

9.4 Logfile-Analyse

EDUARD HEINDL

Einleitung

Webserver zeichnen während des Betriebs viele Informationen über die Besucher und ihre Wünsche auf. Diese detaillierte Aufzeichnung von Nutzerverhalten, das nicht nur eine statistische Stichprobe ist, wie in der Marktforschung üblich, liefert bei kluger Auswertung ungewöhnlich reiche Information zum Verbessern des Internetauftritts im Speziellen und zum Optimieren des E-Business im Allgemeinen. Allerdings sind die Informationen tief in den Logfiles, die auch mit heutigen Maßstäben gemessen eine enorme Größe erreichen können, vergraben. Sie werden erst durch den Einsatz geeigneter Software sichtbar und können dann weitergehend interpretiert werden. Im Folgenden werden die Punkte besprochen: Was kann aufgezeichnet werden? Wie wird ausgewertet? Welche Interpretation ist sinnvoll?

Informationen über
das Nutzerverhalten

Aufzeichnen der Daten

Jedes Mal, wenn ein Internetsurfer mit seinem Browser ein Internetdokument aufruft, sendet der Browser einen Header. Dieser Header ist ein kurzer Text, in dem Informationen über die verwendete Softwareplattform, den Browser, seine spezielle Konfiguration, etwa die gewünschte Sprache, und natürlich das gewünschte Dokument beschrieben sind.

Der Header

Personalisierung mit Cookies

Unter bestimmten Umständen wird zusätzlich personalisierte Information versendet. Dies geschieht mit den so genannten Cookies, kleinen Textfiles, in denen Informationen stehen, die der Internetnutzer bei seinem

letzten Besuch am Server zugesendet bekam und im Browser gespeichert sind. Zusätzlich kann der Webserver weitere Informationen anfordern oder der Browser sendet Zusatzinformationen auf Grund eines kleinen Scripts, das in die Webseite eingebaut ist.

Aufzeichnung aller Aktivitäten

Bemerkenswert ist die Tatsache, dass diese Detailinformation für jedes einzelne abgerufene File vom Webserver aufgezeichnet wird. Da eine Standard-Webseite aus vielen Einzelementen wie Bildern, Sound, Formatierungsfiles, Scripten und anderem besteht, werden im Normalfall bereits heute etwa 40 Einzelemente angefordert, um die Webseite im Browser multimedial erscheinen zu lassen.

Elemente eines Standardlogfiles

- IP Adresse des anfragenden Rechners (Browser)
- Benutzer, falls Identifikation stattfindet (selten)
- Zeit und Datum relativ zu GMT
- Angefragtes Dokument
- Statusmeldung, z.B. 200:ok, 404:Fehlermeldung
- Anzahl gesendeter Bytes
- Referer, letzte betrachtete Webseite (ggf. mit Suchworten)
- Verwendeter Browser und Betriebssystem

Speicherung in
Textdateien

Aus Gründen der Systemperformance sind diese Logfiles im Webserver lange Textdateien, bei denen jede Zeile eine einzelne Dokumentanfrage eines Internetclients enthält.

Nachträgliche Auswertung

Nach Ablauf einer festen Zeit, täglich, wöchentlich oder monatlich, werden diese Logfiles vom Webserver geholt und die Auswertung mit Spezialsoftware kann beginnen. Alternativ ist auch eine Echtzeiterfassung in einer Datenbank und synchrone Ausgabe der Ergebnisse möglich. Dies dient aber meist nur dazu, Zugriffssummen zu ermitteln. Hier werden nur die a posteriori Auswertungen betrachtet, da diese wesentlich tiefer gehen können.

Analyse der gesammelten Daten

Für größere Websites stellt die Analyse der Logfiles einerseits eine außergewöhnlich ertragreiche Quelle über das Nutzerverhalten dar, aber auch eine enorme Herausforderung an die Auswertungssoftware und die Hardware, auf der die Auswertung läuft. Dies wird an einem kleinem Beispiel sofort klar: Auf einer gut besuchten Webseite kommen täglich 10.000 Besucher vorbei (es gibt Websites mit tausendmal mehr Besuchern!), wobei jeder drei Seiten betrachtet und jedes Dokument aus 50 Einzelementen besteht. Wöchentlich wird das Logfile ausgewertet, der sinnvollste Zeitraum. Dieses Logfile hat etwa eine Größe von einem Gigabyte. Für die Auswertung muss daher ein Rechner mit ausreichend Festplattenkapazität und Hauptspeicher zur Verfügung stehen.

