

Controlling Kennzahlen für Web-Sites

Vom Logfile zum Kunden

Seminar Business Intelligence WS 2001|02
Prof. Dr. Kurt Bauknecht
Institut für Informatik und Wirtschaftsinformatik
BWZ, Universität Wien
<http://www.dke.univie.ac.at/>

Anna Saller

9604708

asaller@hotmail.com

Armin Rudelstorfer

9702326

armin@rudelstorfer.net

Intro

Datenerhebung

Ebene 1: Digit Counter

Ebene 2: Logfile

Ebene 3: Wiedererkennung der Besucher – Möglichkeiten und Probleme durch Userprofile

Ebene 4: Personalisierung pur: Login

Weitere Formen der Datenerhebung

Definition der Messgrößen

Logfileanalyse

Kenngrößen aus Rohdaten

Kombination von Rohdaten

Kennzahlen mit Zeitbezug

Kenngrößen mit Einbezug externer Daten

Neue Ansätze der Erfolgskontrolle

Instrumente der Erfolgskontrolle

Web-Controlling

2002 oder 1984

Big Brother Inside

Doubleclicks Griff in die Keksdose

Literaturverzeichnis

Intro

Im Electronic Commerce können die Kunden so gut wie nie zuvor beobachtet werden. Das Internet liefert dem Betreiber mit Hilfe von Logfiles, E-Mails, Web-Formularen und Netzüberwachung eine Fülle von Daten die den Marktforschern und den Unternehmen zur Verfügung stehen. Kann man von der Erfüllung aller Marketing-Wünsche sprechen oder stellt die Datenflut einen Albtraum dar? Die Fülle der möglichen Daten verlangt intelligente Analysen, ehe wirklich die Web-Präsenz beurteilt werden kann. Erst die hohe Schule des Web-Controlling öffnet den Weg in nachhaltig profitable Gefilde. Das E-Commerce-Controlling ist zu einem entscheidenden Wettbewerbsfaktor geworden. Ziel ist die Steigerung der Effektivität bzw. der Effizienz der Marketingmaßnahmen innerhalb und außerhalb von E-Business-Anwendungen.

Wie für andere Werbeträger benötigt man für das Internet natürlich auch Media Daten. Primär wollen Werbekunden noch sehr oft wissen, wie viele Internet-Nutzer es gibt und ob ihre Zielgruppe im Internet vertreten ist. Demographische Daten der Nutzer sind vom höchstem Interesse. Unternehmensexterne Daten werden in die Analyse einer Web Site integriert. Das E-Commerce-Controlling beschäftigt sich also mit der Sammlung und Auswertung möglichst umfangreicher Datenbestände, die dann zu aussagekräftigen Kennzahlen verdichtet werden.

Kaum ein Händler eröffnet eine neue Filiale und bewertet den Erfolg allein danach, wie viele Menschen den Laden betreten. Die Zahl der Besuche oder der betrachteten Seiten gibt erste Anhaltspunkte, aber für den Return on Investment ist die Frage entscheidend, wie viele Transaktionen erfolgreich abgewickelt werden. Antworten werden gesucht: Welche Seiten führen tatsächlich zu Verkäufen? Welche Anzeigen oder Links bringen qualifizierte Besucher zur Website? Bleibt es beim einmaligen Kontakt oder kann die

elektronische Filiale die Kunden dauerhaft binden? Was sind eigentlich die wirklichen Attraktionen der Website für die Besucher?

Die Fragen sind nicht neu, Marktforscher stellen sie nicht erst, seit IBM das „e“ vor das „Business“ stellte. Durch neue Instrumente ergeben sich neue Strategien und diese verlangen neue Daten und Kennzahlen.

Das Web unterscheidet sich durch traditionelle Medien durch seine technischen Strukturen und seine Geschichte. Keines der bisherigen Medien liefert eine Sammlung an Funktionen wie das Internet. Bisher war bidirektionale Kommunikation oder gar Interaktivität nur rudimentär möglich. Eine Fülle an Daten, die für Controllingmaßnahmen genutzt werden können, liefert das Internet mittels der eingesetzten Protokolle und Software automatisch. Auch wenn dies verlockend klingt, darf die Entwicklungsgeschichte nicht außer Acht gelassen werden. Das Internet ist ein Container für verschiedene Anwendungen, welche zum Großteil von Wissenschaftlern für Wissenschaftlern entwickelt wurden. Die meisten der heutigen Entwicklungen basieren noch auf der Steinzeit des Internets und wurden nie für heutige Anwendungen konzipiert. Eine Fülle an Daten entsteht automatisch, ohne jedes Zutun von Anbieter oder Konsument. Grundsätzlich wird jede Bewegung des Benutzers protokolliert. Ein Traum für Controller und Direct-Marketer? Wie reagieren Datenschützer und welche Strategie wird sich als nachhaltig erweisen?

Datenerhebung

Ein werbefinanziertes Online-Angebot braucht, wie auch alle anderen werbefinanzierten Medien, Verbreitungs- und Kontaktmessungen. Gleich, ob nun im Umfeld eines redaktionell gestalteten Werbeträgers oder durch eine eigenständige WWW-Präsenz geworben wird, Kontaktmessung ist notwendig und auch möglich.

Bei Rundfunk und Presse muss man dazu ein eigenes Meßsystem aufbauen. Beim Internet ist das Medium selbst das Meßsystem.

Online-Meßsysteme beruhen auf der Tatsache, dass jede Person, die sich eine bestimmte Seite eines Online-Anbieters ansehen möchte, diese individuell abrufen muss.

Da Logfiles nur bedingt Aufschluss über Motive und Wünsche der Kunden und Besucher geben, gilt es andere Wege zu finden, um den Surfer besser kennen zu lernen. Aufgrund technischer Einschränkungen des zugrundeliegenden Systems (Protokolle, Server- und Clientsoftware,..) bleiben eine Fülle an Daten undeutbarer Datenmüll. Woran dies liegt und wie dem Abhilfe geboten werden kann, soll in diesem Kapitel behandelt werden.

Grundlegend muss für anspruchsvollere Datensammlung zwischen reaktiven und nonreaktiven Verfahren unterschieden werden. Reaktiv bedeutet ein Mitwirken des Surfers (Eingabe von Daten,...), als nonreaktiv werden die Daten bezeichnet, die ohnehin entstehen, wie zum Beispiel die Einträge in Logfiles.

Zu den nonreaktiven Verfahren zählen zum Beispiel:

- Logfiles
- Cookies
- Session ID
- Auswertung von Shop/DBMS

Reaktive Methoden sind:

- Formulare
- Aktiv geführte Profile
- Feedbacksysteme wie Mail oder Votingsysteme
- Communitystrategien
-

Eine einfache Variante zur Gewinnung von Daten des Users ist es, einfach danach zu fragen. Auf diese Weise ist es nicht immer möglich die Daten, die den klassischen Direkt-Marketer interessieren (vom Namen bis zum Einkommen) zu erheben, durch effizienten Einsatz von Response-Elementen können jedoch unpersonalisierte qualitative Daten gewonnen werden. In Verbindung mit Personalisierungsmechanismen können diese auch zu komplexeren Datenprofilen entwickelt werden.

Als Beispiele hierfür dienen Voting Systeme, Diskussions-Boards oder einfache Mailbuttons. Sie zeichnen sich dadurch aus, dass die Besucher der Page aktiv Daten hinterlassen, ohne dass er hierfür seine Identität preisgibt. Identität und persönliche Daten sind ein wertvolles Gut, dies wissen Kunden und Anbieter. Soll der Kunde zusätzlich zu den Spuren, die er hinterlässt auch noch persönliche Daten hinterlassen, ist Kreativität gefragt.

Daten können auf vier Ebenen gewonnen werden:

Ebene 1: Digit Counter

Das Verfahren gibt an, wie oft eine bestimmte Seite abgerufen wurde. Für diesen Vorgang muss noch keine Information über den Nutzer erhoben werden. Beispielsweise findet man auf vielen Seiten des Web Einträge wie: „Sie sind der 002674. Besucher dieser Seite seit dem 1.1.1996. Mit diesem Verfahren werden Hits auf HTML-Seiten gemessen. Oft wird der Wert auch als PageClick bezeichnet. Dabei handelt es sich um alle zu den Browsern der Nutzer übertragenen HTML-Dokumente, andere Dateien, die in diese HTML-Dokumente eingebunden sind, werden nicht berücksichtigt.¹

Ebene 2: Logfile

Da der Server aber wissen muss, wohin die Seite versendet werden soll, ist eine Bezugsadresse anzugeben, die ebenfalls dokumentiert werden kann. Bei dieser Bezugsadresse handelt es sich in der Regel nicht um die Adresse des Users, sondern um die Anschrift des Rechners, über den der Benutzer den Zugang zum Internet herstellt. Er wählt zum Beispiel einen Rechner einer Universität an, über diesen wird dann die Verbindung hergestellt und die Anschrift des Universitätsrechners und einige weitere Daten werden in einem sogenannten Logfile (oder Serverlog) erfasst.

Logfiles sind automatisch erstellte Dateien, die alle Anfragen an einen Webserver und deren Ergebnis in einer Protokolldatei aufzeichnet. Diese enthält zu jeder Anfrage die IP-Nummer des zugreifenden Rechners, Datum und Uhrzeit des Zugriffs, die angeforderten Dateien und das Ergebnis der Übertragung. Auch die zuvor besuchte Adresse sowie Betriebssystem und Browser des Benutzers können erfasst werden. Logfiles sind damit die Basis für eine Bewertung und Optimierung einer Website in Bezug auf Effizienz und Nutzerverhalten.²

¹ Stephan, Werner: *Marketing Instrument Internet*, S. 176

² Ahrens & Behrent Agentur für Kommunikation GmbH, <http://www.a-b-logfile.de/>

Bei jeder Anfrage eines Clients (meist ein WWW-Browser) an einen Webserver wird ein Eintrag in den Logfiles generiert. Oder genauer gesagt : Jeder "Hit" erzeugt eine Zeile in den Logfiles. Insgesamt können im Logfile nur solchen Daten erfasst und ausgewertet werden, die das HTTP-Protokoll übermittelt. Im Folgenden werden die gebräuchlichsten Logfile-Arten und -Formate aufgeführt:

Access Log

Es ist das grundlegende Logfile und protokolliert, wer wann welche Datei abgerufen hat. Die meisten Server verwenden dafür das „Common Logfile Format“ (CLF). Das sieht so aus:

```
host_des_nutzers rfcname logname [datum] "METHODE /datei.htm" Statuscode  
Übertragene_Bytes
```

Die meisten Server legen zusätzlich auch noch folgende Log-Dateien an³:

Referer-Log

Ein Referer-Log protokolliert, von welcher Seite aus der Zugriff erfolgte.

Error-Log

Ein Error-Log notiert, welche Zugriffe scheiterten und warum („File does not exist“).

Agent-Log

Ein Agent-Log führt Buch über die Browser, mit denen die Leser die Site besuchen. Zum Beispiel: **Mozilla/3.0 (Win95; I)**

Kombi-Log

Das sogenannte „NCSA combined log“ verknüpft Access-, Agent- und Referer-Log zu einem einzigen Eintrag.

³ Vogt: *Erfolgreiche Präsenz im Internet*, S. 227

Hier ist eine Beispielzeile aus einem Server-Logfile angeführt⁴ :

```
62.158.8.73 - - [04/Mar/2001:12:41:59 +0100] "GET / HTTP/1.1" 200 4425
www.usability-test.net "http://ideenreich.com/surftipps-allerlei.shtml" "Mozilla/4.0
(compatible; MSIE 5.5; Windows 98; Win 9x 4.90)"
```

IP Adresse des Clients	62.158.8.73
Datum und Zeit der Anfrage	04/Mar/2001:12:41:59 +0100
Das verwendete Übertragungsprotokoll und die Methode	GET / HTTP/1.1
HTTP Status Code	200
Anzahl der übermittelten Bytes	4425
Refer (vorherige URL)	http://ideenreich.com/ surftipps-allerlei.shtml
Browser und Betriebssystem	" Mozilla/4.0 (compatible; MSIE 5.5; Windows 98; Win 9x 4.90)

Die IP-Adresse lässt mehr oder weniger genaue Rückschlüsse auf die Herkunft der Nutzer (Firma bzw. Zugangsprovider) zu, und auch die vom Besucher verwendete Browsersoftware wird ermittelt. Allerdings werden von vielen Providern IP-Adressen dynamisch vergeben, d.h. dass ein und derselbe Nutzer bei einem erneuten Besuch mit einer anderen IP-Adresse auf die Web Site zugreift⁵.

