

2. Grundgesamtheit und Stichprobenziehung

STEFAN MEIER/CARSTEN WÜNSCH/
CHRISTIAN PENTZOLD/MARTIN WELKER

Auswahlverfahren für Online-Inhalte

Auswahlverfahren in der empirischen Sozialforschung

Empirische Untersuchungen werden durchgeführt, um systematisch und nachvollziehbar zu Aussagen über die Realität bzw. soziale Praktiken zu gelangen. Dies gilt gleichermaßen für Untersuchungen, die eher wissenschaftlichen oder eher praktischen Fragestellungen dienen, und auch unabhängig von der eingesetzten Erhebungsmethode. Die Fragestellungen, die dabei verfolgt werden, beziehen sich häufig auf eine große Anzahl von Elementen bzw. einen großen Ausschnitt der Realität. Der Untersuchungsgegenstand, über den bei einer Online-Inhaltsanalyse Aussagen gemacht werden, sind Online-Inhalte. Und auch hier gilt: Oft geht es um Aussagen über eine Vielzahl von Online-Angeboten, bspw. über die Gesamtheit aller deutschsprachigen Blogs, über die Homepages aller kleinen und mittleren Unternehmen in Deutschland oder über Foren zu gesundheitsbezogenen Themen.

Um zu belastbaren Aussagen über eine große Anzahl an Elementen (die Grundgesamtheit) zu kommen, stehen prinzipiell zwei Strategien zur Verfügung:

- (1) Man untersucht tatsächlich all diese Elemente in Bezug auf die interessierenden Merkmale oder
- (2) man untersucht einen Teil der Elemente und greift anschließend auf eine Regel zurück, mit deren Hilfe man von den Ergebnissen der Teiluntersuchung auf alle Elemente schließen kann.

Es ist offensichtlich, dass die erste Strategie zwar scheinbar¹ zu sichereren Ergebnissen führt – schließlich wurde jedes Element, über das man

1 Je umfangreicher Untersuchungen sind, desto häufiger und wahrscheinlicher tre-

Aussagen macht, auch untersucht. Gleichzeitig ist aber ein solches Vorgehen aus Zeit- und Kostengründen und aufgrund der fehlenden Erreichbarkeit aller Elemente oft nicht umsetzbar.

Durch den technischen Fortschritt und die gestiegenen Verarbeitungskapazitäten von Computern ist seit Mitte des Jahrzehnts nun ein Trend zu beobachten, größere Korpora frei verfügbarer Texte, Bilder und ganz allgemein Daten analysieren zu wollen. Dies gilt insbesondere für das der Online-Inhaltsanalyse verwandte Verfahren der automatisierten, computergestützten Textanalyse, die oftmals auf textstatistischen Verfahren basiert (vgl. den Abschnitt 6 zur statistisch-automatisierten Inhaltsanalyse). Diese statistischen Verfahren sind oftmals umso besser, je größer die Korpora ausfallen. Diese können beispielsweise viele Millionen Wörter umfassen (vgl. ZÜLL 2009). Allerdings weisen die Verfahren (noch) Schwächen auf, wenn es um die Feststellung inhaltlicher Kategorien bzw. komplexerer Konstrukte oder Konzepte sowie impliziten Bedeutungen (wie Gewaltbereitschaft, Ausländerfeindlichkeit, Politikstile etc.) geht. Daher spielt die Teilerhebung – die Untersuchung einer Stichprobe – eine zentrale Rolle in der empirischen Forschung. Das Grundprinzip aller Stichprobenverfahren besteht darin, bei der Ziehung der Stichprobe – also der Auswahl der zu untersuchenden Elemente der Grundgesamtheit – so vorzugehen, dass diese die Grundgesamtheit repräsentieren. Repräsentativität ist somit neben Genauigkeit und Präzision (vgl. ADM 1999: 27ff.) die zentrale Anforderung an Stichproben. Bei der Ziehung der Stichprobe sind wiederum zwei grundlegend verschiedene Vorgehensweisen zu unterscheiden.

(2a) Zufallsbasierte Verfahren stützen sich auf die Regeln der Wahrscheinlichkeitsrechnung. Hat jedes Element der Grundgesamtheit die gleiche (von null verschiedene) Wahrscheinlichkeit, in der Untersuchung berücksichtigt zu werden, genügt eine hinreichend große Anzahl an untersuchten Elementen, um mit einer hohen Wahrscheinlichkeit auf die ›wahren‹ Ergebnisse in der Grundgesamtheit zu schließen. Sowohl die Stichprobengröße als auch die Wahrscheinlichkeit, mit der das Ergebnis auch auf die Grundgesamtheit zutrifft,² lässt sich mathematisch nach

ten auch Fehler auf, desto größer ist die Chance, dass sich der Untersuchungsgegenstand durch die Untersuchung selbst (im Sinne von Reaktivität) und innerhalb des Untersuchungszeitraums – der bei großen Untersuchungen i.d.R. länger ausfällt – verändert. Aus diesen Gründen sind Vollerhebungen nicht zwangsläufig genauer als Teilerhebungen.

