

Anwendung eines Instruments der Online-  
forschung im Rundfunkbereich

## → „Usertracking“: Nutzungspfade im Webangebot

Von Beate Frees\*, Martin Fisch\*  
und Alexander Ebbes\*\*

**Internetpräsenz bei den meisten Unternehmen inzwischen üblich**

Die weitaus meisten Unternehmen haben inzwischen eine eigene Präsenz im Internet. Dort können Produkt- oder Dienstleistungsinformationen, Firmen- und Kontaktdaten, Zusatzinformationen – oder im Falle der öffentlich-rechtlichen Sender programmbegleitende Informationen abgerufen werden. Nicht selten gehen allerdings die Angebote an den Bedürfnissen der User vorbei: Sie bieten dem User keine Orientierung, Informationen sind wild zusammengestellt, Gesuchtes wird nicht gefunden. Bei der Erstellung von Internetangeboten sollten die Zielsetzung und die Zielgruppe bekannt sein. Die Fragen, auf welche Nutzererwartungen ein Angebot trifft, um welche Inhalte es sich handelt, welche Struktur und welches Design sinnvoll und nutzerfreundlich ist, müssen beantwortet werden.

**Nutzerverhalten beim Aufbau einer Website berücksichtigen**

Zu den Aspekten des Nutzerverhaltens, die es beim Aufbau und Betrieb einer Website zu klären gilt, gehört, auf welcher Seite des Onlineangebots der User seinen Besuch beginnt. Beginnt er ihn im Fall der Website eines Fernsehsenders wie des ZDF direkt auf der Startseite oder auf der programmbegleitenden Webseite, die in der Fernsehsendung eingeblendet wurde? Gibt es nutzungsabhängige Muster, die es ermöglichen, den User zu kategorisieren? Gibt es zum Beispiel den nachrichtenorientierten Besucher, der sich nur innerhalb des Nachrichtenangebots bewegt? Wie gut sind die Angebote (beim ZDF z.B. zdf.de, heute.de, tivi.de) untereinander verlinkt? Können über den Nutzungszeitpunkt Rückschlüsse auf den Nutzungsort gezogen werden? Und vor allem: Wie viel Zeit verbringen die User mit den jeweiligen Angeboten? Zeit ist ein begrenztes Gut, eine Steigerung des Faktors Zeit geht immer auf Kosten anderer Aktivitäten und damit ggf. auch auf Kosten anderer Medien.

**Prototyp des Usertracking beim ZDF**

Auf die Beantwortung solcher Fragen richtet sich das so genannte Usertracking, mit dem Ziel, den anonymen User bei seinen Nutzungsgewohnheiten zu beobachten. Im ZDF wurde ein Prototyp für das Usertracking entwickelt. Mit Hilfe von Workshops zwischen Onlineredaktion und Medienforschung werden Datenbestände einzelner Monate analysiert. Die Schwerpunkte und Ziele werden vorab vereinbart. Ergebnis dieser Workshops sind immer wieder auch redaktionelle Anforderungen an das

Tool. Diese Anforderungen werden umgesetzt und mit dem nächsten Workshop validiert. Die Einführung des Usertracking ist also ein iterativer Prozess mit offenem Ergebnis.

Die Ziele, die mit dem laufenden Projekt verfolgt werden, sind evident: Das Usertracking-Tool soll mit definierten Kennzahlen auf einem aktuellen und konsolidierten Datenbestand (vom Vortag), online (über einen Internetbrowser) eine flexible, durch den Redakteur selbstständig durchgeführte Analyse ermöglichen. Das heißt, dass dieses Tool weniger die Basis für ein kontinuierliches Berichtswesen, sondern vielmehr die Grundlage für Ad-hoc-Abfragen der Redaktion ist.

### Methoden der Onlineforschung – zwei unterschiedliche Ansätze

Um die Resonanz auf ein Internetangebot bewerten zu können, werden kontinuierliche und auf dem Markt vergleichbare Leistungswerte benötigt – auch und gerade dazu, um die Akzeptanz der Internetseiten im Vergleich mit den Mitbewerbern oder im Trend einschätzen zu können. Für die Akzeptanzmessung von Onlineangeboten müssen bestimmte Maßstäbe gelten. Aus diesem Grund werden Kennzahlen für die Leistungsmessung von Internetangeboten entwickelt, die mit einem einheitlichen Zählverfahren gemessen werden. Seitenabrufe (PageImpressions) und Besuche (Visits) sind solche Kennzahlen, die auf dem deutschen Markt eingeführt sind. Bei Seitenabrufen handelt es sich um „Kontakte“ mit einzelnen Seiten. Diese Informationsgröße entstand originär im Printbereich und wurde für den Onlinebereich übernommen.

