statistik). Die allgemeinen Serverstatistiken befinden sich unter
http://www.hu-berlin.de/statistik/www/ und
http://www2.hu-berlin.de/statistik/www2/.

Daniel Rohde
d.rohde@rz.hu-berlin.de