- 2 Genau genommen wird die Wahrscheinlichkeit berechnet, mit der man sich beim Schluss auf die Grundgesamtheit irrt.

den Regeln der Wahrscheinlichkeitsrechnung bestimmen. Für dieses Vorgehen muss vorausgesetzt werden, dass zum einen die Grundgesamtheit erreichbar ist und zum anderen weiterhin eindeutig definiert ist, was als *ein* Element der Grundgesamtheit betrachtet werden soll. Beides muss gegeben sein, um sicherzustellen, dass alle Elemente die gleiche (von null verschiedene) Auswahlwahrscheinlichkeit haben.

(2b) Die Gruppe der nichtzufälligen Verfahren bedient sich sehr verschiedener Argumentationen und Strategien, um eine Grundgesamtheit möglichst repräsentativ abzubilden.³ Die meisten greifen dabei zumindest implizit auf a priori vorhandene Informationen über die Grundgesamtheit wie bspw. deren Struktur bezüglich relevanter Merkmale zurück. Es werden dann systematisch solche Elemente in die Stichprobe aufgenommen, welche für die bereits bekannten Eigenschaften der Grundgesamtheit ›typisch‹ sind. Diese Strategie wird entweder auf jedes Element einzeln angewandt (z. B. bei der Auswahl typischer Fälle) oder auf die Aggregatmerkmale der Stichprobe (bspw. das Quotenverfahren). Voraussetzung für solcherlei Verfahren ist demnach die Kenntnis der relevanten Eigenschaften der Grundgesamtheit. Der Nachteil solcher Verfahren ist, dass dabei nicht wie bei den wahrscheinlichkeitsbasierten Auswahlen Repräsentativität sichergestellt werden kann, sondern nur als Zielgröße angestrebt wird. Dies fällt umso mehr ins Gewicht, da die Repräsentativität solcher Stichproben eng mit der Güte der Vorinformationen über die Grundgesamtheit verbunden ist und mit der Entscheidung darüber steht und fällt, was typische Fälle und was relevante Auswahlmerkmale sind. Diese theoriebasierten Entscheidungen der Forscher sind nur schwer objektivierbar. Meier und Pentzold (in diesem Band) schlagen bspw. ein Prozedere vor, welches auch bei solchen Auswahlverfahren ein möglichst hohes Maß an Systematik und Nachvollziehbarkeit garantieren soll.

Wir können soweit festhalten: Die Beschränkung auf Stichproben ist beim empirischen Zugang zu einer großen Menge an Untersuchungselementen unumgänglich. Obwohl hierbei nur ein Teil der Elemente tatsächlich untersucht wird, ist es dabei das Ziel, Aussagen über die Grundgesamtheit machen zu können. Voraussetzung für eine gute – repräsentative – Stichprobe ist die Kenntnis und die prinzipielle Erreichbarkeit der

3 Nichtzufällige Auswahlverfahren, welche Repräsentativität nicht anstreben, berücksichtigen wir an dieser Stelle nicht.

Grundgesamtheit. Bei zufallsbasierten Auswahlverfahren wird zudem die trennscharfe Abgrenzung einzelner Elemente der Grundgesamtheit notwendig, während bei nicht zufallsbasierten Verfahren i.d.R. Vorabkenntnisse über relevante Merkmale der Grundgesamtheit erforderlich sind.

Auswahlverfahren der Inhaltsanalyse

Inhaltsanalysen liefern keine Einsichten über den Einzeltext, sondern verallgemeinerbare Aussagen basierend auf der systematischen Analyse von Textmengen. Sie zielen auf die Herausarbeitung interessierender Muster ab, womit ein Informationsverlust einhergeht. Systematisch ist ihr Vorgehen, wenn eine konkrete Forschungsstrategie strukturiert und invariant angewendet sowie offengelegt wird, damit die Ergebnisse intersubjektiv nachvollziehbar sind (vgl. FRÜH 2007: 39f.). Während die Inhaltsanalyse zu Beginn ihrer theoretischen und methodischen Ausarbeitung noch als Forschungstechnik verstanden wurde (vgl. BERELSON 1952), wird sie nunmehr als eigenständige Methode angesehen (siehe RÖSSLER 2005: 18), die aus einzelnen Schritten besteht. So kann die Auswahl des Materials selbst auf verschiedene »empirische Erhebungstechniken« (FRÜH 2007: 28), »Verfahren« (MERTEN 1995: 48) bzw. Formen der »Generierung von Untersuchungsmaterial« (WOLLING/KUHLMANN 2003: 136) zurückgreifen. Früh (2004: 137ff.) verortet die Auswahl im Grundmodell der Themen-Frequenzanalyse in der Planungsphase des Forschungsprozesses. Demnach werden zunächst die Problemstellung sowie theoriegeleitete Hypothesen aufgestellt und in einem zweiten Schritt die entsprechende Grundgesamtheit festgelegt und ggf. eingegrenzt.⁴ Der noch zur Grundgesamtheitsdefinition gehörende Schritt der Bestimmung der Auswahlinheiten (*sampling units*) wird von Rössler (2005: 50ff.) als fünfstufiges Vorgehen vorgestellt, in welchem der Zeitraum, der räumliche Geltungsbereich, die Mediengattungen (Print, Fernsehen, Internet etc.), die Medienangebote und die Ressorts/Formate berücksichtigt werden sollen. Eine zusätzliche Eingrenzung kann sich zudem aus der inhaltlichen Ausrichtung der Forschungsfrage ergeben.