Für die Akzeptanzmessung des Webangebots eines Fernsehsenders sind solche Kennzahlen nicht ausreichend, da Online als Trägermedium nicht nur Printinhalte, sondern auch Audio- und Videoinhalte transportiert. Deshalb ist es notwendig, für Online auch Kennziffern aus dem Fernsbereich zu verwenden, wie zum Beispiel die durchschnittliche Reichweite oder den Marktanteil. Nur so können Anbieter, die ihre Sendungen auf unterschiedlichen Medien anbieten, den Erfolg der einzelnen Marken medienübergreifend einschätzen. Voraussetzung hierfür sind entsprechende Kennzahlen für das Internet, die einen intermedialen Vergleich ermöglichen. „Zeit“ und „Köpfe“ wären solche Indikatoren, die in der Fernsehforschung bekannt sind und im Internet derzeit fehlen. Die Messung der Verweildauer und der Anzahl Besucher für Onlineangebote ist zur Zeit nicht möglich bzw. unbefriedigend. Mit den gängigen Zählverfahren werden momentan nur die Besuche unabhängig von der Nutzungszeit und nicht der Besucher gezählt. Ein Anwender kann pro Tag mehrmals zdf.de aufrufen und sich dort zum Beispiel über die aktuelle Nachrichtenlage informieren. In diesem Fall werden durch einen Besucher mehrere Besuche auf zdf.de gezählt. Dies entspricht der Bruttoreichweite im Fernsehen für ein bestimmtes Zeitintervall (zum Beispiel einen Tag).

Usertracking ist einer von vielen Ansätzen, um Nutzungsdaten von Onlineangeboten zu gewinnen

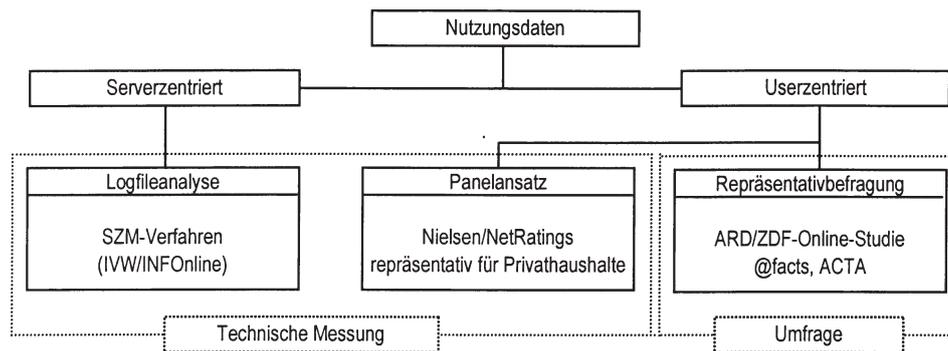
**Basis für Ad-hoc-Anfragen der Redaktion**

**Kennzahlen für Leistungsmessung im Onlinesektor**

**Webseiten von Fernsehsendern mit höheren Anforderungen an Akzeptanzmessung**

\* Medienforschung, Zweites Deutsches Fernsehen, Mainz;  
\*\* GIP Research Institute.

Abb. 1 Methoden der Onlineforschung



Quelle: ZDF.

und zu verarbeiten. Gerade das Medium Internet bietet heute sehr unterschiedliche, mitunter stark von einander abweichende Methoden zur Leistungsmessung, die für das so genannte „Finetuning“ eines Onlineangebots notwendig sind. Gefordert sind Kennzahlen, mit deren Hilfe die Angebote ständig optimiert werden können. Im Folgenden wird das Usertracking in die Methoden der Onlineforschung eingeordnet und ein Überblick über weitere Möglichkeiten der Leistungsmessung gegeben.

#### Grundsätzlich zwei Ansätze in der Onlineforschung

In der Onlineforschung werden zwei grundsätzliche Ansätze zur Erhebung von Nutzungsdaten unterschieden (vgl. Abbildung 1). Daten zur Akzeptanz von Onlineangeboten lassen sich auf Seiten des Servers (dem Anbieter eines Webangebotes) und auf Seiten des Users (dem Internetnutzer) erheben. Grundsätzlich kann die Nutzung erfragt (z.B. durch repräsentative Umfrage) oder auch technisch gemessen werden. Der Fokus dieses Beitrags liegt auf der technischen Messung als Basis für das so genannte Usertracking. Usertracking geht über die Erhebung einfacher Kennzahlen hinaus, es geht vor allem um die Nutzungspfade der User. Wie navigiert ein User, der an Nachrichten interessiert ist? Wie lange, wie viele Übergänge braucht ein User, um von der Startseite auf ein aktuelles Thema, zum Beispiel zu einer ZDF-Sendung wie „WISO“, zu gelangen? Zur Beantwortung dieser und weiterer Fragen müssen zwei Bedingungen erfüllt sein:

1. Die Daten müssen aus den Logfiles entsprechend aufbereitet und in Datenbanken abgelegt sein.
2. Die Daten müssen online zur Verfügung stehen, so dass sie dynamisch und flexibel ausgewertet werden können (Datamining-Tool).

Um den Prozess und die Methode des Usertrackings verstehen zu können, folgt zunächst eine kurze Einordnung in die Verfahren zur Nutzungsmessung.