Herausforderung für die Auswertungssoftware

Die Kenngrößen einer Website

Für die weitere Betrachtung ist es notwendig, einige Begriffe der Webstatistik kurz zu beschreiben, damit keine Verwechslungen auftreten.

Begriffe der Webstatistik

Hits

Die Anzahl der Zeilen im Logfile entspricht der Anfragen aus dem Internet nach einzelnen Files auf dem Webserver. Diese Zahl wird als Hits bezeichnet und ist die höchste Zahl aller relevanten statistischen Kenngrößen. Ihre Aussagekraft ist aus mehreren Gründen für das Marketing gering. Einmal hängt diese Zahl empfindlich vom Design einer Webseite ab, bereits kleine Änderungen können Schwankungen um den Faktor zwei verursachen. Weiterhin werden von den meisten Programmen auch dann Hits gezählt, wenn keine erfolgreiche Aussendung eines Dokuments stattfand, Error 404! Die wichtigste Bedeutung dieser Größe liegt in der Beurteilung der technischen Serverauslastung.

Datenvolumen

Aus den abgerufenen Dokumenten und der jeweiligen Filegröße wird das abgerufene *Datenvolumen* bestimmt. Dies scheint auf den ersten Blick eine sehr technische Größe zu sein, es spricht aber einiges für die Bedeutung, da möglicherweise der Umsatz auf einem

Webserver proportional zum Datentransfer ist, Näheres gegen Ende des Artikels.

Page-Impression

Weit verbreitetste
Kenngröße

Die weit verbreitetste Größe stellt die Anzahl der Page-Impressions (PI) dar. Hier werden die Dokumente gezählt, die dem Besucher als eigenständige Webseite im Browser erscheinen. Diese etwas komplizierte Umschreibung von gesehenen Dokumenten muss wegen der komplexen Arbeitsweise von Frameseiten erfolgen, die aus mehreren einzelnen HTML-Dokumenten zusammengesetzt sind. Für den Werbetreibenden ist diese Größe auch für die Anzahl der Bannereinblendungen relevant.

Page-View

Eine sehr ähnlich Größe sind die Page-Views (PV). Falls keine Frames verwendet werden, ist diese Zahl identisch mit dem Page-Impressions, andernfalls ist sie höher und beschreibt die Anzahl der versendeten HTML-Dokumente. Die Auswertung von Frameseiten wird daher sehr komplex, insbesondere wenn unterschiedlich umfangreiche Frameseiten verwendet werden.

Visit

Besucherzahl
einer Website

Für die Beurteilung der Besucherzahl einer Website dient die Anzahl der Visits, in der die Anzahl unterschiedlicher Sitzungen gezählt wird. Diese Größe ist allerdings nicht direkt aus dem Logfile zu ermitteln, falls keine Cookies verwendet werden, da keine feste Verbindung zwischen Browser und Webserver aufgebaut wird. Man behilft sich daher mit der Auswertung der anfragenden Internetadresse (IP), die vom Browser (Betriebssystem) übermittelt wird. Hier ist es von Nachteil, dass bei vielen Providern die IP Adresse nicht fest vergeben wird und daher hinter einer Nummer mehrere Nutzer verborgen sein können. Um dies auszuschalten, werden Standardruhezzeiten verwertet, das bedeutet, wenn ein Besucher innerhalb von 10 Minuten kein weiteres Dokument von der Website abrufen geht man vom Ende der Sitzung aus. Ein problematisches Vorgehen, da sich gezeigt hat, dass besonders in der Phase des Onlinekaufens häufig zu Mitbewerbern gesprungen

wird und erst nach mehr als 10 Minuten der User zurückkehrt, der dann scheinbar ein anderer Besucher ist.

Weitere Größen

Es lassen sich für Websites viele weitere Kenngrößen finden, wie neue Besucher, Verweilzeit und Anzahl unterschiedlicher Besucher. Allerdings sind diese Größen oft noch schlechter definiert und zudem mit erheblichen Messfehlern behaftet, da sie indirekt gewonnen werden und Zusatzannahmen zur Bestimmung nötig sind.