Wenn der Nutzer sich mit einem Nutzernamen und Passwort eingeloggt hat, erscheint der Nutzernamen als logname. Wenn beides nicht angegeben wird – was auch im Beispiel der Fall ist – erscheinen nur zwei Bindestriche: "- -". Das Datum wird nach dem Muster [Tag/ Monat/ Jahr/ Stunde/ Minute/ Sekunde] und der Abweichung von der Greenwich Standardzeit angegeben. Als Anfragemethode können GET, PUT, POST und HEAD vorkommen. Die verbreitetste ist GET für die Abfrage des Dokuments. Der Statuscode verrät, ob die Anfrage erfolgreich war.

⁴ Alp Uçkan, <http://magazin.usability-test.net/artikel/logfiles2.htm#oben>

⁵ Bossat, Haite: *Internet für Unternehmer*, S. 200

200 bedeutet dabei: „Dokument wurde erfolgreich abgerufen.“⁶ Weiters werden die übertragenen Bytes angegeben. Der Referer gibt die vorherige Seite an. Schlussendlich erhält man noch Auskunft über den Browser und das Betriebssystem des Rechners.

Fast jeder Provider bietet außer den Rohdaten auch ausgewertete Statistiken an. Nachfolgend werden zwei Beispiele angeführt, wie diese dargestellt werden können. Auf die Interpretation solcher Statistiken wird unter dem Kapitel „Logfileanalyse“ näher eingegangen.

Tageszeit Übersicht

(**Andere Statistiken:** Anfang: Monatsbericht: Wochenbericht: Tagesbericht: Wochentage Übersicht: Tageszeit Übersicht: Domänen Bericht:

Organisationsbericht: Server Verweisbericht: Suchanfrage Bericht: Suchbegriff Bericht: Browser Übersicht: Betriebssystem Bericht)

Jede Einheit (≡) entspricht 40 Seitenanfragen oder einem Teil davon.

Stunde: #Anf.: Seit.:

-----: -----: -----:

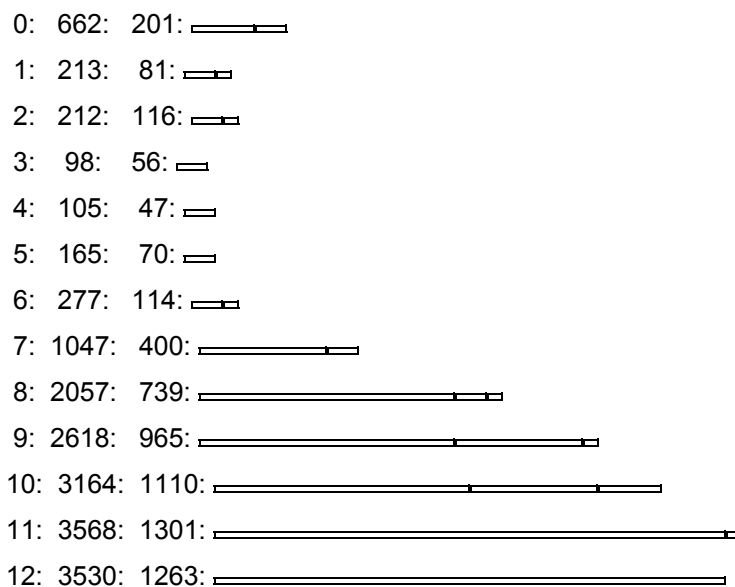


Abbildung 1: Auszug aus der Statistik „Tageszeit Übersicht“

Quelle: Serverstatistiken für Norbert Heydorn: <http://www.n-heydorn.de/report.html>

⁶ Vogt: *Erfolgreiche Präsenz im Internet*, S. 226

Einige WWW-Hoster wie Puretec oder Strato bieten diese Analysen auch im WWW-Format. Der "http-analyzer", den Strato einsetzt, kategorisiert seine Analyse wie in Abbildung 2 :

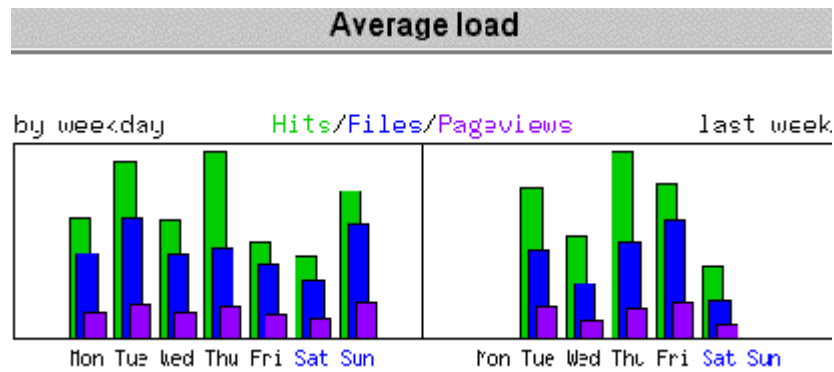


Abbildung 2: Hits nach Wochentagen

Quelle: Alp Uçkan, <http://magazin.usability-test.net/artikel/logfiles2.htm#oben>

Die Zusammenfassung der Hits nach Wochentagen kann Aufschluss darüber geben, ob die Domain eher zu Werktagen oder auch an Feiertagen (und damit überwiegend privat) besucht wird.

Ebene 3: Wiedererkennung der Besucher – Möglichkeiten und Probleme durch Userprofile

Logfiles liefern eine Fülle an Informationen über jede einzelne Bewegung der Surfer, einen Rückschluss auf den einzelnen Besucher lassen sie jedoch nicht zu. Um den Besucher bei mehrmaligen Besuchen wiederzuerkennen wurden einige Verfahren entwickelt. Zuerst aber ein kurzer Blick auf die technischen Probleme hinter den zugrundeliegenden Systemen.

Ein Surfer wird von einem Webserver nicht als solcher erkannt, der Server erhält nur Informationen über den Browser, den Computer und das Netz (den Provider). All diese Informationen sind interessant, wertvoller jedoch sind Informationen über den Menschen hinter dem Computer.

Der Browser liefert mittels dem http-Protokoll Daten über die Browserversion, das Betriebssystem oder die Sprache. Mittels Javascript können weitere Informationen, wie die Auflösung des Monitors oder die Größe des Browserfensters gewonnen werden. Unter <http://privacy.net/analyze/> oder <http://www.gemal.dk/browserspy/> wird gezeigt, welche Daten der Server und die darauf laufende Software ermitteln können. Diese Daten erlauben eine grobe Demografie und sind primär für Webdesigner interessant.

Für weitergehende Datengewinnung müssen folgende Probleme gelöst werden:

Zuerst muss der Computer, beziehungsweise der Browser des Benutzers wiedererkannt werden. Die IP-Adresse bietet hierfür einen Ansatz, ist aber sehr fehleranfällig. Aufgrund der Knappheit der IP-Adressen werden diese dynamisch vergeben, oder es identifizieren sich verschiedene Computer parallel mit der selben Adresse. Dynamische Vergabe bedeutet, dass je nach Internetzugang einem Benutzer nur temporär eine Adresse zugewiesen wird (beispielsweise bei Wählleitungszugängen wie mittels Analog- oder ADSL-Modem, aber auch bei Anmeldung mittels DHCP, wie bei verschiedenen Kabelzugängen) und diese daher nicht an eine Person oder einen Computer gebunden ist. Andererseits wird ein beträchtlicher Teil des Datenverkehrs über Proxy-Server, Firewalls und Router abgewickelt. Diese Systeme verbergen

mehrere Computer und IP-Adressen hinter der Adresse des Proxys oder der Firewall. In solchen Fällen ist es auch nicht möglich die verschiedenen Besucher aus dem selben Teilnetz zu unterscheiden. Studenten der Universität Wien verursachen bei aktiviertem Proxy einen Eintrag im Logfile mit der IP-Adresse 131.130.1.12 (bzw. 131.130.1.135 bei Zugang über Chello). Zur Unterscheidung der einzelnen Studenten müssen weiterführende Maßnahmen getroffen werden.

Session ID

Eine Session ID⁷ ist eine individuell generierte Zahl, die vom Webserver einem Besucher zugewiesen wird. Bei jeder Bewegung auf der Webpage wird diese ID an den Server übermittelt und dieser erkennt somit den Besucher als solchen. Das Verfahren der Session ID dient nicht der Sammlung von Profilen, sondern lediglich dazu, die zuvor beschriebenen, durch Proxys und Firewalls verursachten Probleme zu lösen. Wie der Name verrät, bezieht sich die Session ID nur auf einzelne Sessions und kann daher verwendet werden um Visits zu erkennen. Da bei jedem Besuch eine neue Session ID generiert wird, ist es nicht möglich damit Stammkunden zu erkennen.

Technisch funktioniert dieses System, indem jedem Link des Besuchers die individuelle Session ID zugefügt wird. Bei jeder übermittelten Seite ist die ID an der langen Ziffernfolge in der Adressleiste zu sehen. Bei jedem Link wird diese ID dem Server wieder rückgeliefert und der Link kann somit der zuvor gesandten Page zugeordnet werden. Somit wird ein anonymer Pfad durch die Page erkennbar. Eine Session ist dadurch aber beschränkt auf einen ununterbrochenen Besuch eines Surfers mit einem Browser. Besucht der Surfer kurzzeitig eine andere Seite oder stürzt der Browser ab, kann die Fortsetzung der Session nicht erkannt werden. Abhilfe gegen dieses Problem schafft die Kombination mit Cookies.

Weitere Quellen zum Thema: Entwurf des Standards für Session-Ids

<http://www.w3.org/TR/WD-session-id.html>

⁷ C't 18/2000 - Viele Gesichter Personalisierte Websites stellen sich auf Besucher ein

Cookies

Keines der angewandten Verfahren erhitzt die Gemüter so sehr wie Cookies⁸. Der unscheinbare Name beinhaltet ein Verfahren, das es ermöglicht wiederkehrende Besuche zu erkennen.

Ein Cookie ist lediglich eine kleine Textdatei, die vom Server auf dem Clientbrowser hinterlegt wird. In dieser Datei steht nicht mehr als eine Nummer, die den Browser identifiziert. Dadurch ist es möglich den Browser bei neuerlichen Besuchen zu identifizieren und ein Profil anzulegen. Grundsätzlich ist diese Methode anonym, geriet aber ins Kreuzfeuer der Datenschützer, da einzelne Unternehmen bedenkliche Profile anlegten. Die Problematik der Cookies wird im Kapitel Datenschutz ausführlicher dargelegt. Hier soll vorerst nur die Funktion dargestellt werden.

Mit einem Cookie ist es damit erstmals möglich Besucher wiederzuerkennen, jedoch mit der Einschränkung, dass Cookies nur auf Wunsch des Besucher personalisiert werden können und ebenfalls auf einen Browser beschränkt sind. Der Surfer kann jederzeit seine Cookies löschen und das Anlegen von Profilen somit unterbinden. Durch die starke Kritik an Cookies wurden in allen ernstzunehmenden Browsern Möglichkeiten zur Ablehnung von Cookies implementiert. Cookies bieten ebenso wie Session IDs keine Möglichkeit den Besucher zu erkennen. Es kann lediglich der Browser wiedererkannt werden. Ein Browser ist aber nicht zwangsläufig einem Benutzer zugeordnet. Teilen sich mehrere Benutzer einen Computer, so teilen sie sich eventuell auch den Browser. Bei einer Neuinstallation von Browser oder Betriebssystem, gehen ebenfalls sämtliche Cookies verloren.

Die Kernproblematik bleibt also bestehen. Abhilfe können serviceorientierte Loginverfahren schaffen.