#### Usertracking basiert auf der Logfile-Analyse – ist also eine technische Messung

Usertracking ist ein serverzentriertes Verfahren. Hier wird das so genannte Logfile eines Webseitenbetreibers als Grundlage für die Analyse verwendet. Im Logfile wird jeder Zugriff auf die Internetseite protokolliert. Jedem Anbieter von Onlineangeboten

steht sein eigenes Logfile für das hausinterne Berichtswesen oder tiefer gehende Analysen zur Verfügung. Logfile-Analysen können nach unterschiedlichen Verfahren durchgeführt werden. Ein bekanntes Analyseverfahren ist das SZM-Verfahren der INFOOnline, das ursprünglich von der IVW zur Leistungsmessung auf dem deutschen Markt entwickelt wurde. Beim SZM-Verfahren handelt es sich um ein abgestimmtes, standardisiertes Zählverfahren. Messgrößen sind die Kennzahlen: PageImpressions (Seitenabrufe) und Visits (Besuche). Problem des serverzentrierten Verfahrens ist hauptsächlich, dass nur die eigenen Logfiledaten für die Analyse zur Verfügung stehen. Ein Vergleich mit anderen Anbietern kann nur dann durchgeführt werden, wenn auch diese ihre Logfiles nach dem identischen bzw. einem vergleichbaren Verfahren analysieren.

Das IVW-Ranking nach dem SZM-Verfahren enthält derzeit die Nutzungsdaten von über 400 Onlineangeboten. Es fehlen aber zahlreiche bedeutende Onlineofferten, wie zum Beispiel Ebay oder Google, die nicht am Messverfahren teilnehmen, ebenso wenig wie die Gruppe der Nichtwerbetreiber, zu denen auch die öffentlich-rechtlichen, programmbegleitenden Onlineangebote gehören. Es ist somit keine Markttransparenz gegeben. Ein weiterer Nachteil der serverzentrierten Messung ist der, dass keine soziodemographischen Daten über die Nutzer erhoben werden. Im Logfile wird die Anfrage eines technischen Users, also eines PCs („Unique Client“), der auf das Internet zugreift, gespeichert. Es fehlen die Informationen, wer (z.B. Alter, Geschlecht) oder auch wie viele Personen sich hinter diesem so genannten Unique Client verbergen.

Der userzentrierte Ansatz setzt dagegen auf der Seite des Nutzers an. Der User wird zu seiner Internetnutzung entweder befragt oder sie wird auf seinem PC über den Einsatz einer Software gemessen. Vorab wird der User zu seinen soziodemogra-

#### Nachteile des serverzentrierten Verfahrens

#### Beim userzentrierten Ansatz ist der Nutzer mit seinen soziodemographischen Daten bekannt

phischen Merkmalen, seiner Einstellung oder auch seinem Lebensstil befragt. Damit kann die Nutzung bestimmten Zielgruppen oder auch Nutzertypologien zugeordnet werden. Eine bekannte, bevölkerungsrepräsentative Befragung zur Internetnutzung ist die ARD/ ZDF-Online-Studie, deren Ergebnisse jährlich in den Media Perspektiven veröffentlicht werden. (1)

**Nielsen/Netratings  
derzeit einziges  
Onlinepanel**

Eine andere Möglichkeit, die Nutzung auf Seiten des Users zu messen, erfolgt analog zur Fernsehreichweitenmessung über ein repräsentatives Onlinepanel. Die Panel-Teilnehmer sind in ihrer Gesamtheit ein repräsentatives Abbild der bundesdeutschen Personen ab zwei Jahren, die über einen privaten Internetzugang verfügen. Derzeit gibt es in Deutschland nur einen Anbieter eines solchen repräsentativen Onlinepanels, Nielsen//NetRatings.

**Userzentrierte  
Verfahren liefern  
keine Daten zu  
Unterangeboten und  
Einzelseiten**

Beide userzentrierten Ansätze (Repräsentativbefragung und Onlinepanel) erfassen also den Nutzer und ermöglichen es, soziodemographische Angaben zu den Usern zu erheben. Userzentrierte Verfahren besitzen aber einen Nachteil, der wiederum die Stärke des serverzentrierten Ansatzes darstellt: Sie liefern keine Leistungswerte für kleine Onlineangebote, Unterangebote bzw. Einzelseiten. Hierzu sind die befragten oder gemessenen Fallzahlen viel zu gering. Gerade die Information zur Nutzung von Unterangeboten oder Einzelseiten ist für die tägliche Optimierung eines Onlineangebotes aber wesentlich.

**Verfahren des User-  
trackings analysiert  
Nutzungsmuster**

Das Usertracking steht nun als serverzentriertes Verfahren vor der Aufgabe, die anonymen, aus Logfiles gewonnenen Daten zu sinnvollen Nutzungstypen bzw. Usertypen zusammenzuführen. Das heißt, es wird der Versuch unternommen, das Nutzungsverhalten aus den Massendaten zu schematisieren. Zusammenfassend ist für die Daten, die dem Usertracking zur Verfügung stehen, folgendes charakteristisch:

- Vollerhebung (Problem der Massendatenverarbeitung);
- hohe Genauigkeit der Daten (hohe Granularität);
- keine personenbezogenen Daten (anonyme Daten);
- keine Vergleichsmöglichkeit zu anderen Wettbewerbern, da nur Daten aus den eigenen Logfiles vorliegen.