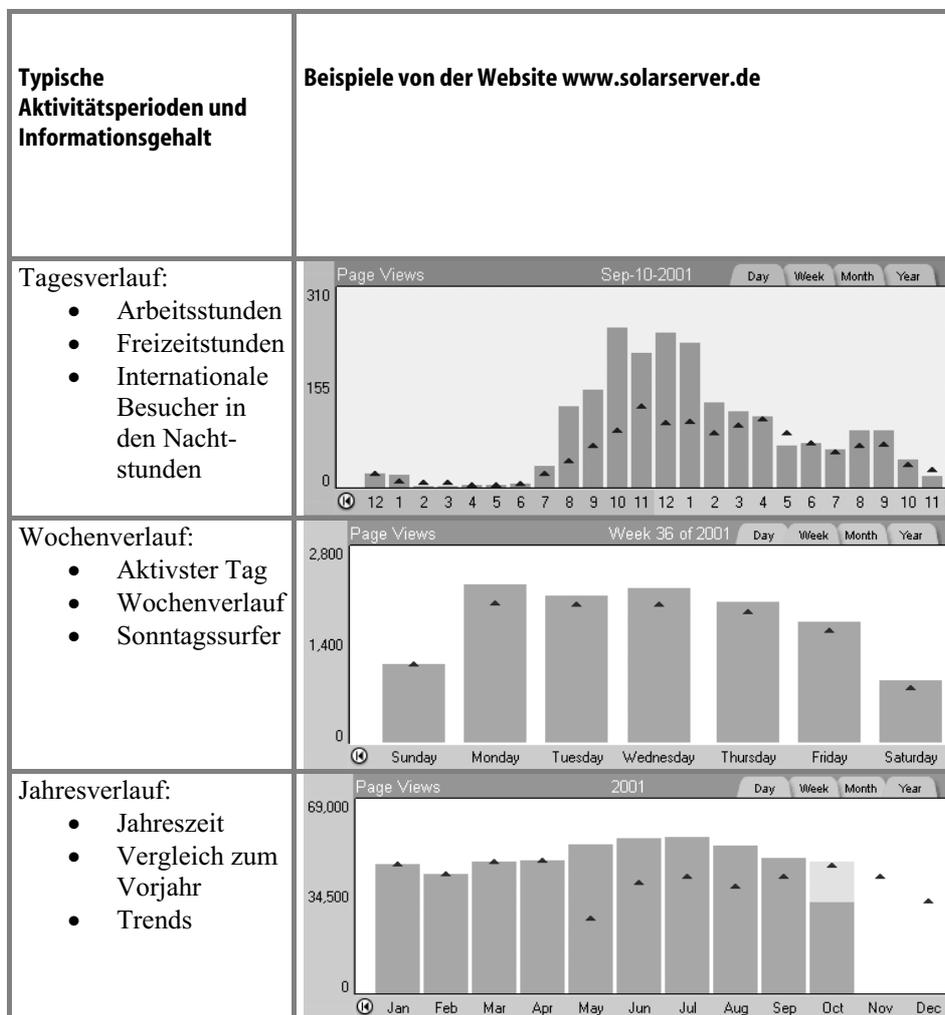


Abb. 1: Beispiel für eine Websiteanalyse

Tagesgang, Wochengang, Ferien

Die zeitlichen Trends

Die Basis einer Websiteanalyse ist die Verfolgung von Trends in den verschiedenen Kennzahlen. Hier zeigen sich drei stark ausgeprägte Perioden. Der Tagesgang, mit einem Minimum in den frühen Morgenstunden und einem Maximum während der Arbeitszeit oder nach Feierabend, der Wochengang, mit großen Unterschieden zwischen den Arbeitstagen, Samstag und Sonntag, und zuletzt unterschiedliche Besucherzahlen aufgrund von Schul- oder Semesterferien und jahreszeitlich unterschiedlichen Interessen.

Wetter und Sportereignisse

Modifikationen der Zyklen

Neben diesen Hauptzyklen unterliegen die einzelnen Zyklen auch noch einer Modifikation auf Grund der übergeordneten Phase. So ist der Tagesgang an einem Samstag völlig anders als an einem Montag. Ein weiterer nicht zu unterschätzender Einflussfaktor stellt das Wetter dar: An schönen Sommertagen beobachtet man wesentlich weniger Besucher als bei Regenwetter. Auch hier können viele andere Detailfaktoren dazukommen, wie regionale Feiertage, konkurrierende Medien bei Sportereignissen. Es bleibt damit die Feststellung, dass es keine zwei gleichen Tage im Internet gibt und daher Trends in den Besucherströmen mindestens über einige Monate beobachtet werden müssen, damit sie als signifikante Veränderung der Nutzerakzeptanz der Website angesehen werden können.