⁸ C't 18/2000 - Viele Gesichter Personalisierte Websites stellen sich auf Besucher ein

Ebene 4: Personalisierung pur: Login

Verfahren bei denen sich ein Besucher einer Seite oder eines Webservices mit Benutzernamen und Kennwort anmelden kann, fallen in den Bereich der reaktiven Systeme. Die Anmeldung einer Person bedeutet, dass mit sehr hoher Wahrscheinlichkeit auch nur diese Person die Services unter dem Login nutzt. Auch wenn dieses Verfahren bei Anwendungen wie Telebanking selbstverständlich ist, muss den Nutzern von durchaus anonym nutzbaren Services vorerst ein schmackhafter Grund für die Preisgabe ihrer Identität geboten werden. Da das Internet als anonymes Medium entstanden ist und gerade im deutschsprachigen Raum Anonymität hoch geschätzt wird und Datenschutz als achtenswertes Recht verstanden wird, ist dies nur mit durchdachter Privacy-Politik und kombiniert mit Services sinnvoll möglich. Als Beispiel für eine Serviceanwendung mit durchdachtem Communitykonzept kann sms.at der UPC AG gesehen werden. Unter www.sms.at können Besucher kostenlos SMS versandt werden. Dies ist nicht weiter spektakulär, da dieser Service auf vielen der starkfrequentierten Pages geboten wird. Zu einem hier betrachtenswerten Service wird sms.at erst durch die Verbindung mit Services auf mehreren Ebenen. Auf der untersten Ebene kann jeder Besucher anonym den Service nutzen und SMS versenden. Nach einem einmaligen Registrierungsvorgang gelangt er in den Genuss von Zusatzleistungen, wie einem persönlichem Telefonbuch. Der Besucher hat die Möglichkeit bei jedem Einstieg auf die Page seinen Benutzernamen und sein Kennwort einzugeben, oder aber einen Cookie setzen zu lassen. Dadurch wird ein Besucher bei wiederkehrendem Besuch erkannt und der Anmeldevorgang bleibt erspart. Der Cookie wird nicht automatisch auf dem Client hinterlegt, sondern nur auf Wunsch des Benutzers (Opt-In). Die nächsthöhere Ebene ist ein gebührenpflichtiger Service, der zu geringen jährlichen Kosten weitere Services, wie das Senden längerer Texte oder ein Archiv der gesendeten SMS erlaubt. Das Angebot des kostenpflichtigen Service liegt nicht nur in der Finanzierung des Services begründet, sondern auch darin, dass bei kostenpflichtigen Services die Identität preisgegeben werden muss und somit das Userprofil einen ungleich höheren Wert besitzt. Das kostenlose Login

erlaubt dagegen kaum Rückschluss auf die Person, ebenso wie auch das Anlegen von mehreren Profilen durch eine Person und daraus resultierende Dateileichen, nicht vermieden werden kann.

Andere Beispiele für die Verknüpfung von Services mit Loginmechanismen sind Tageszeitungen wie derstandard.at, die digitale Ausgabe des Standard. Die Teilnahme an Discussionboards oder der Zugang zum Archiv ist nur mittels Anlegen eines Userprofiles möglich, wobei das reguläre Contentangebot ohne Anmeldung zugänglich ist.

Generell soll hier betont werden, dass jegliche Form der Registrierung nur in Verbindung mit Services möglich ist, wobei dies nur auf zusätzliche Services beschränkt sein sollte und nicht das Kernangebot reduzieren darf. Am Beispiel sms.at kann auch betrachtet werden, dass ein Registrierungsmechanismus als Vorbereitung für spätere kostenpflichtige Services benutzt werden kann.

Weitere Beispiele für Systeme mit Login-Mechanismen sind Freemailservices. Eine abgeschwächte Form von Services mit vorangehender Registrierung sind Userprofile für Abodienste wie beispielsweise Mailinglisten. Nicht immer ist ein Service, der eine permanente Anmeldung rechtfertigen kann, realisierbar. In solchen Fällen können durch das Angebot von Abodiensten Kundendaten gewonnen werden. Eine Registrierung kann hier den Vorteil von personalisierten News bieten, wobei nicht die persönliche Ansprache gemeint ist, da diese in der Regel weder den Aufwand der Registrierung, noch die Preisgabe der persönlichen Daten rechtfertigt. Viel mehr könne so Inhalte personalisiert werden, wie dies beispielsweise bei presstext.austria geschieht. Unter www.presstext.at kann ein Benutzerprofil angelegt werden, bei dem Interessensgebiete ausgewählt werden. Je nach Profil erhält der Nutzer Pressemeldungen zu den gewählten Interessensgebieten. Durch Angabe von e-Mail Adresse und Passwort kann dieses Profil jederzeit verändert werden. Der Service wird durch den Versand von Originaltext Meldungen finanziert, für deren Aquise die genaue Kenntnis der Nutzer des Services sicher kein Nachteil ist.

Durch die Analyse mittels Data-Mining Systemen können valide Daten über die Nutzer und den Ausbau und die Maßschneidung der Inhalte gewonnen werden.

Weitere Formen der Datenerhebung

Votingsysteme

Ein beliebtes Instrument zur Evaluierung von Angeboten ist die Frage nach der Qualität des Angebots („Wie fanden Sie dieses Buch“), welche von diversen Shop- und Contentsystemen angeboten wird (amazon.com, microsoft.com,...). Besuchern wird die Frage nach dem Wert der aktuell besuchten Seite oder der angebotenen Information gestellt. Auch wenn manche Systeme eher fragwürdig sind, da ein Buch zuerst gelesen werden muss, um Auskunft über das Gefallen des Inhalts geben zu können, können zum Beispiel FAQ Systeme qualitativ überprüft werden. Die klare Aussage, dass die gesuchte Information gefunden oder nicht gefunden wurde, liefert bessere Informationen über die Qualität oder die Mängel einer Seite als Logfiles. Das Logfile beinhaltet lediglich die Information über den Ort an dem der FAQ-Bereich verlassen wurde, nicht aber den Grund.

In Verbindung mit Profiling-Methoden können die hier gewonnenen Daten auch wertvolle Information über den einzelnen Benutzer liefern.

Discussion-Boards

Die Möglichkeiten reichen von ausgeprägten Diskussionssystemen wie die von Tageszeitungen, bis hin zu einfachen Gästebüchern.

Durch intelligente Platzierung wird den Besuchern die Möglichkeit zur Abgabe der Meinung über die Website oder deren Inhalte, bis hin zur aktiven Teilnahme an Inhalten und deren Entstehung. Letzteres wird beispielsweise von beinahe jeder Tageszeitung geboten. DerStandart.at bietet ein, an Login gekoppeltes Diskussion Board, dessen Inhalte die des redaktionell geführten teils quantitativ nicht selten übertreffen. Die inhaltliche Qualität ist durch Selbstregulation durch

die Teilnehmer geregelt. Eine redaktionelle Mitarbeit ist möglich, inhaltliche Kontrolle auf jedem Fall nötig. Die hierdurch entstehenden Kosten müssen durch den Mehrwert dieses Services gedeckt sein.

Web-Umfragen, Gewinnspiele und Usability Tests

Fragebogensysteme im Web bieten einige Vorteile gegenüber klassischen Fragebogensystemen, sie können direkt beim Nutzer und direkt am Forschungsobjekt platziert werden. Mittlerweile existieren eine akzeptable Anzahl an Anwendungen, die von Datenerhebung bis zur Auswertung softwaregestützte Arbeit ermöglichen. Nachteile sind mangelhafte Repräsentativität und die generelle Flüchtigkeit der Internetsurfer (bei sinkendem Interesse verschwindet der Surfer schnell wieder zu Yahoo). Für den Einsatz als Instrument zur Analyse einer Website eignen sie sich primär zur Gewinnung von qualitativen Daten und können die Lücke im Bereich der Motivforschung füllen. Logfiles und sonstige nonreaktiv gewonnene Daten liefern Daten die keinen Aufschluss über die tatsächlichen Beweggründe der Besucher geben. Dazu sind aufwendige Verfahren wie die des Usability Tests oder der Umfrage nötig. Umfragen sind mittels digitaler Fragebogen im kleinen Rahmen realisierbar und zeigen dem Besucher, dass der Service an den Wünschen der Nutzer orientiert und um permanente Verbesserung bestrebt ist. Fragebögen stellen außerdem ein wichtiges Instrument zur Gewinnung von Kundendaten dar, da ein Fragebogen ohne Gratifikation kaum Sinn macht und die sinnvollste Möglichkeit sich für die Mithilfe des Surfers zu bedanken ist die Teilnahme eines Gewinnspiels oder die Gewährung eines Rabatts für spätere Käufe. Ersteres verlangt vom Nutzer die Preisgabe seiner Daten, der Rabatt führt den potentiellen Kunden ein Stück näher zum Kauf.

Weitere Quellen zum Thema: Online Research, Bernard Batinic/Andreas Werner/Lorenz Gräf, Verlag Hogrefe, 1999

Collaborative filtering

Unter Collaborative filtering⁹, oder auch Group Filtering werden lernfähige Systeme verstanden, die basieren auf dem Verhalten von einzelnen Nutzern Regeln bilden und Vorhersagen treffen. Verwendung finden solche Systeme beispielsweise bei Amazon.com. Jeder Interessent eines Produktes findet bei der jeweiligen Detailbeschreibung eine Liste der Produkte, die von bisherigen Käufern ebenfalls erworben wurden.

Systeme dieser Art sind einfach zu betreiben, da sie nur einmal initialisiert werden müssen, jedoch sind hohe Zugriffszahlen für detaillierte Regelbildung nötig. Systeme dieser Art sind beispielsweise Firefly (entwickelt am MIT) oder GroupLens der Universität Minnesota.

Weiter Quellen zum Thema:

<http://www.kontext.org/>

www.sims.berkeley.edu/resources/collab/

⁹ C't 18/2000 - Viele Gesichter Personalisierte Websites stellen sich auf Besucher ein

Definition der Messgrößen

Man unterscheidet einige Standardgrößen, die die wesentlichen Abrufaktivitäten der Nutzer beschreiben. Diese Rohdaten lassen sich aus den Logfiles des Webservers mit geringem Aufwand generieren.

Um vergleichbare Messgrößen zu schaffen, haben sich in Deutschland schon frühzeitig verschiedene Verbände wie der Deutsche Multimedia Verband (DMMV), der Bundesverband Deutscher Zeitungsverleger (BDZV) und der Verband Deutscher Zeitschriftenverleger (VDZ) auf gemeinsame Standards geeinigt. Die Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern (IVW) soll mit ihrem Messverfahren die Verlässlichkeit der Nutzerzahlen gewährleisten.

Hit

Zugriff des Browsers auf ein einzelnes WWW-Element, z. B. auf eine einzelne Seite oder eine Graphik. Da Dokumente meistens aus mehreren miteinander verknüpften Elementen bestehen, handelt es sich um eine unqualifizierte Größe, die keine verlässlichen Aussagen über die Anzahl der Kontakte zulässt. Hits werden im Logfile des WWW-Servers verzeichnet.¹⁰

PageImpression

PageImpressions (bisher PageViews) bezeichnen die Anzahl der Sichtkontakte beliebiger Benutzer mit einer potentiell werbeführenden HTML-Seite. Sie liefern ein Maß für die Nutzung einzelner Seiten eines Angebotes. Liegt dem Web-Angebot ein Layout zugrunde, das Werbeanteil und Inhalt (Content) trennt, so zählt jeder Sichtkontakt mit einer Content-Seite als zusätzlicher Page-View für den zugehörigen Werbeanteil. Als Content-Seiten gelten alle Seiten, bei denen es sich nicht um reine Werbe-

¹⁰ Stephan, Werner: *Marketing Instrument Internet*, S. 211

oder Navigationsseiten handelt. Die Summe aller PageImpressions gibt Aufschluss über die Attraktivität eines Angebots.¹¹

Visit

Ein Visit bezeichnet einen zusammenhängenden Nutzungsvorgang (Besuch) eines WWW-Angebots. Er definiert den Werbeträgerkontakt. Als Nutzungsvorgang zählt ein technisch erfolgreicher Seitenzugriff eines Internet-Browsers auf das aktuelle Angebot, wenn er von außerhalb des Angebotes erfolgt.¹²

Ein Visit ist zwar aussagekräftiger als ein Hit oder eine PageImpression, er ist aber auch schwieriger zu messen. Ein großer Anteil der Visits besteht aus nur einem Klick, d. h. ein Nutzer betritt die Web Site und verlässt diese gleich wieder. Visits mit mehr als einem Klick dokumentieren ein höheres Interesse und Bindungspotential des Besuchers.