Die Basis für das Usertracking ist der Visit (Besuch), also der Besuch zusammenhängender Seiten einer Website. Der Visit wird im Weiteren auch äquivalent als Session bezeichnet. Es handelt sich bei Visits um anonyme Nutzungsvorgänge innerhalb eines Webangebotes. Der Besucher bleibt mit seinen soziodemographischen Merkmalen unbe-

kannt. Beim Usertracking werden Einstiegs- und Ausstiegsseite eines Visits markiert. Einstiegsseiten geben zum Beispiel Aufschluss darüber, ob der Besuch auf der Startseite des Angebots oder direkt auf einer Sendungsseite (über Suchmaschine oder Eingabe der Internetadresse) begonnen hat. Darüber hinaus dient das Usertracking der Validierung der Verweildauer von Onlineangeboten. Dieser Indikator wurde bereits als wesentlich für die Akzeptanzmessung von Onlineangeboten eines Fernsehsenders bezeichnet. Da die Verweildauer auf Basis der Logfiledaten berechnet wird, handelt es sich streng genommen um die gemessene Zeit zwischen zwei Seitenabrufen eines Visits. Es wird also die Betrachtungszeit einer Webseite bis zum Aufruf der nächsten Webseite ermittelt.

Für den Vergleich mit den Mitbewerbern sind entsprechend standardisierte Zählverfahren eine Voraussetzung. Für das Usertracking im ZDF gelten zwei Grundsätze: Orientierung an standardisierten Verfahren und Transparenz bzw. Nachvollziehbarkeit des Verfahrens. Für Seitenabrufe (PageImpressions) und Visits hat sich seit einigen Jahren das SZM-Verfahren der INFOnline als Standard im deutschen Markt etabliert. Auf dem deutschen Markt existiert kein Standard für die Messung der Verweildauer nach einem serverzentrierten Verfahren. Entsprechend können Nutzungszeiten, die mit unterschiedlichen Methoden berechnet werden, deutlich voneinander abweichen. In diesem Fall gilt das Prinzip der Nachvollziehbarkeit. Entsprechend wurden unterschiedliche Verfahren zur Berechnung der Verweildauer ausprobiert, um eine Annäherung an die reale Verweildauer zu erhalten.

**Usertracking auf Basis des SZM-Verfahrens**

Das ZDF setzt seit Juli 2002 das standardisierte Zählverfahren der INFOnline (SZM-Verfahren) für die interne und externe Akzeptanzmessung ein. Für die interne, über die INFOnline hinausgehende Leistungsmessung orientiert sich das ZDF darüber hinaus an dem SZM-Verfahren. Es werden weitere Kennzahlen mit zusätzlichen Modulen entwickelt, die der Methodik des SZM-Verfahrens entsprechen. Zum Beispiel hat das ZDF als Fernsehsender andere Anforderungen an ein Zählverfahren als ein Verlag: Es werden unter anderem Bewegtbilder (Streams) im Internet angeboten, deren Akzeptanz gemessen und dokumentiert wird. Wie in anderen Unternehmen auch, die ihre Marken auf unterschiedlichen medialen Plattformen anbieten, ist eines der Ziele, diese unterschiedlichen Kennzahlen zu harmonisieren und vergleichbar zu machen, zum Beispiel die Sehbeteiligung im Fernsehen mit durchschnittlichen Kontakten im Internet oder Sichtungen von Streams. Diese Weiterentwicklungen (u.a. zur Messung von Streaming Media oder der Verweildauer) entsprechen zwar nicht einem standardisierten Zählverfahren, sind aber nach dem gleichen Verfahren (Algorithmus) aufgebaut und mit dem SZM-Verfahren verifizierbar. Dies ermöglicht es, interne Koinzidenzchecks durchführen und so die Validität der gemessenen Daten prüfen zu können. Aus diesem Grund wurde das Usertracking conse-

**Vergleich im  
Wettbewerb erfordert  
Standards**

**Anpassung des  
Zählverfahrens an  
Anforderungen eines  
Fernsehsenders**

quenterweise ebenfalls dem SZM-Algorithmus angepasst.