Statistische Ungenauigkeit

Dies gilt natürlich im Besonderen, wenn die erhobenen Datensätze starkem statistischem Rauschen unterliegen, so wird die Besucherzahl einer Website, die nur hundert Besucher pro Tag hat, im Mittel um 10 % schwanken, bei noch kleineren Besucherzahlen ist eine tagesgenaue Auswertung praktisch nutzlos. In der Praxis hat es sich bewährt, einzelne Wochen zu vergleichen, um kurzfristige Entwicklungen zu erkennen und längerfristig die Daten mehrerer Jahresverläufe übereinander zulegen, um das Wachstum einer Site zu beurteilen.

Abweichungen versus Standardwerte

Erkenntnisse aus den Abweichungen

Die kurzfristigen Abweichungen von den Erwartungswerten, ist oft aufschlussreicher als die reinen Zahlen.

Schnellen am Montag Morgen zu Arbeitsbeginn die Besucherzahlen hoch, dann sind es oft die neuen Meldungen, die zu dieser Zeit ins Netz gestellt werden. Umgekehrt steigen die Zugriffszahlen von Wetterseiten gegen Ende der Woche immer an. Beobachtet man während einer Sportveranstaltung im Fernsehen keinen Rückgang der Besucherzahl, kann man daraus schließen, dass die Besuchergruppe, die man anspricht, nicht daran interessiert ist. Wird die Seite am Samstag Vormittag besucht, sind die Besucher zu dieser Zeit nicht im Supermarkt.

Analyse von Marketingmaßnahmen

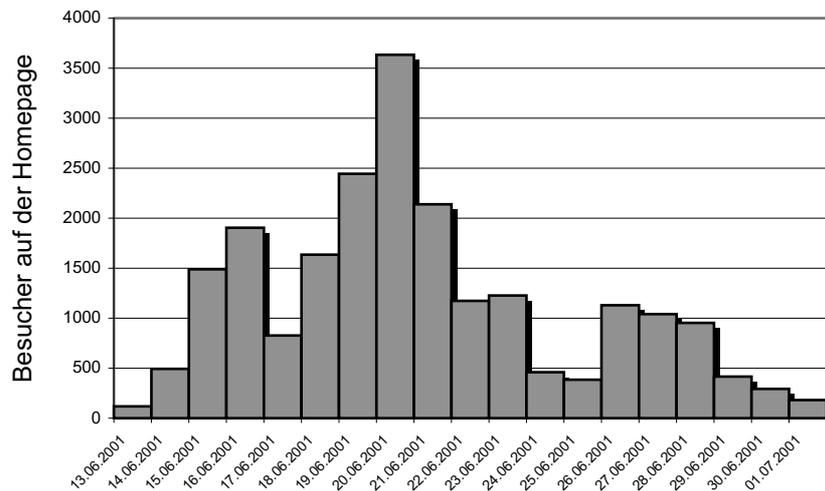


Abb. 1: Wachstum und Rückgang der Seitenaufrufe nach einer Briefaktion mit 60.000 Aussendungen in Deutschland (Tägliche Auswertung).

Ähnlich kann eine begleitende Marketingmaßnahme ausgewertet werden: Nach Start der Aktion wird die Zunahme der Seitenbesuche verfolgt. Bei Versenden von Briefen erzeugt dies eine etwa einwöchige Anomalie in der Besucherzahl, zuerst ein rasches Ansteigen und danach ein Abklingen der Besucherzahlen. Bei der Verwendung von Serien E-Mails, ist der Anstieg wesentlich schärfer, so dass oft die dabei auftretende Serverbe-

lastung das Limit für die Spitze bildet, anschließend sinken die Besucherzahlen wieder rasch ab, und wenn die Kampagne ein Erfolg war, bleibt die Zahl oberhalb der Ausgangswerte.

Die Quelle der Besucherströme

Für den kommerziellen Erfolg einer Website ist eine hohe Besucherzahl unabdingbar, da immer nur ein kleiner Teil der Besucher aktiv Geschäfte tätigen wird und weil die Anzahl der Page-Impressions für den bannerfinanzierten Inhalt benötigt wird.

