

Dirk Fox

Webtracking

Hintergrund

Die Analyse der Nutzung von Internet-Angeboten (Webtracking) spielt für die Weiterentwicklung des Aufbaus und der Gestaltung von Webseiten eine wichtige Rolle. Nutzungsanalysen geben nicht nur Auskunft über besonders stark und selten frequentierte Seiten eines Internet-Auftritts, sondern auch darüber, wie der Besucher die Seite gefunden hat (z. B. über eine Suchmaschine, durch direkte Eingabe der URL, durch Aufruf eines individualisierten Links einer Marketing-E-Mail oder einen Klick auf eine Online-Anzeige), welche auf einer Seite platzierten Links angeklickt werden, wie lange Besucher auf einer Seite verweilen, wie viele Seiten sie aufrufen und an welcher Stelle des Kaufprozesses in einem Webshop Bestellungen besonders häufig abgebrochen werden (*Bounce Rate*). Auch die wichtigsten „Pfade“ eines Webseitenbesuchs lassen sich bestimmen, die Herkunft der Besucher (durch Geolokalisierung der IP-Adresse¹) und, sofern Besucher „wiedererkannt“ werden können, auch die Häufigkeit von Seitenbesuchen. Analysen dieser Art ermöglichen auch eine Erfolgs- und Reichweitemessung von Marketing- oder Werbemaßnahmen für Internet-Angebote.

Methoden

Die unterschiedlichen Methoden zur Analyse von Webseitenbesuchen lassen sich in zwei Hauptklassen unterteilen: Server-basierte Verfahren, die im wesentlichen die auf dem Web-Server angelegten Protokolldaten (Logdateien) auswerten, und Client-basierte Verfahren, die über den Aufruf von nicht sichtbaren „Ein-Pixel-Bildern“ Zugriffsdaten erzeugen oder durch aktive Elemente im Quellcode der Seite (z. B. in JavaScript) Informationen des Besuchersystems (Browser-, Rechner- und Bildschirmtyp, Mausbewegungen, Tastatureingaben etc.) erheben. Mit der Speicherung einfacher Co-

kies lassen sich Besucher wieder erkennen, sogar dann, wenn sie zwischendurch den Internet-Zugang wechseln (Laptop in der Firma, im Hotel, zu Hause).

Während Server-basierte Analysen ihre Stärke z. B. bei der schnellen Identifikation von „toten Links“ und anderen Fehlerseiten (Tippfehler bei der URL-Eingabe o. ä.) besitzen, erlauben sie keine zuverlässige Besucher-Identifikation, da die IP-Adresse eines Seitenaufrufs kein hinreichend genaues Merkmal ist – in Unternehmen sorgt *Network Address Translation* (NAT) dafür, dass der besuchte Webserver nur die IP-Adresse der Unternehmensfirewall oder eines Web-Proxy-Servers sieht. Auch wird der Besucher nicht wiedererkannt, wenn die IP-Adresse sich z. B. durch einen Wechsel des Internet-Zugangs ändert.

Client-basierte Analysen hingegen erlauben eine sehr umfangreiche Analyse und die Wiedererkennung des Besuchers, können jedoch vom Nutzer unterbunden werden, z. B. durch die Sperrung von Cookies, Pixel-Bildern oder der Ausführung aktiven Codes.

Auswertungen

Viele Unternehmen nutzen die Dienstleistungsangebote Dritter zur Webseitenanalyse. Server-basierte Analysen werden in der Regel von Webseiten-Hostern angeboten, Client-basierte Analysen dagegen von spezialisierten Drittanbietern, deren Code in den Quelltext der Webseiten integriert werden muss. Dabei werden die Nutzungsdaten an den Dienstleister übermittelt, der die Daten in Form von statistischen Auswertungen und grafischen Übersichten aufbereitet und dem Unternehmen zur Verfügung stellt.

Aber es gibt auch deutlich weiter gehende Nutzungsmöglichkeiten. So existieren Dienstleistungsangebote, die die erhobenen Besucherdaten mit Daten aus dem *Customer Relationship Management* (CRM), also Einträgen in der Kundendatenbank verknüpfen: Besucht ein Kunde die Webseite, werden ihm Kaufempfehlungen oder Angebote passend zu früheren Käufen präsentiert, oder es wird im

CRM vermerkt, dass er ein Angebot der Webseite genauer betrachtet, und daraufhin z. B. ein telefonischer Vertriebskontakt mit ihm aufgenommen.

Probleme und Gefahren

Zahlreiche der heute in großer Zahl praktizierten Auswertungen sind – sofern ungekürzte und damit personenbeziehbare IP-Adressen verwendet werden oder Profilerstellungen durch Verknüpfung mit CRM-Daten erfolgen – ohne Einwilligung des Betroffenen (die in der Regel nicht vorliegt) unzulässig.²

Die Auswertung durch Dienstleister, insbesondere aus Staaten mit bekanntem niedrigem Datenschutzniveau, birgt zudem weitere Gefahren, da sich die Nutzungsdaten grundsätzlich auch über Firmengrenzen hinweg verknüpfen und auswerten lassen.

So dürfte beim Marktführer Google inzwischen eine sehr umfassende Profilbildung möglich sein, die nicht nur die Verfolgung aller Seitenbesuche von mit Google Analytics verbundenen Websites erlaubt, sondern auch die Anreicherung um das Nutzungsverhalten bei anderen Google-Diensten, wie die in Google Search gewählten Suchbegriffe, die aus den Suchergebnissen angewählten Seiten, die mit Google Maps angezeigten Adressen und berechneten Fahrrouten, die über Google News gelesenen Nachrichten, die in YouTube betrachteten Videos, die in Picasa hochgeladenen Privatfotos und, bei Vorliegen eines Google Mail Accounts, auch die versandten und empfangenen E-Mails, ggf. einschließlich älterer, über das „Umzugstool“ hinzu gespeicherten Nachrichten.

Die zunehmende Nutzung von Internet-Diensten auf Smartphones erlaubt die Erhebung eines weiteren interessanten Datums: des genauen Aufenthaltsorts. Damit ist es nicht nur möglich, gezielt Anzeigen von lokalen Anbietern einzublenden, sondern das Besucherprofil durch Verdichtung der Aufenthaltswahlungen um Bewegungsprofile (und ggf. sogar –muster) zu ergänzen.

¹ Das gelingt sogar mit gekürzten IP-Adressen, siehe Kühn, DuD 12/2009, S. 747-751.

² Siehe Knopp, in diesem Heft.