**Das SZM-Verfahren wird für das Data-mining nachgebildet**

Für das Datamining hat die Firma GIP Exyr GmbH das SZM-Verfahren weitgehend nachgebildet, um die Logfiledaten des ZDF aufzubereiten. Dies geschieht vor dem Hintergrund der Vergleichbarkeit und Validierung mit den aus dem SZM-Verfahren gewonnenen Daten. Der Abruf einer Seite wird als PageImpression oder Seitenabruf bezeichnet. Etwas komplizierter wird es bei der Zählung von Visits. Bei dem Abruf einer ZDF-Webseite wird ein so genannter temporärer Cookie auf dem PC eines Nutzers gespeichert. Gleichzeitig werden die mit dem SZM-Verfahren definierten Informationen über Art und Inhalt der Webseite zusammen mit der Cookie-Kennung an eine so genannte SZM-Box, die den Seitenabruf zählt, übertragen. Wird vom gleichen Nutzer eine weitere Webseite des ZDF-Angebots aufgerufen, kann über den Cookie und zusätzliche userspezifische Parameter (IP-Adresse in Kombination mit einem bestimmten Betriebssystem und einer bestimmten Browserversion) der zweite Seitenabruf eines Visits ermittelt werden. Ein Visit stellt die Gesamtzahl der PageImpressions dar, die ein Benutzer innerhalb eines definierten Zeitrahmens, ohne lange Pausen zwischen zwei Nachfragen, erzeugt. Der eigentliche Nutzer bleibt für das Messverfahren unbekannt. Zwei Visits können nicht miteinander in Verbindung gebracht werden. Entsprechend können auch keine Profile (Nutzungsmuster) für Besuchergruppen definiert werden. Darüber hinaus fehlen die Nutzungszeiten. Zur Beantwortung der hausinternen Fragen reichen also die eingeführten Kennziffern PageImpressions und Visits nicht aus.

**Weitere Kennzahlen, die über PageImpressions und Visits hinausgehen**

Ein hypothetischer Visit auf dem ZDF-Onlineangebot soll die weiterführenden Kennzahlen als Basis für das Usertracking verdeutlichen: Ein Nutzer ruft die Website zdf.de auf, von dort klickt er den Teaser zu „WISO“ an und kommt über einen für ihn interessanten Artikel auf den Themenbereich Ratgeber „Ärztliche Behandlungsfehler“. Dort lässt sich der Anwender treiben und erreicht über das Thema „Gesund leben“ die Serviceseite von „Volle Kanne Service täglich“. Danach beendet er seinen Besuch. Es fragt sich nun nach Protokollierung dieses Nutzungsvorgangs, ob der Nutzer zufrieden ist, ob er gefunden hat, was er sucht, oder ob er verirrt oder gar frustriert ausgestiegen ist. Die Antwort auf diese Fragen ist ein dynamischer Prozess mit weiteren Kennzahlen, wie zum Beispiel Verweildauer, Einstiegs- und Ausstiegsseiten und Seitenübergängen. Für jede Seite wird die durchschnittliche Schrittzahl für den Übergang von dieser Seite auf eine andere Seite berechnet. Es wird also nachvollzogen, wie viele Schritte ein Besucher im Schnitt benötigt, um von der Startseite von zdf.de zu „WISO“ zu gelangen oder von „WISO“ auf die Sendungsseite von „Volle Kanne“. Dies ergibt Indikatoren für die thematische Nähe und Tiefe der programmbegleitenden Seiten oder auch für eine erfolgreiche bzw. weniger erfolgreiche Navigation bzw. Verlinkung. Weitere Indikatoren sind die Nut-

zungszeit: Wie lange dauert im Durchschnitt ein Besuch oder wie lange verweilt der User auf den Seiten des programmbegleitenden Ratgeberangebots von „WISO“ im Internet?

Daten aus den Logfiles zu extrahieren, zu konsolidieren und kennzahlenrelevant aufzubereiten, ist in der Regel ein einmaliger Aufwand und stellt kein Problem dar. Aufwändig ist die Konzeption des Datenmodells und die anschließende Interpretation der Daten. Für die Interpretation der beschriebenen Kennzahlen ist das Wissen über die Datenstruktur eine Grundvoraussetzung. Deshalb werden zu Beginn, also zur Einführung eines solchen Instruments, Workshops mit Spezialisten (Dienstleister, Medienforschung und Onlineredaktion) organisiert und ad-hoc am System praxisbezogene Fragen gestellt, analysiert und beantwortet. Dies setzt Expertenwissen voraus: Wissen über die Inhalte des Onlineangebots, über die Verlinkung der einzelnen Seiten und die gespeicherten Kennzahlen.

Die eigentliche Auswertung, also die flexible Analyse der Kennzahlen, erfolgt online mit einem so genannten Datamining-Tool. (2) Datamining verfolgt das Prinzip, aus riesigen Datenbankbeständen (Massendaten) Kundeninformationen oder wie in diesem Fall Nutzungsstrukturen zu strategischen Zwecken herauszufiltern. Die Analyse erfolgt also dynamisch und online am PC, da alle möglichen Auswertungsvarianten nicht in Berichtsform abgespeichert werden können. Darüber hinaus kann ein solches Dataminingsystem auch für einen konsolidierten Datenbestand genutzt werden und als Online-Berichtswesen der Geschäftsführung zur Verfügung gestellt werden. Das ZDF verwendet das Datamining-Tool von GIP „Xyna Experimentierplattform“.