Ein Problem bei der Ermittlung von Page Impressions und Visits kann durch Proxy-Server hervorgerufen werden. Diese dienen im Internet zur Zwischenspeicherung von häufig abgerufenen Webseiten. Der Browser des Nutzers ruft die angeforderten Seiten dann oftmals nicht vom eigentlichen Webserver ab, sondern erhält diese von dem dazwischenliegenden Proxy-Server. Für die gemessenen Abrufzahlen bedeutet dies wiederum, dass diese gegenüber der tatsächlichen Nutzung zu niedrig ausgewiesen werden. Bei IVW-geprüften Websites wird diese Problematik entschärft, indem der Abruf einer speziellen Zähldatei zwangsweise erfolgt und damit der Seitenabruf erfasst wird.¹³

¹¹ Informationsgemeinschaft zur Kontrolle der Verbreitung von Werbeträgern e.V.: <http://www.ivw.de/>

¹² Informationsgemeinschaft zur Kontrolle der Verbreitung von Werbeträgern e.V.: <http://www.ivw.de/>

¹³ Informationsgemeinschaft zur Kontrolle der Verbreitung von Werbeträgern e.V., <http://www.ivw.de/>

View Time (Nutzungsdauer)

Für dynamischen Content (darunter fallen etwa Animationen, Chats oder Spiele) könnte es sinnvoll sein, die Messgröße View Time (Nutzungsdauer) einzuführen. In einer vorläufigen Definition bezeichnet die View Time die Zeit, in der ein potentiell werbeführender Teil eines Internet-Angebotes während eines Nutzungsvorganges sichtbar war (Zeit pro Visit). Die gesamte View Time eines (Teil-)Angebotes ergibt sich als Summe der View Time für alle Nutzungsvorgänge des (Teil-)Angebotes. Letztlich ist aber noch offen, ob sich der Begriff View Time durchsetzen wird, zudem ist seine Definition noch im Fluss. Eine Reihe offener, vor allem technischer Fragen ist hier noch zu klären. Ein Beispiel: Die zur Erfassung des Parameters notwendigen Java-Scripte können vom aktuellen MS Explorer nur begrenzt interpretiert werden, und vor allem hat jeder User die Option, Java-Script in den Browser-Options zu deaktivieren (außerdem werden immer noch ältere Browser eingesetzt, die Java-Script überhaupt nicht umsetzen können).¹⁴

AdImpression

AdImpressions sind der messbare Werbemittelkontakt. Sie bezeichnen die Anzahl der von den Clients der Nutzer abgerufenen (requested) Werbemittel vom Server eines Werbeträgers oder anderen AdServers.¹⁵

Erste offizielle Spezifikationen des Deutschen Multimedia Verbandes sind:

- Die abgerufenen AdImpressions werden über das erste zählbare Element eines Werbemittels mittels Redirect-Verfahren erfasst. Ein solches Element muss enthalten sein.
- Der AdImpression erfasst nicht nur den Abruf klassischer (GIF-)Banner, sondern grundsätzlich auch neue Werbeformen.
- Ein automatischer Refresh von Bannern sollte nicht unter 30 Sekunden durchgeführt werden.

¹⁴ Deutscher Multimedia Verband: http://www.dmmv.de/de/7_pub/homepagedmvm/themen/emarketing/media/zielemedia/messgroessenemarketing.cfm

¹⁵ Deutscher Multimedia Verband: http://www.dmmv.de/de/7_pub/homepagedmvm/themen/emarketing/media/zielemedia/adimpressiondefinition.cfm

Eine AdImpression sagt aber nur, dass der User eine Seite im Graphikmodus angesehen hat und das Banner vollständig herausgeladen wurde. Befindet sich das Banner jedoch im unteren Seitenbereich der nur durch Scrollen erreicht werden kann, so ist es leicht möglich, dass der Benutzer das Banner gar nicht gesehen hat. Sicherer messbar ist hingegen der AdClick. Der User klickt auf ein Banner und wird gleich mit einem Werbeangebot verbunden.

Ad Clicks

Die Zahl der Klicks auf ein werbetragendes Objekt (Banner oder Button). Banner sind mit einem Link ausgestattet und führen zu einer weiterführenden Information des Werbetreibenden, z.B. auf dessen Homepage (die nicht verlinkten und in diesem Sinne statischen Buttons verlieren zunehmend an Bedeutung). Ad Clicks indizieren die Anzahl der tatsächlich realisierten Werbemittelkontakte. Während Page Views und Visits die Attraktivität des Werbeträgers widerspiegeln, lässt sich mit den Direktkontakten die Effizienz des Werbemittels bestimmen. Ad Clicks sind aber - auch wenn dies in der Öffentlichkeit zum Teil anders dargestellt - im Unterschied zu den (Media-) Planungsgrößen Page Views und Visits keine Währung, sondern ein Preismodell (Abrechnung primär nach "Pay-per-Click" und nicht nach den Page Views für die Seite, auf die ein Advertising Banner geschaltet wurde).¹⁶

Unique Visitor

Mit Unique Visitor bezeichnet man einen einzelnen identifizierten Besucher einer Website. Dieser wird erfasst, indem beliebig viele Kontakte (PageImpressions) bzw. Zugriffe (Hits) des gleichen Rechners innerhalb eines festgelegten Zeitraumes als Einheit gesehen werden.¹⁷ Die Ermittlung der Zahl von tatsächlichen Besuchern (nicht Besuchen), also die Identifikation individueller Nutzer, ist bisher nur sehr eingeschränkt möglich, z.B. über eine Anmeldeprozedur oder mittels der

¹⁶ Deutscher Multimedia Verband: http://www.dmmv.de/de/7_pub/homepagedmmv/themen/emarketing/media/zielemedia/messgroessenemarketing.cfm

¹⁷ Ahrens & Behrent, Agentur für Kommunikation GmbH, <http://www.a-b-uniquevisitor.de>

umstrittenen Cookies, die vor allem bei Shoplösungen gern verwendet werden und bei Warenkorbsystemen meist sogar unentbehrlich sind.

Die im IVW-Verfahren genutzten Messgrößen "PageImpressions" und "Visits" stellen eine gute und mittlerweile erprobte Planungsbasis für die Mediaplanung dar. Weitere webadäquate Parameter sind jedoch gefordert. Aktuell beschäftigt sich der AK Media mit der Entwicklung neuer Messgrößen in Bereichen Visitor, Flash und Streaming Media sowie den weiteren Werbeformen Chat, E-Mail und Newsletter.¹⁸

Logfileanalyse

Die Methode, mit der die Rohdaten eines Logfiles ausgewertet werden, nennt man Logfile-Analyse. Die Analyse des Logfiles kann wichtige Hinweise auf die Stärken und Schwächen einer Web Site geben. Oft führen Logfiles jedoch ein Schattendasein und ihre Informationen bleiben ungenutzt.

Im Rahmen der Studie "Interaktivität und die Bedeutung für e-Commerce"¹⁹ der Intouch GmbH in Bad Homburg wurden Unternehmen zu den Zielen von Interaktion auf WebSites und Möglichkeiten zur Messung der Zielerreichung befragt. Es sind gerade mal 57 % der befragten Unternehmen, die ihre Server-Logfiles und damit die Nutzung ihrer Website analysieren. 43 % der Betreiber von Internetpräsenzen verzichten darauf, ihre Logfiles auszuwerten und lassen die Chance, mit Hilfe dieser Nutzungsdaten ihr Webangebot zu optimieren, ungenutzt verstreichen (Stand 5/99)²⁰.

In unbearbeiteter Form haben Messgrößen, die in einem Logfile generiert werden, relativ wenig Informationsgehalt. Sie müssen daher erst zu aussagekräftigen Kennzahlen verdichtet werden, um eine umfassende Analyse

¹⁸ Deutscher Multimedia Verband: http://www.dmmv.de/de/7_pub/homepagedmmv/themen/emarketing/media/zielemedia.cfm

¹⁹ integra Deutschland: http://www.integra-europe.at/dialog_studien.html

²⁰ infotrust: <http://www.infotrust.de/leistungen/webcontrolling.html>

der Web Site möglich zu machen. Es ist sinnvoll, mehrere, sachlich zusammenhängende Kennzahlen zu einem Kennzahlensystem zu verbinden.

Im ersten Teil wird beschrieben, welche Informationen sich von den Rohdaten ableiten lassen. Kombiniert man absolute Zahlen entstehen Verhältniszahlen mit einem betriebswirtschaftlichen Zusammenhang. Anschließend wird den Kennzahlen ein zeitlicher Aspekt hinzugefügt. Einerseits werden bestimmte Zeitintervalle einander gegenübergestellt, um die Veränderung der gemessenen Kennzahl feststellen zu können und andererseits wird das Nutzungsverhalten über längere Zeiträume hinweg beobachtet. Weiters werden Kennzahlen unterschieden, die sich nicht nur auf Daten aus dem Logfile stützen, sondern auch Informationen einbeziehen, die außerhalb der Web Site erhoben werden.

Kenngrößen aus Rohdaten

Welche Kennzahlen nun zur Beurteilung der eigenen Website benötigt werden, hängt in erster Linie von der Funktion ab, die die Webpräsenz für den Anbieter erfüllt. Die Auswahl der benötigten Kennzahlen hängt von der jeweiligen Branche, dem Unternehmen, dem Adressaten oder dem Verwendungszweck ab. Websites von Medien, aber auch PR-orientierte Unternehmenspräsenzen, werden vorwiegend an der Messung des publizistischen Erfolgs interessiert sein: Welche Inhalte werden wie oft von wievielen Nutzern zu welcher Zeit abgerufen? Betreiber von Online-Shops wünschen sich dagegen vor allem valide Daten über den Absatz und die Umsatzentwicklung ihrer per Web vertriebenen Produktpalette.

Eine hohe Zahl von Seitenabrufen pro Besuch belegt nicht unbedingt, dass es sich um ein interessantes Angebot handelt, sondern kann auch durch eine umständliche Navigation verursacht werden. Grundsätzlich sollte sich die Anzahl der Informationsebenen bzw. der nötigen Klicks bis zum Ziel im Interesse der Nutzer in Grenzen halten, auch wenn dies zu einer geringeren Zahl abgerufener Seiten führt. Das gilt besonders bei Online-Shops. Hier stellt

der direkte Weg zum Ziel einen wesentlichen Vorteil für den Kaufinteressierten dar.

Nicht immer ist beispielsweise eine hohe Anzahl von Page Views auch positiv zu bewerten. In einem Servicebereich werden z. B. niedrige Nutzungszahlen angestrebt, denn somit kann man davon ausgehen, dass auch die Kunden zufrieden sind. Vergleichsweise hohe Zahlen sind dagegen bei Produktinformationen und Online-Bestellungen wünschenswert.²¹ Wird ein Link im oberen Bereich der Homepage platziert wird dieser dementsprechend oft angeklickt, was aber nicht heißen muss, dass dieser auch besonders interessant für die Benutzer ist.

Aus Marketingsicht ist natürlich besonders interessant, wie die Nutzer zur Website gelangen, über welche Seite sie das Angebot wieder verlassen und welche Aktionen sie in der Zwischenzeit ausführen bzw. welche Pfade sie durch die Website nehmen. Die Referrer geben Auskunft darüber, von welcher Website der Besucher auf das eigene Angebot überwiesen wurde.²² Nicht überraschend ist, dass die Homepage meistens am Anfang eines Nutzungsvorganges steht. Interessanter scheint da schon zu sein, jene Seiten zu überprüfen, die wiederholt am Ende eines Nutzungsvorgangs stehen. Gründe für den Ausstieg können z. B. einfach nur das Beenden ein Kaufvorganges sein, jedoch auch mangelnde Attraktivität von Seitendesigns oder ein unverständlicher Inhalt die Ursache sein. Bei Suchmaschinen kann man zusätzlich feststellen, welche Suchbegriffe die Nutzer dort eingegeben haben, um zur entsprechenden Website zu gelangen.

Das Logfile gibt weiters Auskunft über die Häufigkeit der Benutzung von Hyperlinks. Man sollte über die Schaltung einer Bannerwerbung nachdenken, wenn ein Link besonders stark frequentiert wird.

Es ist möglich, über den Domain-Namen jene Länder herauszufinden, aus denen die Nutzer kommen bzw. welchen Organisationen sie angehören. Wird

²¹ Pispers: Web-Controlling: Internet-Erfolge planen, steuern und analysieren, S. 219

die Web Site in größerem Umfang auch von fremdsprachlichen Nutzern besucht, sollte man darüber nachdenken, den Inhalt auch in anderen Sprachen anzubieten. Bezeichnungen der Subdomains weisen oft auf Organisationen hin und enthalten z. B. Namen von Unternehmen, Universitäten oder Zugangsprovidern. Mit Hilfe dieser Information kann man den Inhalt noch besser an seine Benutzergruppe anpassen.²³

Sinnvoll ist der Einsatz von Auswertungssoftware wenn es um das Aufspüren von Fehlern in der Website geht, z. B. bei Nutzerzugriffen auf nicht mehr vorhandene Seiten. Schreibfehler und Links, die sich auf veraltete Adressen beziehen sind die häufigsten Ursachen und lassen sich teilweise durch geeignete Gegenmaßnahmen beheben. Die Web Site sollte außerdem dahingehend überprüft werden, wie sie von den am häufigsten genutzten Browsern und Betriebssystemen angezeigt wird. Ältere Versionen sind unter Umständen nicht fähig, bestimmte Gestaltungselement zu verarbeiten.

Tabelle 1 fasst die Kenngrößen aus Rohdaten zusammen.