Suchmaschinen

Optimierung der
Einträge und Anmeldung

In den Logfiles wird heute immer aufgezeichnet, welche Internetseite der Besucher als letztes Dokument betrachtet hat, bevor er die betrachtete Seite aufgerufen hat. Damit ist es den Auswerteprogrammen möglich, exakt zu erkennen, wie viele Besucher über Suchmaschinen, auf Grund von statischen Links und über die direkte Eingabe der Webadresse auf eine Website gelangen. Die Verteilung der verschiedenen Besucherquellen kann anhand dieser Information optimiert werden, so sollte die Verteilung der aktiven Suchmaschinen für die Website etwa der Bedeutung der verschiedenen Suchmaschinen entsprechen. Fehlt ein wichtiges Suchsystem oder ist stark unterrepräsentiert, kann durch Optimierung der Website und Anmeldung eine Verbesserung erreicht werden, die sich dann wiederum aus den Webstatistiken ablesen lässt.

Externe Links

Externe Links als Vorbote
für große Besucherzahlen

Die Bedeutung externer Links auf die eigene Website darf nicht unterschätzt werden, anhand der Statistik kann exakt festgestellt werden, wie viele fremde Seiten auf das eigene Dokument verweisen. Dabei ist besonders auf die Veränderung zu achten. Wächst die Zahl rasch, ist das ein eindeutiger Hinweis auf eine gute Akzeptanz innerhalb der Interessengruppe und ein Vorbote für große Besucherzahlen und damit entsprechende Umsatzperspektiven.

Favoriten

Als dritte Gruppe erscheinen Besucher, die über die Eingabe der Internetadresse direkt auf die entsprechen-

de Website kommen. Dabei ist allerdings zu berücksichtigen, dass nicht alle Programme zwischen der Verwendung der Favoriten und manueller Eingabe unterscheiden können. Trotzdem kann gerade damit der Erfolg einer Werbekampagne in einem anderen Medium (Radio, Fernsehen, Print) aufschlussreich verfolgt werden.

Bedeutung der einzelnen Dokumente

Aufgrund der Logdaten ist es einfach, die Häufigkeit, mit der einzelne Seiten abgerufen werden, auszuwerten. Hier beobachtet man zumeist die klassische 20/80 Regel, dass 20 Prozent der Dokumente 80 Prozent der Seitenbesuche verursachen. Eine genauere Analyse kann allerdings wesentlich mehr aufdecken, da es mindestens drei verschiedene Typen von Webseiten gibt.

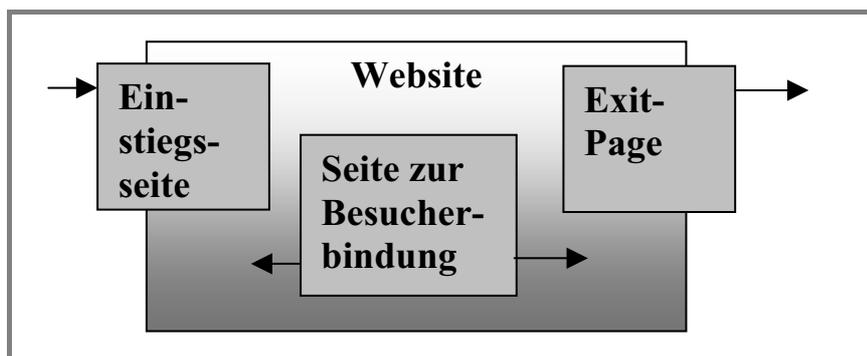


Abb. 2: Die wichtigsten Seitentypen einer Website

Einstiegsseiten

Einstiegsseiten sind jene Dokumente, die einen relativ hohen Anteil von Besuchern haben, die nicht über einen eigenen Link kamen. Dies sind oft Dokumente, die gut in Suchmaschinen auffindbar sind oder einen so guten Inhalt haben, dass sie direkt verlinkt oder gemerkt werden. Einige Websites sind speziell so aufgebaut, dass diese sogenannten Doorway-Pages gezielt neue Besucher anlocken. Auf keinen Fall sollte man diese Dokumente ohne triftigen Grund ändern oder löschen, sondern mit genügend Verknüpfungen zu anderen Dokumenten der Site versehen und selbstverständlich mit der Homepage verknüpfen.