Zu Beginn des Workshops wurden die aufbereiteten Daten des Datamining-Tools mit den Leistungswerten des standardisierten Zählverfahrens INFOnline validiert. Die ZDF-Angebote: zdf.de, heute.de, tivi.de, 3sat.de und phoenix.de erreichten im November 2004 rund 83 Millionen PageImpressions nach INFOnline. Die aufbereiteten Daten des Instituts GIP beliefen sich im November 2004 auf rund 86 Millionen Seitenabrufe und neun Millionen Sessions bzw. Visits. Die Differenz zwischen INFOnline und GIP beträgt demnach rund drei Millionen Seitenabrufe (+3,6%). Dies erklärt sich durch den Algorithmus des SZM-Verfahrens, der nicht im Detail abbildbar ist. Darüber hinaus enthält die Summe der Gesamtseitenabrufe nach INFOnline und dem Datamining nach GIP Nutzungsdaten, die für das userbezogene Datamining zu bestimmten Inhalten nicht wesentlich sind. Einige Aspekte sind für die Analyse nicht relevant, wie zum Beispiel hausinterne Zugriffe, Downloads oder überlange Sessions, die durch Robots (Suchmaschinen, die im Netz nach Schlagworten suchen) erzeugt werden. Diese

**Expertenwissen wird in Workshops vermittelt**

**Auswertung per Datamining-Tool**

**Prüfung der Datenvalidität und Reduktion auf userbezogene Zugriffe zu bestimmen Themen**

Zugriffe werden für das Usertracking herausgefiltert. Die Zielsetzung ist dabei, so nah wie möglich an typische userbezogene Abrufe zu gelangen, die sich auf konkrete Inhalte beziehen. Im November 2004 wurden entsprechend 59 Millionen von 86 Millionen Seitenabrufen näher betrachtet.

**Datamining-Tool bietet vielfältige Auswertungsmöglichkeiten**

Die Fragen der Redaktion an das Datamining-System sind sehr vielfältig und offerieren viele Möglichkeiten. Interessant ist zum Beispiel die Verlinkung zweier Angebote. Es könnte ein Ziel sein, die Verlinkung zwischen zdf.de und tivi.de zu verbessern. Hierbei kann überprüft werden, ob der Teaser von zdf.de auf tivi.de funktioniert. Darüber hinaus ist für den Onlineauftritt des ZDF wichtig, die programmbegleitenden Sendungsangebote im Internet zu optimieren und die Vernetzung untereinander zu gewährleisten: Bietet das Onlineangebot Orientierung? Wird die Struktur erkannt? Wie sucht der User die Inhalte: Über die Sendung oder das Thema?

**Programmbegleitendes Onlineangebot zu „Bianca“ als Beispiel für die Analyse eines Sendungsangebots**

**„Bianca“: Im ersten Monat vier Millionen Seitenabrufe**

Am anschaulichsten vermitteln Ergebnisse einer Sendungsanalyse das Prinzip des Dataminings. Am 1.11.2004 startete die erste deutsche Telenovela im ZDF, dies war der Anlass, das programmbegleitende Angebot zu „Bianca“ im Netz näher zu betrachten. „Bianca“ erreichte im November rund 4,15 Millionen Seitenabrufe der Besucher. Die Abrufe nach dem SZM-Verfahren von INFOnline stiegen danach kontinuierlich, im März 2005 erreichte das Onlineangebot zu „Bianca“ rund 15,6 Millionen Page-Impressions. Zum Vergleich: Das gesamte Onlineangebot des ZDF erzielte im März 2005 über 106 Millionen Seitenabrufe und elf Millionen Visits.

Für das Usertracking ist jeder Visit innerhalb des Onlineangebots mit seiner Einstiegs- und Ausstiegsseite markiert. Es lässt sich daher genau ermitteln, wie viele Sessions auf den Onlineangeboten des ZDF begonnen wurden (beliebiger Einstieg), ob während der Session mindestens eine Seite des Onlineangebots von „Bianca“ aufgerufen wurde und auf welcher Seite die Session innerhalb des ZDF-Angebots beendet wurde (beliebiger Ausstieg). Dies lässt sich im Weiteren konkretisieren. Wie viele Sessions beginnen im ZDF-Angebot (beliebiger Einstieg) und enden innerhalb des „Bianca“-Onlineangebots? Mit Hilfe des Datamining-Tools kann diese Frage beantwortet werden. Hierzu einige Beispiele:

**Verschiedene Zugänge zum Onlineangebot „Bianca“**

Rund 3 Prozent (entspricht 167 000) von 5,2 Millionen Besuchen von zdf.de kommen nicht über die Startseite von zdf.de oder einem Link auf zdf.de, sondern rufen direkt die Angebotsseiten zu „Bianca“ auf: www.Bianca.zdf.de. Diese Information kann aus der Anzahl Sessions gewonnen werden, die innerhalb des Sendungsangebots beginnen. Es lässt darauf schließen, dass „Bianca“-Fans ihre „Bianca“-Seite direkt aufrufen oder sie als Favoriten in ihrem Browser abgespeichert haben. Denkbar ist

auch, dass ein Teil der Besucher über Suchmaschinen auf „Bianca“ gelangt. Es lässt sich aber auch die Anzahl der Sessions ermitteln, die auf der Startseite von zdf.de (www.zdf.de) beginnen und direkt im Angebot von „Bianca“ fortgeführt werden (über einen Link, einen Teaser). Dieser Zugang wird von rund 1,9 Prozent (rund 100 000 direkte Übergänge) der zdf.de-Besucher genutzt. Das heißt, dass die meisten Besuche von „Bianca“ direkt auf dem Onlineangebot „Bianca“ beginnen oder über die Startseite auf zdf.de erfolgen.