Kenngrößen		
Hits	Registrierte Nutzer	Suchbegriffe
Page Views	Nutzungszahlen von Seiten	Top-Level-Domains
Visits	Einstiegsseiten	Subdomains
User	Ausstiegsseiten	Übertragenes Datenvolumen
Ein-Klick-Besuche	Linknutzung	Übertragungsfehler
Weitergehende Besuche	Reichweite Bannerwerbung	Browser-Typen, -Versionen
Mehrfachbesucher	Click-Throughs der Banner	Betriebssysteme

Tabelle 1: Kenngrößen aus Rohdaten der Web-Site-Nutzung

Quelle: Schwickert, Wendt: *Controlling-Kennzahlen für Web-Sites*, S. 655

²² Chase, Hanger: *Essential Business Tactics for the Net*, S. 219

²³ Schwickert, Wendt: *Controlling-Kennzahlen für Web-Sites*, S. 654

Kombination von Rohdaten

Verwendet man ausschließlich Kennzahlen auf der Basis von Rohdaten, ist die Analyse der Web Site wenig aussagekräftig, da den Werten Vergleichsmaßstäbe fehlen. Verhältniszahlen, die die absoluten Zahlen in Form von prozentualen Anteilswerten wiedergeben, relativieren die Angaben über die Nutzer und deren Besuche.

Eine wichtige Größe in diesem Zusammenhang ist das Verhältnis zwischen der Anzahl der Page Impressions und Visits. Diese sagt einiges über das Nutzungsverhalten der User aus, insbesondere über die Intensität der Nutzung. Ein hoher Prozentsatz an Ein-Klick-Besuchen deutet zwar auf effektive Promotion-Maßnahmen hin, doch die Attraktivität des präsentierten Inhalts scheint vergleichsweise niedrig zu sein. Web-Site-Anbieter sind daher an einem hohen Anteil weitergehender Nutzungsvorgänge interessiert. Bei den General-Interest-Sites werden meist Werte im Bereich von 4 zu 1 erreicht, d.h. pro Visit werden durchschnittlich vier Seiten abgerufen.

Auch im Online-Handel steht der Kunde im Mittelpunkt. Zur Bestimmung von Einkaufsmustern und Präferenzen liefern die Logfiles mehr als erste Anhaltspunkte. Man will die Vorlieben der Kunden und ihre Gewohnheiten besser kennenlernen - mit dem Ziel, ein genau angepasstes Sortiment zum günstigen Preis anzubieten.

Die wirklich „wertvollen“ Nutzer im Sinne des Marketings sind jene, die mehrfach wiederkehren und vor allem jene, die sich registrieren. Hohe Anteile dieser Nutzer im Vergleich zur Gesamtzahl der Nutzer werden angestrebt. Der von den Usern erwartete Nutzen muss jedoch schon erheblich sein, dass diese ihre Daten preisgeben.

Die Bedienungsfreundlichkeit der Web Site, die sogenannte Usability, ist ein Thema, dass bei Rentabilitätsüberlegungen nicht außer Acht gelassen werden darf. Eine Analyse der meistgenutzten Klickpfade kann Aufschluss über das Nutzerverhalten geben. Referrer-Angaben geben Auskunft darüber, in welcher Reihenfolge die Seiten aufgerufen wurden. Somit kann man beispielsweise

feststellen, ob sich die User vor der Bestellung eingehend über die Produkte informiert haben oder ob plötzliche Abbrüche auf Navigationsprobleme zurückzuführen sind. Besondere Muster, die in einem Klickpfad auftreten, sind etwa Schleifen, die darauf hindeuten, dass die Orientierung schwierig ist oder Informationen unübersichtlich über mehrere Seiten verteilt sind.

So empfiehlt Sari Kalin vom Cio Web Business Magazine,²⁴ sich nicht zu früh über steigende Zahlen bei den Zugriffen auf die siteinternen Suchmöglichkeiten zu freuen. Häufig ist dies eher ein Zeichen dafür, dass die Website für die Besucher nicht durchsichtig ist. Untersuchungen der User Interface Engineering²⁵ belegen, dass Besucher, die sich mit Hilfe der feststehenden Links durch eine Site bewegen, mit mehr als 50 Prozent höherer Wahrscheinlichkeit die gewünschten Inhalte finden, als die Besucher, die die Suchfunktion nutzen. Auch hier kann die Logfileanalyse wichtige Anhaltspunkte zur Verbesserung liefern.

Um die Effektivität einer Werbung zu bestimmen, ist es hilfreich, die Klickrate (Click-Through-Rate) zu bestimmen. Diese errechnet sich aus dem Verhältnis von Click-Throughs zur Anzahl der Abrufe einer werbetragenden Seite. Der Wert der Klickrate liegt zwischen einem und zwei Prozent.²⁶

Im Zusammenhang mit Fragebögen stellt sich immer die Frage wie repräsentativ die Ergebnisse sind. Erst die Antwortquote, also das Verhältnis ausgefüllter Fragebögen zur Gesamtzahl der Nutzer, zeigt, wie erfolgreich eine Fragebogenaktion ist und wie sehr die Nutzer bereit sind, persönliche Informationen weiterzugeben.

²⁴ Sari Kalin: <http://www.cio.com/>

²⁵ User Interface Engineering: <http://www.national.com/>

²⁶ Werner: *Site Promotion: Werbung auf dem WWW*, S. 59

Tabelle 2 fasst die Kenngrößen aus der Kombination von Rohdaten zusammen.

Kenngrößen aus der Kombination von Rohdaten		
Anteil der Ein-Klick-Besuche	=	Ein-Klick-Besuche / Visits gesamt
Anteil weitergehender Besuche	=	Weitergehende Besuche / Visits gesamt
Besuchstiefe	=	Page Views / Visits
Anteil der Mehrfachnutzer	=	Mehrfachnutzer / User gesamt
Anteil registrierter Nutzer	=	Registrierte Nutzer / Nutzer gesamt
Klickpfade	=	Reihenfolge der Seitenaufrufe
Click-Through-Rate	=	Click-Throughs / Reichweite
Antwortquote von Fragebögen	=	Ausgefüllte Fragebögen / Page Views der Seite

Tabelle 2: Kenngrößen aus der Kombination von Rohdaten

Quelle: Schwickert, Wendt: *Site Promotion: Werbung auf dem WWW*, S. 658

Kennzahlen mit Zeitbezug

Auch die Analyse des Faktors Zeit spielt eine zunehmend wichtige Rolle. Zu welcher Tageszeit oder an welchen Wochentagen wird die Website am stärksten frequentiert? Diese Information lässt sich leicht aus den Zeitangaben eines ausgewerteten Logfiles ablesen.

Wie lange dauert jedoch etwa ein durchschnittlicher Besuch oder werden bestimmte Seiten betrachtet? Diese Fragen sind schon etwas schwieriger zu beantworten und es ist daher nötig, den Kennzahlen eine zeitliche Komponente hinzuzufügen. Grundsätzlich misst man die Zeitintervalle im Nutzungsverhalten oder das Verhalten über längere Zeiträume hinweg.

Ein grundlegendes Zeitintervall zur Beschreibung der Web-Site-Nutzung ist die Verweildauer. „Die Verweildauer ist die Zeitspanne, die ein Nutzer auf einer bestimmten Website verbringt. Dabei wird die Zeit zwischen dem ersten und letzten Zugriff auf die Seite gemessen. Die Verweildauer gibt Auskunft darüber, wie intensiv die Besucher das Angebot der Website nutzen.“²⁷ Betrachtet man

²⁷ Ahrens & Behrent, Agentur für Kommunikation GmbH, <http://www.a-b-verweildauer.de/>

diese Kennzahl im Zusammenhang mit dem Inhalt einer Seite, kann man feststellen, ob die Information vom User erfasst werden konnte oder nur flüchtig registriert wurde. Man darf nicht generell davon ausgehen, dass eine lange Verweildauer etwas Positives darstellt. Seiten des Kundenservices sollten nur kurz genutzt werden. Um zu messen wie lange ein Seitenbereich die Aufmerksamkeit der Nutzer auf sich zieht, wird die durchschnittliche Verweildauer aller Nutzer für die betreffende Seite ermittelt.

Ebenso wie die Nutzungsdauer kann für jeden User der Zeitraum ermittelt werden, in dem er der Web Site fernbleibt. Wie groß dieser Zeitraum üblicherweise ist, hängt auch vom Angebot der Web Site ab. Die Online-Version einer Tageszeitung hat einen regelmäßigen Nutzer wahrscheinlich verloren, wenn dieser über eine Woche nicht auf das Angebot zugreift. Auf Service-Seiten trifft dies natürlich nicht zu.

Die Besuchsfrequenz stellt dar, wie oft ein Nutzer in einem bestimmten Zeitintervall auf eine Web Site zugreift. Wieder ist dies abhängig vom Angebot der Web Site. In der Regel wird sich der Kunde kurz vor einer Bestellung vermehrt über das Produkt informieren. Die Besuchsfrequenz steigt an und der Anbieter kann, sofern er dies rechtzeitig bemerkt, darauf reagieren und den potentiellen Kunden mit einem Anreiz endgültig überzeugen.²⁸

Mit Hilfe der durchschnittlichen Server-Antwortzeit wird angegeben, wie schnell ein HTML-Dokument durchschnittlich an einen Nutzer weitergeleitet wird. Dies lässt die Beurteilung der Leistung eines Service-Providers zu bzw. ermöglicht den Vergleich mit anderen.

Bietet man Downloads an, sollte man die durchschnittliche Download-Zeit ermitteln. Ist die Übertragungszeit dem Nutzer nicht zumutbar, ist es sicher sinnvoll die Datei aufzuteilen.

²⁸ Schwickert, Wendt: *Controlling-Kennzahlen für Web-Sites*, S. 659-660

Beobachtet man das Nutzungsverhalten über längere Zeiträume hinweg, sprich Tage, Wochen, Monate und sogar Jahre, lassen sich Prognosen über die zukünftige Nutzung ableiten.

Die tägliche Nutzung wird durch den Verlauf stündlicher Kennzahlen dargestellt. In der Regel steigen die Nutzungszahlen bis zum Nachmittag an, sinken bis zum späten Abend leicht und zeigen niedrige Werte in der Nacht um am frühen Morgen.

Eine Betrachtung des wöchentlichen Nutzungsverlaufes zeigt eine überwiegend geschäftliche Nutzung während der Woche und eine hauptsächlich private Nutzung am Wochenende. Weiß man über solche Nutzungsverläufe Bescheid, kann man den Bedarf an Kapazitäten dementsprechend berücksichtigen.

Die Kenngrößen Störungsrate und Störungsintensität geben Auskunft über die Anzahl der und die Ausfallzeit durch Störungen. Besonders in Branchen die eine hohe Systemverfügbarkeit garantieren müssen (z. B. Online-Banken) ist die Bewertung der Zuverlässigkeit des Servers wichtig.

Tabelle 3 fasst die Kenngrößen mit Zeitbezug zusammen.

Kenngrößen mit Zeitbezug		
Verweildauer eines Besuches	=	\sum Verweildauern einzelner Page Views
Durchschnittliche Verweildauer	=	\sum Verweildauern einzelner Nutzer / Nutzeranzahl
Dauer des Fernbleibens	=	Aktuelles Datum – Datum letzter Besuch
Besuchsfrequenz	=	Besuche eines Nutzers / Zeitraum
Durchschnittliche Server-Antwortzeit	=	Zeit zwischen Client-Request und Server-Response
Durchschnittliche Download-Zeiten	=	\sum Übertragungszeit / Anzahl Downloads
Tägl., wöchentl., monatl., jährlicher Nutzungsverlauf	=	Nutzerzahlen, Seitenaufrufe, Server-Auslastungen über bestimmte Zeiträume
Störungsrate	=	Anzahl Störungen / Nutzungsdauer System
Störungsintensität	=	Ausfallzeit durch Störungen / Nutzungsdauer

Tabelle 3: Kenngrößen mit Zeitbezug

Quelle: Schwickert, Wendt: *Site Promotion: Werbung auf dem WWW*, S. 661

Kenngrößen mit Einbezug externer Daten

Manche Kennzahlen können nur ermittelt werden, indem Daten, die außerhalb der Web Site erhoben werden, einbezogen werden (z. B. Umsatz).

Eine Kenngröße dieser Kategorie ist die Anzahl der Web-Kunden. Das ist jener Teilbereich der Web-Site-Nutzer, der ein Produkt kauft, also Umsatz generiert. Den höchsten Wert haben loyale Kunden, da sie regelmäßig wiederkehren und dem Unternehmen somit auch in Zukunft Umsätze verschafft. Ehemalige Kunden gilt es, durch geeignete Marketingmaßnahmen, wieder zum Kauf zu bewegen. Sie stellen also das reaktivierbare Kundenpotential dar.