Doorway-Pages

Analyse guter Seiten	<p>Besucherbindung</p> <p>Seiten, die zum Weiterlesen verführen, sind jene Dokumente innerhalb einer Site, bei denen der Besucher nach dem Betrachten ein weiteres Dokument innerhalb der Site aufruft. Diese Seiten zeichnen sich oft durch geschickten Inhalt und passend gesetzte Links auf eigene Seiten aus. Auf jeden Fall sollte man diese Seiten einer genauen Analyse unterziehen, und mit dem gewonnenen Wissen andere Dokumente so gestalten, damit die Besucher lange auf der Site verweilen.</p>
	<p>Exit-Pages</p> <p>Die Exit-Pages sind jene Dokumente, die viele Besucher zum Verlassen der Website veranlassen. Eine dieser Seiten ist klassischerweise die Linkliste, womit der Zweck der Liste erfüllt ist, aber oft sind es Seiten, die fehlerhaft sind, eine ungewöhnlich lange Ladezeit haben und im besten Fall ist es die Vielen-Dank-Seite am Ende eines Kaufprozesses. Durch sorgfältige Verbesserung der ungewünschten Exit-Pages kann die Zahl der Page-Views pro Visit erheblich erhöht werden.</p>
	<p>Besonders aufschlussreiche Dokumente</p> <p>Einige Dokumente innerhalb der Website können zusätzliche Informationen liefern, die über die Besuchersfrequenz hinausgehen.</p>
Die Suchmaschinen leiten	<p>Robotfile</p> <p>Das File robots.txt wird nicht von menschlichen Besuchern, sondern von Suchmaschinen aufgerufen wenn diese eine Website für ihre Datenbank einlesen. Damit gibt die Anzahl der Abrufe ein genaues Bild, wie häufig Suchmaschinen die Website besuchen und welche Suchmaschinen nie den Dokumentenbestand analysieren.</p>
Das Favoriten-Icon	<p>Aktivieren von der Favoritenfunktion</p> <p>Will man wissen, welche Informationen von den Besuchern als nützlich eingeschätzt werden, muss man zählen, wie viele Besucher das entsprechende Dokument zu ihren Favoriten hinzufügen. Dies kann man an der Anzahl der Fileabrufe des Dokuments favicon.ico erkennen, das immer angefragt wird, wenn ein Besucher einen Favoriten setzt. Hat man die Website sorgfältig in verschiedene Verzeichnisse eingeteilt, und in jedem Ver-</p>

zeichnung das File bereitgestellt, kann man daraus differenziert auf die tatsächliche Beliebtheit der Rubriken schließen, dies ist wesentlich aussagekräftiger, als die reine Seitenabrufstatistik!

Ladezeitmessung

Bei großen Webseiten kommt es dazu, dass Besucher vorzeitig den Ladeprozess abbrechen, dies kann man einfach daran erkennen, wenn man auswertet, wie häufig das letzte Bild auf der Webseite relativ zum HTML-Dokument aufgerufen wurde. Sind mehrere Bilder auf der Seite, kann dies sogar bis in das Detail untersucht werden, dazu ist es allerdings notwendig, die Statistiksoftware so zu konfigurieren, dass auch die Bildaufrufe mitanalysiert werden, was häufig nicht in der Standard-einstellung der Fall ist.

Besuchspfade

Analysiert man das Kaufverhalten auf einer Website, stellt man fast immer fest, dass der Käufer zuerst eine Vielzahl von Dokumenten auf der Site besucht, einerseits um das Produkt auszusuchen, aber auch um Vertrauen in den Händler zu gewinnen. Es ist daher hilfreich den Weg des Besuchers zu analysieren und an Hand dieses Wissens gegebenenfalls die Navigation zu optimieren.

Aussagen über
das Kaufverhalten

Hyperlinks

Für jedes Dokument kann exakt bestimmt werden, welche Hyperlinks wie häufig angeklickt wurden. Dies kann man auch mit geeigneter Software visualisieren, damit der Webdesigner die Dokumente weiterentwickeln kann. Es zeigt sich regelmäßig, dass Hyperlinks am Anfang einer Liste wesentlich häufiger angeklickt werden als in der Mitte, aber auch das letzte Element hat eine überdurchschnittliche Trefferwahrscheinlichkeit. Damit kann man die Pfade der Besucher stark beeinflussen, indem geeignete Angebote richtig platziert werden, anschließend kann man die Verhaltensänderungen statistisch belegen.

Erfolgreiche Pfade

Die gesamten Pfade der Besucher sind allerdings statistisch schwer zu analysieren, da bereits wenige, gut ver-

Pogosticking

knüpfte Seiten eine unendliche Zahl von möglichen Wegen erlauben. Hier sollte der Schwerpunkt der Untersuchung in Teilpfaden liegen, etwa dem sogenannten Pogosticking, bei dem ein Besucher immer wieder zwischen zwei Seiten wechselt, weil er sich in der Information nicht zurecht findet. Ein anderer Schwerpunkt kann die Rückverfolgung der Pfade sein, die schnell zu einen erfolgreichen Kauf geführt haben und daraus die Ableitung der notwendigen Änderungen, die es jedem Besucher erlaubt, diesen Pfad leicht zu finden.