Die Informationsgrößen lassen sich mit dem Datamining-Tool natürlich auch für einzelne Tage (Trends) oder auch Stunden detaillieren. Zur Interpretation der Daten sind in einigen Fällen Zusatzdaten (Metadaten) notwendig. Im konkreten Beispiel wird untersucht, ob der Besuch auf „Bianca“ über die zdf.de-Startseite oder direkt im Sendungsangebot beginnt. In diesem Fall müssen Metadaten zur Interpretation herangezogen werden, um Informationen über den Aufbau der zdf.de-Startseite (prominenter Teaser zu „Bianca“) und eventuelle Einblendungen zum Onlineangebot in der Fernseh-sendung zu erhalten. Mit Datamining können also auch crossmediale Effekte betrachtet werden. Neben den Einstiegsseiten (Beginn der Session) werden auch die Ausstiegsseiten (Ende der Session) untersucht: Wie ist der Nutzungspfad nach dem Besuch der sendungsbegleitenden Seiten zu „Bianca“? Gibt es weitere Sendungsbereiche, die ein typischer „Bianca“-Besucher nutzt?

Auffällig ist, dass ein hoher Anteil der Besucher von „Bianca“ auch innerhalb des Angebots bleibt. Dies lässt darauf schließen, dass das Angebot mit seinen umfangreichen und interaktiven Inhalten in sich geschlossen ist. Darüber hinaus funktioniert die Vernetzung zu anderen Sendungsangeboten, die Crosspromotion zu „Bianca“ betreiben. Überdurchschnittlich viele direkte Übergänge lassen sich dementsprechend zu folgenden Sendungsformaten finden: „Hallo Deutschland“, „Johannes B. Kerner“ und „Wetten, dass..?“ Interessant ist, dass auch überdurchschnittlich hohe direkte Übergänge von inhaltlich ähnlichen Formaten zu erkennen sind. Im Fall von „Bianca“ trifft dies besonders auf das Format „Reich und Schön“, eine tägliche Serie aus dem Vormittagsprogramm des ZDF, zu.

Die Verweildauer auf den einzelnen Seiten des „Bianca“-Onlineangebots ist im Durchschnitt etwas geringer als die durchschnittliche Verweildauer auf zdf.de-Seiten. Dies erklärt sich durch die häufig genutzten und mit Inhaltsangaben aufwändig gestalteten Bildergalerien zu den Einzelfolgen sowie das interaktive Modul „Bianca's Welt“. Insbesondere umfangreiche Bildergalerien senken die Verweildauer eines Angebots, da sie im Verhältnis zu Text weniger Zeit beanspruchen.

An Werktagen werden die Seiten zu „Bianca“ fast doppelt so häufig abgerufen wie an Wochenenden. Dies entspricht im Wesentlichen dem werktäglichen Ausstrahlungsrhythmus von „Bianca“ im Fernsehen.

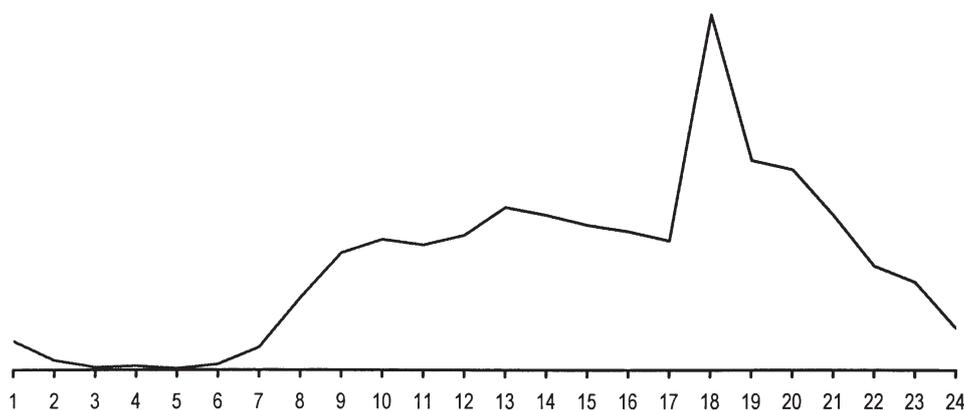
**Auch crossmediale Effekte werden beobachtet**

**Datenanalyse gibt Aufschluss über die Nutzung vernetzter Angebote**

**Bildergalerien senken durchschnittliche Verweildauer**

**Nutzung der „Bianca“-Webseite zeitlich eng verbunden mit TV-Ausstrahlung**

Abb. 2 Stundenverlauf der Seitenabrufe zu dem programmbegleitenden Onlineangebot „Bianca“ am 16.11.2004



Quelle: ZDF.