4 Gehrau/Fretwurst (2005: 19) unterteilen diesen Schritt nochmals: die Festlegung der Untersuchungsobjekte (die Elemente), der Menge aller untersuchbaren Untersuchungsobjekte (Auswahlgesamtheit) und der Menge aller Untersuchungsobjekte, über die Aussagen gemacht werden.

Darüber hinaus ist es möglich, das Material der Grundgesamtheit mittels verschiedener Auswahltechniken auf einen analysierbaren, den Forschungsressourcen (Zeit, Geld, Ausstattung) und Erkenntnisinteressen gemäßen Korpus zu reduzieren. Das zuverlässigste Auswahlverfahren ist, wie erläutert, die Zufallsstichprobe, die unterschieden werden kann in einfache, geschichtete und Klumpenstichproben. Wird eine einfache zufallsbasierte Auswahl durch Umstände erschwert, so bietet sich zudem die systematische Zufallsauswahl an, bei welcher nach einer festgelegten Systematik, die aber unabhängig von den zu untersuchenden Merkmalen sein sollte, das Material ausgewählt wird (z. B. Random Walk, Last Birthday, jeder n-te Artikel einer Zeitungsseite). Neben den Zufallsauswahlen finden sich bei Inhaltsanalysen darüber hinaus bewusste Auswahlen, die zwar anders als willkürliche ebenfalls auf Auswahlkriterien beruhen, doch hinsichtlich der Repräsentativität hinter den Zufallsauswahlen zurückfallen (vgl. MERTEN 1995: 283ff.). Beide aber – die Zufalls- und die bewusste Auswahl – streben eine Verallgemeinerbarkeit der in der Stichprobe beobachteten Werte auf die Grundgesamtheit an (vgl. KOPS 1977: 97). Bei letzterer ist allerdings ein statistischer Repräsentationschluss nicht möglich (vgl. WOLLING 2005: 140). Für bewusste Auswahlen werden weitere Möglichkeiten unterschieden: die Bestimmung typischer Fälle, extremer Fälle (mit dem Spezialfall des Schneeballprinzips), die Auswahl nach dem Konzentrationsprinzip (Stichprobenziehung nur in einem definierten Teil der Grundgesamtheit) und die Quotenstichprobe (Auswahl gemäß Quoten in der Grundgesamtheit). Außerdem können die verschiedenen Verfahren in mehrstufigen Prozessen auch miteinander verbunden werden.

Zwar ist die Wahrscheinlichkeits- bzw. Zufallsauswahl im Blick auf die Repräsentativität die anzustrebende Erhebungstechnik, doch bilanziert Merten (1995: 300), dass in der Praxis zeitliche, technische und ökonomische Gründe diese selten zulassen würden. Konsequenterweise sieht er bewusste Auswahlen dem Motto »So gut es geht« (ebd.) folgend in der Überzahl. Nicht selten würden sogar keine Angaben zur Sample-Bildung gemacht. Angesichts dessen stimmt er Kops' (1980) Einsicht zu, dass man als Alternative zur Zufallsauswahl auf Strategien zur Absicherung der Qualität bewusster Auswahlen zurückgreifen sollte. Einen ähnlichen Schluss ziehen Gehrau und Fretwurst (2005). Zum einen argumentieren sie, dass eine eingehendere Diskussion zu Auswahlverfahren in den einschlägigen Einführungsbänden unterbleibt. Zum