Aber nicht nur das Verhältnis von Besuchern zu Käufern ist aufschlussreich, auch die Suche nach Ladenhütern, die in den virtuellen Regalen verstauben, macht Sinn. Auf der Grundlage von Logfileanalysen nahm z.B. die National

Semiconductor Corp.²⁹ in 1998 über 3.500 Artikel, wie ältere Halbleiter und andere elektronische Bauteile, aus dem Online-Programm ihres Business-to-Business Angebotes. Sie fanden kaum Nachfrage, aber die entsprechenden Seiten mussten dennoch gepflegt werden. Diese Kosten wurden eingespart.

Eine Analyse der Online-Bestellungen zeigt nicht nur, welche Produkte besonders gern über die Web Site verkauft werden bzw. welche die eben benannten Ladenhüter sind, sondern lässt auch Schlüsse über Beziehungen zwischen den Produkten zu. Werden gewisse Produkte meist in Verbindung mit anderen verkauft, dann bestehen sicherlich Potentiale für Cross-Selling (Verkauf von Produktarten, die in keinem direkten Zusammenhang stehen). Besteht jedoch ein Zusammenhang zwischen den Produkten (z. B. PC und zusätzliche Hardware) spricht man vom sogenannten Up-Selling. Um diese Potentiale optimal zu nutzen, sollte man häufige Kombinationen möglichst nah aneinander platzieren, d. h. wenn schon nicht auf der gleichen Seite, dann zumindest nur wenige Klicks voneinander getrennt.³⁰

Besonders aufschlussreich kann es sein, herauszufinden, warum potentielle Käufer ohne Aufgabe einer Bestellung ihren elektronischen Warenkorb stehen lassen. Gegebenenfalls ist auch hier eine Optimierung notwendig. Ursachen für verlassene Warenkörbe sind beispielsweise komplizierte oder umfangreiche Bestellverfahren oder Fehler in der Datenübermittlung.

Abschließend wäre sicher ein Vergleich der eigenen Daten mit jenen der Konkurrenz aufschlussreich, doch sind Daten der Wettbewerber kaum verfügbar. Unternehmen sind kaum bereit sich in die Karten sehen zu lassen und veröffentlichen ihre Informationen daher üblicherweise nicht.

Die Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern (IVW)³¹ bietet daher ein standardisiertes Verfahren zur

²⁹ National Semiconductor Corp.: <http://www.national.com/>

³⁰ Schwickert, Wendt: *Controlling-Kennzahlen für Web-Sites*, S. 659-662

³¹ Informationsgemeinschaft zur Kontrolle der Verbreitung von Werbeträgern e.V.: <http://www.ivw.de/>

Bewertung einer Web Site an, dessen Ergebnisse allgemein zugänglich sind. Es werden aber nur die beiden Basiskennzahlen PageImpression und Visit angegeben, die nur wenig über den Erfolg einer Web Site aussagen.

Weitere externe Daten liefern Studien von Marktforschungsinstituten, die Branchenvergleiche anstellen oder die allgemeine Entwicklung der Internet-Nutzung betreffen.

Tabelle 4 fasst die Kenngrößen mit Einbezug Web-Site-externer Daten zusammen.

Kenngrößen mit Einbezug Web-Site-externer Daten	
Anzahl Web-Kunden	= Anzahl der Nutzer, die auch als Käufer auftreten
Web-Site-Kundenquote	= Anzahl der Web-Site-Kunden / Gesamtzahl Kunden
Loyale Kunden	= Regelmäßig wiederkehrende Kunden
Reaktivierbares Kundenpotential	= Ehemalige Kunden
Anteil verlassener Warenkörbe	= Verlassene Warenkörbe / Gesamtzahl Warenkörbe
Cross- und Up-Selling-Potentiale	= Verkaufsvolumina nicht-/verbundener Produkte
Anzahl Klicks/Transaktion	= Anzahl Schritte bis zur erfolgten Bestellung
Standardisierte IVW-Ergebnisse	= Standard. Meßverfahren für Page Views, Visits

Tabelle 4: Kenngrößen mit Einbezug Web-Site-externer Daten

Quelle: Schwickert, Wendt: *Site Promotion: Werbung auf dem WWW*, S. 663

Die Auswertung der Logfiles sollte zum Pflichtprogramm jeder professionell betreuten Website gehören. Und dies nicht nur bei Online-Werbeträgern, deren Anzeigenkunden immer detaillierte Informationen über die Nutzung der Website verlangen, bevor sie Budgets fürs Online-Advertising locker machen. Für jede Website ist die Erfassung und Auswertung von Besucherzugriffen von großer Bedeutung, wenn es darum geht, die Auswahl und Präsentation der eigenen

Inhalte zu optimieren. Nur wer weiß, wie seine Site genutzt wird, kann die richtigen Schritte zur kontinuierlichen Verbesserung einleiten und so der Konkurrenz immer einen Mouseclick voraus sein.

Neue Ansätze der Erfolgskontrolle

Mit dem Zeitalter des E-Commerce ist auch ein neues Zeitalter der Erfolgskontrolle entstanden, da alle auf dem digitalen Weg vom Werbemittel bis zur Kaufentscheidung erzeugten Informationen in Form von strukturierten Daten vorliegen. Nie zuvor war es möglich, den Kunden schon beim Kontakt mit dem Werbemittel „kennenzulernen“ und ihn bis in den Warenkorb hinein zu begleiten. Und sich anschließend auch noch nach Tagen und Wochen noch an seine Kaufgewohnheiten zu erinnern. Die intelligente Analyse aller vorhandenen Daten entlang der digitalen Wertschöpfungskette ermöglicht es heute, dem Online-Kunden eine persönliche Ansprache und individualisierte Angebote zu offerieren. Die Personalisierung von Angebot und Inhalten ist als der kritische Erfolgsfaktor im Internet auszumachen. Voraussetzung für eine solche Strategie ist die detaillierte Analyse der Wertschöpfungskette im E-Commerce, die durch ein konsequentes Web-Controlling erst ermöglicht wird.

Instrumente der Erfolgskontrolle

Die heute eingesetzten Instrumente der Erfolgskontrolle, wie beispielsweise AdServer, Logfile-Analyse und Cookie-Technologie bieten nach Meinung vieler Autoren keine zufriedenstellenden Ergebnisse. Somit werden Millionenbeträge in das Internet investiert, ohne dass ein fundierter Controlling-Ansatz implementiert wäre. „Viel mehr hat man es mit Analysen ohne konkrete Ergebnisse zu tun“, schreibt Ralf Pispers, Vorstand der mindfact interaktive medien ag, Pulheim.³² Im folgenden werden die bisher eingesetzten Ansätze beleuchtet:

AdServer, die zur Verwaltung von Banner-Platzierungen eingesetzt werden, liefern zwar Werte wie Werbemittelkontakte (AdViews) und Klickrate. Eine Aussage darüber, ob die Besucher die über das Werbemittel (Banner) auf die

³² Pispers: *Web-Controlling: Internet-Erfolge planen, steuern und analysieren*, S. 214

Site gelangen sind, auch zu Kunden geworden sind, liefert sie in der Regel nicht.

Logfile-Analysen haben mit Problemen wie Frames, Proxy-Servern und dynamischen IP-Adressen zu kämpfen und stoßen sehr schnell an ihre Grenzen, wenn es um fundierte Aussagen zu speziellen Fragestellungen geht. Diese quantitativen Analysen geben darüber hinaus einerlei Auskunft über den betriebswirtschaftlichen Erfolg der Online-Aktivitäten. Als zusätzlicher Nachteil erweist sich, dass die Ergebnisse zumeist erst mit einiger zeitlicher Verzögerung vorliegen. Eine Reaktion in Form von Maßnahmen kann also oft erst viel zu spät erfolgen. Die Folge ist das Verschwenken wertvoller Zeit, in der keine Kunden und Umsätze generiert werden oder in der sich entscheidet, ob die Internet-Gemeinde einen Ansatz annimmt oder nicht.

Cookie-Konzepte beinhalten neben datenschutzrechtlichen Unklarheiten Akzeptanzprobleme auf der Benutzerseite. Darüber hinaus besteht das Problem der Inkonsistenz, da viele Anwender die inhaltlich zum Teil immer noch recht zweifelhaften Dateien gerne von ihrem Rechner entfernen. Kommt der Anwender beim zweiten Besuch ohne Cookie wieder, scheint es sich um einen neuen Besucher zu handeln. Der Cookie verliert so seine Bedeutung.

Ein weiteres Problem bestehender Analysen ist, dass die verschiedenen Ergebnisse der eingesetzten Methoden lediglich isoliert betrachtet werden können, d.h., eine Verknüpfung der Logfile-Analyse mit den Ergebnissen des AdServers existiert zumeist nicht. Wertvolle Informationen gehen verloren bzw. können nicht generiert werden. Die hier aufgezählten Probleme stellen nur einen kleinen Ausschnitt der heute unbefriedigenden Ergebnisse der Erfolgskontrolle von Online-Aktivitäten dar.

Fazit: Die bisherigen Methoden der Auswertung im Internet genügen in keinem Fall den Ansprüchen, die für eine fundierte Entscheidungsgrundlage benötigt werden.

Web-Controlling

Aufbauend auf diesen Problemen und Unzulänglichkeiten der Erfolgskontrolle von Online-Aktivitäten wurde das Konzept des Web-Controlling entwickelt.

Dieses Konzept stellt eine völlig neuartige Kombination aus Software-Produkt und Dienstleistung dar. Es besagt, dass die Lösung der bestehenden Probleme in der vollständigen Analyse der Online-Aktivitäten mit Hilfe von Data Warehouse-Technologie liegt.

Web-Controlling dient der Planung, Überwachung, Kontrolle und Steuerung von Inter-/Intra- und Extranet-Maßnahmen. Web-Controlling liefert unter anderem valide und flexible Informationen über Zugangswege, Zugriffshäufigkeiten und nachfolgend ausgelöste Transaktions- bzw. E-Commerce-Aktivitäten, wie etwa Bestellungen.³³

Web-Controlling erlaubt es erstmals, die gesamte E-Commerce Wertschöpfungskette integrativ zu analysieren. Vom Klick auf das Werbebanner über die Site in den Shop und sogar bis in das Warenwirtschaftssystem kann das Verhalten der Kunden erfasst und analysiert werden. Die einzelnen Daten-Pools werden in einem Data Warehouse gebündelt und bilden somit die Informationsbasis für die daraus abzuleitenden Entscheidungen und Maßnahmen (siehe Abbildung 3).

³³ Framfab: <http://www.framfab.de/>



Abbildung 3: Vollständige Analyse der Online-Wertschöpfungskette

Quelle: Framfab: *Web-Controlling*, <http://www.webcontrolling.de/loesung.html>

Web-Controlling kombiniert hochqualifizierte Dienstleistung mit Standard-Software und bezieht externe Einflussfaktoren mit ein, die dann eine Ursache-Wirkungsanalyse möglich machen. Konnten bisher die Ursachen von auffälligen Veränderungen des Nutzerverhaltens auf der Web Site nicht ermittelt werden, liefert die Einbeziehung externer Einflussgrößen nun konkrete Ergebnisse. Die Erfassung derartiger Parameter erlaubt das Konzept der Acti- und EnviBase. Während in der ActiBase (Maßnahmendatenbank) alle den Erfolg der Web Site beeinflussenden Aktivitäten und Maßnahmen – wie bspw. Banner-Kampagnen, Offline-Promotionen und Mailing-Aktionen – erfasst werden, handelt es sich bei der EnviBase (Umfelddatenbank) um eine Datenbank, die nicht beeinflussbare Faktoren wie Aktivitäten des Wettbewerbs, Urlaubszeiten und Wetterdaten enthält. Diese beiden Datenbanken werden kombiniert und die Anwender erhalten ausführliche Stärken- und Schwächenprofile sowie

aussagekräftige Wirtschaftlichkeits-Analysen, die eine profunde Basis für die künftige Strategie darstellen.³⁴

Der Zugriff auf die Analyse-Reports erfolgt dabei über eine einfache Browser-Oberfläche und ist somit leicht bedienbar. Die Reports werden dabei zielgruppen- bzw. anwenderspezifisch administriert. Jeder Anwender erhält nur die Ansicht auf seine Berichte. Zahlen anderer Abteilungen bleiben ihm verschlossen.