Die Verweildauer unterscheidet sich jedoch nicht an Werk- oder Wochenendtagen. Der Stundenverlauf zu „Bianca“ zeigt die höchste Nutzung am Tag zwischen 17 und 19 Uhr (vgl. Abbildung 2). Im ZDF wird „Bianca“ in der Zeit von 16.15 bis 17.00 Uhr ausgestrahlt. Die Onlinenutzung des programmbegleitenden Angebots erhöht sich also direkt nach der Ausstrahlung im Fernsehen. Die Nutzung von „Bianca“ im Internet erfolgt hauptsächlich zwischen 9 und 22 Uhr. Nach Ausstrahlung im Fernsehen verdoppelt sich die Nutzung und bleibt bis 20 Uhr auf relativ hohem Niveau. Dies ist zum einen mit dem Hinweis auf das Onlineangebot während der Fernsehsendung und zum anderen mit aktuellen Highlights, wie zum Beispiel Chats mit den Schauspielern oder dem interaktiven Modul „Bianca's Welt“, welches täglich den aktuellen Folgen inhaltlich angepasst wird, zu begründen. Das Angebot ist also aktuell und dynamisch auf den „Bianca“-Fan zugeschnitten.

#### Fazit

Ein Onlineangebot muss sich mit seiner Zielsetzung an den Bedürfnissen seiner Nutzer orientieren. An wen sich ein Webangebot richtet und welche Nutzererwartungen durch die Inhalte erfüllt werden, muss ständig überprüft und im Sinne der jeweiligen Zielsetzung optimiert werden. Um diesem Anspruch gerecht zu werden, sind eine klare Struktur sowie eine kontinuierliche Optimierung des Angebots erforderlich. Diesem Ziel dienen aussagekräftige Nutzungsdaten, die den Vergleich mit anderen Angeboten zur Validierung der Leistungswerte ermöglichen und darüber hinaus das Verfahren für weitere Leistungswerte offen legen. Zur Beantwortung der vielfältigen Fragen der Redaktionen in Fernsehsendern mit umfangreichem Webangebot reicht das standardisierte, im Markt eingeführte Auswertungsverfahren nicht aus. Für das operative Geschäft eines Website-Betreibers wie beispielsweise des ZDF sind weit mehr Informationen notwendig, die zum Beispiel eine Logfileanalyse liefern kann. Jeder Zugriff (Vollerhebung) auf eine Internetseite wird im Logfile protokolliert. Die Informationen liegen also vor, sie müssen nur

entschlüsselt und zur Verfügung gestellt werden; auf diesem Prinzip basiert das Usertracking.

Dadurch kann zum Beispiel beim Webangebot eines Fernsehsenders festgestellt werden, ob der Besucher seinen Visit auf der Startseite oder direkt auf einer programmbegleitenden Seite beginnt, zu der er über eine Suchmaschine oder über einen Favoriten gelangt ist. Weitere Kennzahlen sind die Anzahl der Übergänge zwischen zwei Seiten oder die Verweildauer einer Seite oder eines Besuchs. Die individuelle und flexible Auswertung dieser Kennzahlen unterstützt den Onlineredakteur bei seinem operativen Geschäft, indem ihm zum Beispiel Informationen zu Designentscheidungen ad-hoc zur Verfügung stehen. Er erhält Hinweise über die Positionierung von Teasern oder die Verlinkung von Sendungsangeboten. Darüber hinaus lassen sich alle Kennzahlen für einzelne Tage oder auch Stunden darstellen, um Trends oder crossmediale Effekte (Verknüpfungen zwischen Fernsehen und Internetangebot) zu beobachten. Hierfür sind zur Interpretation der Kennzahlen teilweise Metainformationen über die Fernsehsendung notwendig, über die der Onlineredakteur verfügt: zum Beispiel, wann die Sendung zum Onlineangebot ausgestrahlt wurde oder welche Hinweise im Fernsehen auf das Onlineangebot erfolgten. Das beim ZDF in Erprobung befindliche Usertracking-Verfahren erlaubt damit vielfältige Analysemöglichkeiten zur täglichen Optimierung des Internetangebots des Senders.

#### Anmerkungen:

- 1) Zuletzt von Eimeren, Birgit/Heinz Gerhard/Beate Frees: Internetverbreitung in Deutschland: Potenzial vorerst ausgeschöpft? ARD/ZDF-Online-Studie 2004. In: Media Perspektiven 8/2004, S. 550–570.
- 2) Datamining ist die Analyse und Auswertung von Daten mittels Verfahren und Techniken, welche es erlauben, aus sehr großen Datenmengen bisher unbekannte Zusammenhänge, Muster und Trends zu erkennen. Datamining gewinnt im Marketingbereich zunehmend an Bedeutung. Anhand der Interpretation von Nutzerprofilen (Alter, Geschlecht, Surfgewohnheiten usw.) lassen sich äußerst wirksame Werbestrategien entwickeln und Marktsegmente bestimmen.