Konversionsrate

Das Verhältnis der Besucher, die einen Kauf tätigen, zu allen Besuchern auf der Website wird als die Konversionsrate bezeichnet, eine markante Kennziffer, mit der der Nutzen für das E-Business bestimmt werden kann.

Kritische Betrachtung zu Statistiken

Sinnvolles Filtern der Daten

Statistiken unterliegen immer einer wesentlich größeren Willkür, als man gemeinhin vermutet. Dies liegt zum einem am endlichen Zahlenmaterial, das zur Verfügung steht, an dem „sinnvollen“ Filtern der Daten und in der unterschiedlichen Interpretation der Resultate. Diese Aspekte sollen hier speziell auf Webserverstatistiken bezogen werden.

Signifikanz

Das Zahlenmaterial erscheint auf den ersten Blick außerordentlich umfangreich, wie oben beschrieben, soweit es die Rohdaten betrifft. Völlig anders ist die Situation oft bei Detailfragen, etwa dem Verhalten der Käufer an Sonntagen. Selbst gut besuchte Server haben dann oft weniger als hundert Datensätze, die das spezielle Kriterium erfüllen und man kann allgemein feststellen, dass der damit verbundene Fehler von über 10 Prozent keine verwertbaren Resultate liefert. Das Problem wird sogar noch verschärft, da viele Untersuchungen nicht mit statistisch unabhängigen Größen arbeiten, das bedeutet etwa, wenn ein Besucher an unterschiedlichen Tagen das gleiche Verhalten zeigt, liegt eine abhängige Größe vor und die Signifikanz einer Beobachtung verschlechtert sich dadurch merklich.

Schiefe Statistiken

An vielen Stellen kann es unbemerkt zu schiefen Statistiken kommen. Ist etwa eine Homepage von einem be-

stimmten Browser oder ein Plugin abhängig, darf man sich nicht wundern, dass man viele Besucher mit der gewünschten Systemkonfiguration findet. Hier wurde dann eben nicht berücksichtigt, dass andere Besucher den Server natürlich meiden. Leider sind solche Zusammenhänge oft nicht einfach zu erkennen und es ist dann professionelle Unterstützung bei der Untersuchung der Webserver-Logfiles sinnvoll.

Proxyserver

Eine besondere Veränderung der Ergebnisse entsteht durch den Einsatz von Proxyservern. Das sind Rechner, die Internetseiten zwischenspeichern, weil viele Besucher die Dokumente nicht direkt vom Webserver holen, sondern zwischengespeicherte Dokumente erhalten. Damit ergeben sich zwei Verschiebungen, zum einen wird damit die Zahl der Seitenabrufe unterschätzt, da kein Logfile-Eintrag erfolgt, weiterhin werden häufig besuchte Seiten öfter in Proxyservern zwischengespeichert und damit die relative Häufigkeit der Seitenabrufe verfälscht. Dies kann besonders bei stark frequentierten Websites einen erheblichen Einfluss haben. Es gibt technische Verfahren, die Zwischenspeicherung zu unterbinden, dabei entstehen aber zusätzliche Probleme, die dieses Vorgehen nicht rechtfertigen.

Verfälschung
der Ergebnisse

Pseudobesucher

Die Rohdaten müssen vor der Verarbeitung nach mehreren Kriterien gefiltert werden, damit eine aussagekräftige Beobachtung entsteht. Für eine Untersuchung der Besucherzahlen sollten alle Besuche auf dem Server gefiltert werden, die nicht der Zielgruppe entsprechen. Dies sind in erster Linie die Suchmaschinen, genommen die Suchrobots und Archivsysteme, alle am Aufbau der Website Beteiligten und zumeist auch die Mitarbeiter des eigenen Unternehmens, die im Unternehmen dorthin surfen. Leider wird dies oft unterlassen, weil danach die Zahlen nicht mehr so „schön“ aussehen. Selbstverständlich sind alle Abfragen, die eine Fehlermeldung verursachen, nicht als erfolgreiche Seitenabrufe zu zählen, was oft übersehen wird. Insbesondere kann die sinnvolle Reduktion dieser Fehler ein scheinbares Absinken der Seitenaufrufe verursachen, was natürlich eine Fehlinterpretation ist.