Im Zuge eines „Active Data Warehouse“ lassen sich Nachrichten selbständig an einen dafür zuvor eingerichteten Empfänger verschicken oder sogar Maßnahmen eigenständig auslösen. Beim Erreichen von bestimmten Schwellenwerten (z. B. Klickrate eines Banners sinkt unter 1 Prozent), erhält der Mediaplaner einen Alert, das Banner auszutauschen. Diese Information kann je nach Bedarf an verschiedene Ausgabemedien wie Handy, E-Mail, Fax, Pager oder Drucker gesendet werden. In der vollständigen Automatisierung des Prozesses kann das Modul den Banner-Austausch auch selbständig vornehmen.³⁵

Web-Controlling bietet gegenüber bestehenden Systemen der Erfolgskontrolle ein Höchstmaß an Benutzerfreundlichkeit und Übersichtlichkeit. Die Bedienung des Systems erfolgt vollständig über den Browser. Verschiedene Abteilungen im Haus können über das Intranet auf die Reports zugreifen und externe Abteilungen (z. B. Media-Agentur) erhalten über das Internet Einsicht in für sie relevante Berichte und Masken. Das Konzept des „Active Data Warehouse“ erlaubt auch die weitgehende Automatisierung verschiedener Prozesse. Die Entscheider erhalten aktuellste Informationen über das Geschehen der Web Site und können so ohne zeitliche Verzögerung eingreifen, wenn bestimmte Tendenzen deutlich werden.

Web-Controlling umfasst somit die permanente Analyse und Steuerung sämtlicher Internet-Aktivitäten auf der Basis eines umfassenden Informations-

³⁴ Framfab: Web-Controlling, <http://www.webcontrolling.de>

³⁵ Pispers: *Web-Controlling: Internet-Erfolge planen, steuern und analysieren*, S. 217

und Kennzahlen-Systems mit dem Ziel einer Optimierung sämtlicher E-Commerce-Aktivitäten. Web-Controlling geht damit weit über die quantitative Logfile-Analyse hinaus.

2002 oder 1984?

„Für den Thesaurus von Microsoft Word ist `Anonymität` synonym zu `Untergrund`, `anonym` gleich `amtlich`. Für freiheitlich denkende Menschen ist Anonymität hingegen ein anerkanntes Grundrecht.“³⁶

Wie in den vorigen Kapiteln dargelegt wurde, ist die Sammlung von Daten der Kunden und Besucher auf verschiedenen Ebenen möglich und kann sehr weitreichend sein. Das Internet entstand als anonymes Medium, weshalb zur Erstellung von Userprofilen neue Instrumente gefunden werden mussten. Diverse der eingesetzten Verfahren sind nonreaktiv, weshalb der Besucher auch oft im unklaren über die angelegten Datenprofile sind. In diesem Kapitel soll anhand von zwei Beispielen gezeigt werden, welche Möglichkeiten durch Ausreizung des Systems bestehen und wie die Benutzer darauf reagieren. Es soll hier nicht die Vor- und Nachteile der Anonymität beschreiben werden. Vollkommene Anonymität existiert ohnehin nicht, da im Falle einer Straftat beinahe jede Spur rückverfolgt werden kann. Über Kundendaten von Providern oder Telefongesellschaften kann beinahe jede Aktivität im Internet zu einer Person rückverfolgt werden. Diese Vermittlungsdaten sind jedoch für kommerzielle Datensammler nicht erreichbar und sind deshalb hier nicht interessant.

Big Brother Inside

Als erstes Beispiel für mangelhaften Umgang mit Verbraucherwünschen soll hier die Intel CPU ID und die Reaktion der Datenschützer auf die Pläne von Intel beschrieben werden.

Anfang 1999 wurde bekannt, dass Intel im kurz darauf erscheinenden Pentium III Prozessor eine eindeutige Seriennummer implementieren werde³⁷. Dadurch wurde jeder Prozessor, und somit auch jeder Computer eindeutig identifizierbar. Die Seriennummer selbst war nicht das Problem, als kritisch wurde vielmehr die

³⁶ c't 6/2000 - Unerkannt surfen

³⁷ c't 3/1999 - Sicherheit kontra Datenschutz - Pentium III: Umstrittene Seriennummer

Tatsache betrachtet, dass diese ID mittels Software auslesbar und für E-Commerce Anwendungen konzipiert war. Kritiker empfanden diese Idee als Orwellschen Überwachungswahn. Einerseits löst das Problem nicht das einige Kapitel zuvor beschriebene Problem, nämlich die mangelnde Rückschließbarkeit von Computer auf Benutzer. Die Problematik wird durch zunehmende Integration des Internet noch verstärkt. Da Intel Prozessoren zwar weit verbreitet ist, aber künftige Internethardware nicht zwangsläufig ein PC sein muss, stellt sich hier schon die Frage nach der Sinnhaftigkeit der Nummer. Ein Computer ist zwar identifizierbar, aber die konkrete Identität des Benutzers muss für sensible Anwendungen trotzdem mittels anderer Verfahren (Login,...) sichergestellt werden.

Laut Intel sollte die Seriennummer softwaremäßig deaktivierbar und die Anonymität des Benutzers somit sichergestellt sein. Kurze Zeit später erschien ein Programm, welches die zuvor deaktivierte ID wieder aktivierte, hundertprozentige Abschaltung war also nicht möglich³⁸. Angesichts der zunehmenden Verbreitung von Viren und Spyware bietet das rein softwareseitige Auslesen auch einen unzureichenden Schutz. Datenschützer befürchteten ein globales Cookie³⁹, das keine Kontrolle durch den Benutzer ermöglicht. Was Intel sich als besonderen Clou für Geschäfte im Internet ausdachte, verdamnte der US-Kongreßabgeordnete Edward J. Markey in einem Brief an Firmenchef Craig Barrett als Bedrohung für den Datenschutz⁴⁰. Datenschutzorganisationen wie das Electronic Privacy Information Center zum Boykott von Intel-Produkten auf, bis Intel die Serialisierung wieder entfernt. Kurz nach bekannt werden der Pläne tauchte im Web folgende Mutation des Intel Logos auf:

³⁸ c't 5/1999 - Pentium-III-Seriennummer doch 'weich' einschaltbar

³⁹ ID-Nummer für Intel-Prozessoren, Florian Rötzer, Telepolis 25.01.1999
<http://www.heise.de/tp/deutsch/inhalt/te/1774/1.html>

⁴⁰ c't 3/1999 - Sicherheit kontra Datenschutz - Pentium III: Umstrittene Seriennummer



Die Intel CPU Seriennummer fand nie wirkliche Anwendung und wurde im Pentium IV wieder entfernt, da sie angeblich zur Veränderung neigte⁴¹ (die Nummer wurde bei jedem Prozessor individuell mittels Laser eingebrannt, indem kleine Sicherungen durchgebrannt wurden. Das Material der Sicherungen neigte scheinbar dazu, den entfernten Kontakt wieder herzustellen). Somit ist die Nummer weder Eindeutig noch zeitlich konstant(!!!) und auch nur bedingt tauglich um Transaktionen sicherer zu gestalten.

Doubleclicks Griff in die Keksdose

Welche Möglichkeiten durch die Verwendung von Cookies bestehen, zeigt Doubleclick (www.doubleclick.com). Wie Surfer, Bürgerrechts- und Datenschutzorganisationen darauf reagieren kann ebenfalls anhand dieses Beispiels gesehen werden. Weiters zeigt das Beispiel, dass Aktienkurse nicht unbedingt direkt proportional zum Umfang der gesammelten Userprofile steigen.

DoubleClick ist ein weltweit tätiger Adserverbetreiber und schaltet auf zehntausenden Websites Banner. Für effektives Targeting sind ausführliche Daten über die einzelnen Pages zwingend nötig, dies verlangt das Prinzip des Adserver. Die Methoden von Doubleclick an Daten zu gelangen, gerieten Anfang 2000 ins Visier der Datenschützer, da nicht einfach nur Mediadaten

⁴¹ c't 24/2000 - Bei Lichte betrachtet

gesammelt wurden, sondern umfangreiche personalisierte Profile getrackt wurden.

Da ein auf einer Kundenseite platzierter Banner direkt vom Doubleclick Server bezogen wird, hat dieser Server die Möglichkeit auf dem Clientrechner ein Cookie zu setzen. Durch die große Anzahl an Kunden (11.000 Kunden weltweit, Stand Anfang 2000, darunter viele Top Sites⁴²) ist umfangreiches Tracking möglich, da jeder Besuch einer weiteren Page anhand des bereits gesetzten Cookies zum Profil des Surfers zugefügt werden kann. Durch die große Anzahl an Kunden aus allen Bereichen, können detaillierte Userprofile gesammelt werden, da in jeder einzelnen Page Impression gezählt, einer Person zugeordnet und vor allem mit einem die einzelne Website übergreifenden Profil verknüpft wird. Durch Kenntnisse über die Inhalte der Kundenpages kann das Profil um soziografische Daten erweitert werden. Lediglich demografische Daten können nicht erhoben werden. Als Gegenwehr gegen diese Methode kann jederzeit das Cookie gelöscht werden, beziehungsweise bei manchen Browsern die Annahme von Cookies generell verweigert werden. Im Web finden sich zahlreiche Anleitungen zur Konfiguration von Proxys und Firewalls, um Banner von Doubleclick generell abzulehnen.

Dies ist aber nur ein Grund warum Doubleclick ein umstrittenes Unternehmen ist. Das Problem stellen die Summe der Details dar: so wurde zum Beispiel bekannt, dass mit dem Aufruf des Banners vom Doubleclick Server auch Formulareingaben der aktuellen Page mitübertragen werden können. So wurden beispielsweise auf den Websites von AltaVista, RealNetworks, HealthCentral, Quicken, und Travelocity Formulardaten wie Alter, Name, Geburtsdatum und e-Mailadresse ohne Wissen des Surfers an die Doubleclick Page übermittelt⁴³.

Auch auf Pages die nicht Doubleclick Banner anzeigt, kann von Doubleclick getrackt werden. Dazu ist lediglich die Kooperation des Websitebetreibers nötig. Welche Datei vom Doubleclick Server übertragen wird ist grundsätzlich nicht

⁴² Interview mit Arndt Groth, Geschäftsführer von Doubleclick Deutschland
<http://www.heise.de/tp/deutsch/inhalt/te/8645/1.html>

⁴³ Datensammler und Kundenjäger, Brigitte Zarzer, in Telepolis 18.06.2000
<http://www.heise.de/tp/deutsch/inhalt/te/8263/1.html>

relevant, so reicht es auch anstatt einem Banner ein kleines 1x1 Pixel großes Bild auf die Page zu laden. Solche Bilder können eingebaut werden, ohne vom Surfer bemerkt zu werden, der Effekt ist jedoch derselbe wie bei einem Banner: Der Surfer hinterlässt eine Spur in den Logfiles von Doubleclick. Dieses Verfahren wird als Web-Bugs⁴⁴ bezeichnet und kann ebenfalls zum Tracking von Mails benutzt werden.

Zum Absturz der Doubleclick Aktie und zur angeblichen Kündigungsdrohung des damals größten Kunden (Altavista, 20% Anteil der Adimpressions⁴⁵) führte aber die Fusion mit Abacus Alliance. Abacus ist ein Direct Marketing Unternehmen und verfügt über personalisierte Daten von 80 Mio US Haushalten (Stand Anfang 2000). Obwohl Doubleclick dementierte, die beiden Datenbestände fusionieren zu wollen, wurde bekannt, dass auf verschiedenen Servern von Doubleclick Kunden ein System installiert wurde, das Daten bei Onlinekäufen an Abacus sandte⁴⁶. Dies war zuviel für Datenschützer. Eine Fusion würde bedeuten, dass die soziografischen Profile auch demografische Daten mit Adresse und Namen erhielten. Kritiker sehen in solch einem Vorgehen den gläsernen Konsumenten realisiert.

Sowohl Datenschützer als auch das amerikanische Kartellamt gingen mit Untersuchungen oder Klagen gegen DoubleClick vor.

Am 10. Februar 2000 reichten die Verbraucherrechtler vom Electronic Privacy Information Center (www.EPIC.org) eine Beschwerde bei der Federal Trade Commission (www.FTC.com) ein. Darin wirft EPIC den Online-Werbern vor, 'gesetzeswidrig die Online-Aktivitäten von Surfern zu verfolgen und diese Aufzeichnungen mit detaillierten Personenprofilen' zu verknüpfen. Neben der FTC ermittelten auch zwei US-Bundesstaaten gegen DoubleClick. Das Center for Democracy and Technology (www.CDT.org) rief eine Online-Aktion unter dem Motto 'I will not be targeted!' ins Leben. Eine Kalifornierin reichte Privatklage gegen die Datensammler ein. Dabei ginge es ihr nicht ums Geld,

⁴⁴ Nach den Cookies die Web Bugs, Florian Rötzer, Telepolis 14.11.1999

<http://www.heise.de/tp/deutsch/inhalt/te/5482/1.html>

⁴⁵ c't 6/2000 - DoubleClicks geordneter Rückzug

⁴⁶ c't 6/2000 - DoubleClicks geordneter Rückzug

ließ sie über ihren Anwalt verlauten, sondern `darum, die unfairen und betrügerischen Handlungen der Firma´ zu stoppen.