Die Webstatistik analysieren

Interpretation

Die Interpretation der ausgewerteten Daten gehört zum schwierigsten Teil der Analyse. Zunächst ist das Wachstum der Besucherzahl nur dann ein erfreuliches Resultat, wenn dieses Wachstum nicht nur durch die schlichte Zunahme der Internetnutzer zustande kommt, sondern durch die hohe Qualität der Website. Wesentlich schwieriger sind oft die Gründe zu finden, warum bestimmte Dokumente gut gefunden werden oder warum Besucher eine Bestellung in einer bestimmten Phase abbrechen. Hier hilft oft der Abgleich der Information aus der Webstatistik mit dem Wissen eines Kommunikationsdesigners. Dieser kennt Eigenarten beim Nutzerverhalten und kann die Ergebnisse auch in Bezug zu anderen Websites setzen. So darf man etwa nicht erschrecken, wenn man erfährt, dass über 80 % eine Kaufvorbereitung abbrechen, wenn man weiß, dass es sich hier um einen internetweiten Durchschnitt handelt.

Die Fragestellung ist maßgebend

Perspektiven der Logfile-Analyse

Welche Größen in der Logfile-Analyse die höchste Aussagekraft haben, hängt von der Fragestellung ab. Betrachtet man das Benutzerverhalten aus der Perspektive des Informationssuchenden, der seine Kaufentscheidung vom realen und emotionalen Gehalt der präsentierten Information abhängig macht, ist neben der Qualität der Information immer auch die Quantität von entscheidender Bedeutung. Dies ist der Grund, warum für größere Anzeigen, längere Werbespots und aufwendige Schaufenster viel Geld gezahlt wird. Im Internet ist die Menge der Information mit einer einfachen Größe zu ermitteln, es sind die übertragenen Bytes, die den Nutzer erreichen.

Quantität der Information

Ausführliche Texte und qualitativ hochwertige Bilder sind jene Daten, die längerfristig die Entscheidung für ein Produkt prägen. Dabei ist heute allerdings die Datenrate bei der Internetnutzung mit Modem oder ISDN so gering, dass ungewöhnlich scharfe Kompromisse zugunsten der Übertragungsgeschwindigkeit getroffen werden müssen. Erst die Einführung der DSL-Technik entschärft diese Fragestellung und fordert den Webser-

verbetreiber genau zu beobachten, wie er die Datenmenge, die den Nutzer erreicht, vergrößern kann.

Umsatzkennziffer für Webseiten

Setzt man den Umsatz einer Website zur Menge der versendeten Daten in Beziehung, findet man, dass jedes Megabyte etwa einen Euro Umsatz generiert. Diese Kennziffer sollte man längerfristig im Auge haben, denn sie ist auch dann aussagekräftig, wenn der abgebrochene Kaufvorgang später wieder aufgenommen wird, dann kommen die bereits versendeten Daten dem Umsatz zugute, der Kunde hat verstanden, was geboten wird. Längerfristig werden die Kosten für den Empfang der Daten sinken und damit werden die Kunden kostengünstiger im Internet einkaufen.

Jedes Megabyte generiert
einen Euro Umsatz

Mangelndes Interesse

Natürlich können für den Internetauftritt auch sehr komplexe Logfile-Analysen erstellt werden und diese sinnvoll für die Verbesserung der Website verwendet werden, aber solange weniger als 40 % der Entscheider überhaupt wissen, wie viele Besucher ihre Website hat, solange sind feinsinnige Statistiken nur für Internetexperten nützlich.

Literatur

Heindl Eduard; Der Webmaster, 3. aktualisierte Auflage, , ISBN 3-8273-1853-X, 330 Seiten, Addison-Wesley, München 2001
Heindl, E., Bücking, J., Emmert, U.; Der IT-Sicherheitsexperte, ISBN 3-8273-1840-8;304 Seiten, Addison-Wesley, München 2001

Geeignete Internetadressen

<http://www.ivw.de/>
Neutrale Ergebnisse von Besucherdaten
<http://www.wemf.ch/>
Internet Medienstudien
<http://www.datango.de/>
Marktstudie "Statusbericht Websites 2001"
<http://www.exody.net/>
Hersteller der nützliche Analysesoftware websuxess
<http://www.heindl.de/webkolumne/index.html>
Betrachtungen des Autors zum Web