Auf die Kritik von allen Seiten hin lenkte Doubleclick ein⁴⁷, Firmenchef Kevin O´Connor gab zu: `Der Plan, Daten von Personen mit deren Surfprofilen zu verknüpfen, war ein Fehler.´⁴⁸[58] Eine Imagekampagne mit etwa 50 Millionen Bannern und ganzseitigen Zeitungsanzeigen wurde gestartet und die Website <http://privacychoices.org> wurde in Leben gerufen. Dort haben US Websurfer die Möglichkeit per Opt-Out das Sammeln von Daten zu verbieten. Diese Möglichkeit gab es bereits zuvor, jedoch ohne ein mit 50 Millionen Banner beworbenes Portal, lediglich gut versteckt auf der Homepage von Doubleclick, konnte das Tracking unterbunden werden. Alle bisher gewonnenen Profildaten blieben dabei bestehen.

Dies reichte jedoch EPIC nicht, so verkündete man einen Tag nach bekannt werden der Offensive von Doubleclick: "Sie ändere nichts am Kernproblem. Solange die Firma auf ihrem Opt-out-Verfahren beharre, werde man weiter gegen sie vorgehen."

Opt-Out bedeutet, dass grundsätzlich Daten gesammelt werden und der Konsumenten die Möglichkeit hat dies unterbinden, indem er den Betreiber über geeignete Schnittstellen (beispielsweise ein Formular auf der Webpage) darüber informiert. Im europäischen Raum ist Opt-Out nicht zulässig, hier müssen in der Regel Opt-In Verfahren angewandt werden, der Kunde muss also explizit zur weiteren Verwendung seiner Daten zustimmen. Die dafür nötige Checkbox darf laut EU-Richtlinie auch nicht standardmäßig ausgewählt werden, sondern muss leer sein. Deshalb werden laut Doubleclick Deutschland auch im europäischen Raum keine Userprofile angelegt⁴⁹. Dies gilt natürlich nur für deutschsprachige Pages. Europäische Surfer die US-Pages besuchen werden entsprechend dem vorgehen von Doubleclick-US getrackt.

⁴⁷ DoubleClick geht in die Offensive, Brigitte Zarzer, Telepolis 05.06.2001
<http://www.heise.de/tp/deutsch/inhalt/te/7809/1.html>

⁴⁸ c't 6/2000 - DoubleClicks geordneter Rückzug

⁴⁹ Interview mit Arndt Groth, Geschäftsführer von Doubleclick Deutschland
<http://www.heise.de/tp/deutsch/inhalt/te/8645/1.html>

Die Problematik der beiden Fälle zeigt, dass Missachtung der Schutzwürdigkeit der Daten von Surfern und Kunden scharf geahndet wird. Kunden kennen den Wert ihrer Daten, sind aber auch bereit Sie zu überlassen, erwarten sich dafür aber eine Gegenleistung. Ob diese im besserem Service oder in Rabatten besteht, bleibt dem Anbieter überlassen. Gegen die anonyme Auswertung von Userprofilen und Logfiles zur Optimierung des Angebots wird kaum ein Konsument Einwände haben. Ein Verkauf dieser Daten oder gar übergreifendes Tracking stößt hingegen auf weniger Sympathie. Das Beispiel Doubleclick zeigte die Mangelhaftigkeit von Opt-Out Systemen. Jedoch können auch Opt-In Systeme durch entsprechendes Design der Privacy-Policy missbraucht werden. Versteckte Checkboxes deren Defaulteinstellungen die Weitergabe von Daten erlauben, entlocken den Kunden und Surfern relativ einfach unfreiwillig Daten. Da der Grossteil der Surfer wenig Bewusstsein über die Konsequenzen seiner Freigiebigkeit besitzt, die zugrundeliegenden Systeme komplex sind und die Verknüpfungen der Systeme kaum nachvollziehbar sind, liegt die Verantwortung für die Daten nicht mehr ausschließlich bei den Kunden. Betreiber von datensammelnden Services sollten im Interesse einer nachhaltigen Kundenbeziehung Verantwortung für die Daten ihre Kunden übernehmen.

Die künftige Problematik liegt nicht in den Daten von einzelnen Services, sondern in der Verknüpfung von Daten aus unterschiedlichen externen Quellen. Die Verknüpfung von personalisierten Geo-Daten (generierbar über Location Based Services), Einkommensinformationen, psychografischer Profile und sonstigen Konsumenteninformationen drängt den Eindruck von Orwellschen Szenarien auf.

Zu den Kunden von Doubleclick zählt beispielsweise Ortho Biotch mit dem Produkt Procrit⁵⁰. Procrit ist ein Medikament gegen Blutarmut und wird unter anderem zur AIDS und Krebs Therapie eingesetzt. Ein käufliches Profil einer Person, mit Daten, die einen Bewerber als Surfer auf der Procrit Webpage identifiziert, würde vermutlich jeden Jobscout interessieren.

⁵⁰ Datensammler und Kundenjäger, Brigitte Zarzer, in Telepolis 18.06.2000
<http://www.heise.de/tp/deutsch/inhalt/te/8263/1.html>

Ob die medizinstudierende Tochter oder die Ärztin, mit der er verheiratet ist, die Procrit Page besuchte ist für einen Personalmanager weniger relevant, da der Bewerber kaum mit dem Profil konfrontiert werden würde und somit auch keinen Grund zur Rechtfertigung erhält.

Das Internet liefert Daten und Instrumente, die mittels entsprechender Kooperationen Daten liefern, von denen das klassische Direct Marketing nur zu träumen wagte. Die Konflikte zwischen Bürgerrechtlern und Doubleclick stellen die Fragen, ob nicht auf manche der Methoden verzichtet werden sollte.

Als Richtlinie für den Umgang mit Kundendaten existieren diverse Policies und Gütesiegel wie P3P. Kurz gesagt geht es beim P3P darum, eine Spezifikation über Datenschutzstandards (formalisierte `Privacy Policies`) zu erstellen, denen sich eine Site verpflichtet. Diese Spezifikation ist somit eine Art datenschutzrechtliche Visitenkarte, die in einer für menschliche Besucher zugänglichen, aber auch in maschinenlesbarer Form vorliegen soll. Nach der derzeitigen Spezifikation muss ein Anbieter beispielsweise seine volle Anschrift nennen. Damit ist er bei Beschwerden etwa für Datenschützer erreichbar. Über einen User-Agent kann der Benutzer feststellen, welche Daten er wo gelassen hat. Das stufen Vertreter der P3P-Gruppe als einen wichtigen Schritt in Richtung informationelle Selbstbestimmung ein.

Sonstige Links und Gütesiegel:

Datenschutz Leitfaden der EU:

http://europa.eu.int/comm/internal_market/en/media/dataprot/news/guide_de.pdf

TRUSTe's Privacy Seal: <https://www.truste.org/>

P3P-Spezifikation: www.w3.org/TR/P3P/

Literaturverzeichnis

Ahrens & Behrent Agentur für Kommunikation GmbH, <http://www.a-b.de/>

Alp Uçkan, <http://magazin.usability-test.net/artikel/logfiles2.htm#oben>, 26.11.2001

Alpar, Paul: Kommerzielle Nutzung des Internet, Springer-Verlag, Berlin, 1998, S. 119

Bei Lichte betrachtet: C't 24/2000

Bossat, Felix; Haite, Sven: Internet für Unternehmer, SmartBooks Publishing AG, Kilchberg, 1999, S. 200

Chase, Larry; Hanger, Nancy C.: Essential Business Tactics for the Net, Wiley Computer Publishing, New York, 1998, S. 219

datascout (Homepage von Gerald Reischl): http://www.datascout.net/intro_frame.html

Datenbankgestützte Web-sites: Philip Greenspun, Hanser 1998

Datensammler und Kundenjäger: Brigitte Zarzer, in Telepolis 18.06.2000
<http://www.heise.de/tp/deutsch/inhalt/te/8263/1.html>

Datensammler und Kundenjäger: Brigitte Zarzer, Telepolis 18.06.2000
<http://www.heise.de/tp/deutsch/inhalt/te/8263/1.html>

Datenschutz Leitfaden der EU:
http://europa.eu.int/comm/internal_market/en/media/dataprot/news/guide_de.pdf

Deutscher Multimedia Verband:
http://www.dmmv.de/de/7_pub/homepagedmmv/themen/emarketing/media/zielemedia/messgroessenemarketing.cfm

DoubleClick geht in die Offensive: Brigitte Zarzer, Telepolis 05.06.2001

<http://www.heise.de/tp/deutsch/inhalt/te/7809/1.html>

DoubleClicks geordneter Rückzug: C't 6/2000

Exody: Lösungen für Controlling von E-Commerce-Aktivitäten,

<http://www.exody.net/ger/company/press/news.cfm?NewsID=21>

Framfab: <http://www.framfab.de/>

Gefährliche Netze: Gerald Reischl, Verlag Carl Ueberreuter, 2001

ID-Nummer für Intel-Prozessoren: Florian Rötzer, Telepolis 25.01.1999

<http://www.heise.de/tp/deutsch/inhalt/te/1774/1.html>

Informationsgemeinschaft zur Kontrolle der Verbreitung von Werbeträgern e.V.:

<http://www.ivw.de/>

infotrust: <http://www.infotrust.de/leistungen/webcontrolling.html>

integra Deutschland: http://www.integra-europe.at/dialog_studien.html

Interview mit Arndt Groth: Geschäftsführer von Doubleclick Deutschland

<http://www.heise.de/tp/deutsch/inhalt/te/8645/1.html>

Lampe, Franz: Unternehmenserfolg im Internet, Vieweg, Braunschweig,
Wiesbaden, 1998, S. 245-248

Nach den Cookies die Web Bugs: Florian Rötzer, Telepolis 14.11.1999

<http://www.heise.de/tp/deutsch/inhalt/te/5482/1.html>

National Semiconductor Corp.: <http://www.national.com/>

Online Research - Methoden, Anwendungen und Ergebnisse: Bernard Batinic/Andreas Werner/Lorenz Gräf,
Verlag Hogrefe, 1999

Pentium-III-Seriennummer doch 'weich' einschaltbar: C't 5/1999

Personenbezogene Surfprofile bei DoubleClick: C't 4/2000

Pispers, Ralf: Web-Controlling: Internet-Erfolge planen, steuern und analysieren, Artikel aus: Heinen, Irene (Hrsg.): Internet – mit E-Commerce auf dem Weg zum wirtschaftlichen Erfolg, Deutscher Internet Kongress Karlsruhe 1999, dpunkt-Verlag, Heidelberg, S. 211-221

Schwickert, A.C./Wendt, P.: Controlling-Kennzahlen für Web-Sites, Information Age Economy, Physica-Verlag, S. 653

Sari Kalin: <http://www.cio.com/>

Sicherheit kontra Datenschutz - Pentium III: Umstrittene Seriennummer: C't 3/1999

Stephan, Roland; Werner, Andreas: Marketing Instrument Internet, dpunkt-Verlag, Heidelberg, 1997, S. 171-195

Surfer im Visier: C't 20/1999

Thunder Lizard Productions: How to Analyze All That Data:

http://www.thunderlizard.com/dwnld_handout.html#webmarket

<http://www.thunderlizard.com/webmarketing.html>

http://dmoz.org/Computers/Software/Internet/Site_Management/Log_Analysis/Articles_and_Information/

<http://hotwired.lycos.com/webmonkey/e-business/tracking/tutorials/tutorial2.html>

TRUSTe's Privacy Seal: <https://www.truste.org/>

Unerkannt surfen: C't 6/2000

User Interface Engineering: <http://www.national.com/>

Viele Gesichter Personalisierte Websites stellen sich auf Besucher ein:

C't 18/2000

Vogt, Petra: Erfolgreiche Präsenz im Internet, SmartBooks Publishing AG,
Kilchberg, 1998, S. 225-234

Welche Werte haben Daten und Netze?: Stefan Krempl, Telepolis 14.05.2000

<http://www.heise.de/tp/deutsch/inhalt/konf/8143/1.html>

Werbebranche sucht Klicks: Christiane Schulzki-Haddouti, Telepolis 01.09.2000

<http://www.heise.de/tp/deutsch/inhalt/co/8648/1.html>

Werner, Andreas: Site Promotion: Werbung auf dem WWW, dpunkt-Verlag,
Heidelberg, 1998, S. 39-61

Zoellick, Bill: Web Engagement: Connecting to Customers in e-Business,

Addison Wesley Longman, New York, 2000, S. 41-55