anderen zeigen sie auf der Basis einer Meta-Inhaltsanalyse inhaltsanalytischer Studien, die 2002/03 in den Fachzeitschriften *Publizistik* und *Medien und Kommunikationswissenschaft* sowie in den Tagungsbänden der DGPK-Jahrestagungen veröffentlicht wurden, dass zwei Drittel (66 %) aller dortigen Studien auf bewusste Auswahlverfahren zurückgriffen und nur 13 Prozent eine Zufallsauswahl bzw. 21 Prozent eine Vollerhebung vornahmen. Die bewussten Auswahlen wiederum orientierten sich in 60 Prozent der Fälle an Merkmalen, in 26 Prozent an einem Spektrum (z. B. das politische Spektrum des Medienangebots) und in 14 Prozent wurde eine willkürliche Auswahl vorgenommen. Außerdem machten die untersuchten Inhaltsanalysen zu 59 Prozent der Fälle keine Angaben über die Grundgesamtheit, und 27 Prozent ließen Informationen zum Umfang der Auswahl vermissen, was den intersubjektiven Nachvollzug erschwert oder unmöglich macht. Eine explizite/implizite Begründung des gewählten Vorgehens wurde nur in knapp einem Viertel (26 %) der Studien geliefert (vgl. FRETWURST/GEHRAU/WEBER 2005; ähnl. RIFFE/FREITAG 1996).

Spezifika von Online-Inhaltsanalyse

Wie deutlich wurde, liegt die Stärke inhaltsanalytischer Arbeiten in ihrem expliziten Ziel, eine »große Zahl von Botschaften« (RÖSSLER 2005: 16) nach intersubjektiv nachvollziehbaren Kriterien formal und inhaltlich zu erschließen, um verallgemeinerbare Aussagen z. B. über deren mediale Quellen, zeitliche Transformationen sowie deren soziale Eingebundenheit treffen zu können. Für die Analyse medialer Inhalte liefert dieses Verfahren einen systematischen Zugriff auf die Fülle, Heterogenität und Komplexität bestehender Medienprodukte. Bei der Auswahl massenmedialer Botschaften konnte und kann die Inhaltsanalyse jedoch häufig auf eine ritualisierte Publikationspraxis (Periodizität und Parallelität) sowie auf eine systematische Archivierung der Medienprodukte durch die Sender, Verlagshäuser oder Bibliotheken zurückgreifen. So gewährleistet die Institutionalisierung der Medienanbieter auch die Strukturierung und Zugänglichkeit des Untersuchungsmaterials. Daran anschließend können sich Forschende in der Bestimmung der Auswahleinheiten an mehr oder weniger institutionalisierte Periodizität sowie genre- und formatspezifische Kriterien halten. In der inhaltsanalytisch

orientierten Online-Forschung erscheinen durch die komplexen medialen Spezifika des Untersuchungsgegenstandes die Auswahl- und Analysepraktiken jedoch eine notwendige Neujustierung erfahren zu müssen. Diese Spezifika wurden in zahlreichen kommunikationswissenschaftlichen Arbeiten (vgl. z. B. RÖSSLER 1997, 2001; RÖSSLER/WIRTH 2001; SEIBOLD 2002; WOLLING/KUHLMANN 2003; FRAAS/PENTZOLD 2008; MEIER 2008) beschrieben. Sie und die ggf. daraus folgenden Probleme für die Auswahl des Untersuchungsmaterials lassen sich schlagwortartig wie folgt zusammenfassen (vgl. WELKER et. al. in diesem Band):

Flüchtigkeit, Dynamik und Transitorik der Inhalte: Online-Inhalte werden kontinuierlich neu erstellt, verändert oder gelöscht. Damit geht auch einher, dass sich die Grundgesamtheit kontinuierlich verändert und sich zeitlich z. B. als ein statischer ›Schnappschuss‹ definieren lässt. Diese Grundgesamtheit ist demnach in der so definierten Form nur für einen kurzen Zeitraum verfügbar. Weiterhin können Schwierigkeiten bei der Abgrenzung verschiedener Angebote (und somit Untersuchungselemente) entstehen, da geklärt werden muss, ab welchem Veränderungsgrad von einem neuen Angebot gesprochen werden kann.

Medialität, Multimedialität bzw. Multimodalität: Durch die Kombination und Integration verschiedener medialer Zeichenträger (z. B. PDFs, Videos, Fotostrecken), Zeichenmodalitäten bzw. -systeme (z. B. Bilder, Filme, Sprachtexte) und Kommunikationsformen (z. B. Chat, Foren, Websites) entstehen komplexe multikodierte Inhalte.

Nonlinearität/Hypertextualität: Textcluster, Webseiten und Websites sind durch Verlinkungen verbunden und können kommunikative Einheiten bilden. Bei der Definition der auszuwählenden Elemente (Auswahleinheiten) steht man daher vor der Aufgabe, ein einzelnes Element als solches eindeutig und einheitlich zu definieren.

Reaktivität und Personalisierung: Aufgrund der aktuellen Generierung von Inhalten, die (auch) interaktiv auf Nutzereingaben basieren und so quasi personalisierte Angebote entstehen lassen, sind Online-Inhalte als reaktiv zu charakterisieren. Unter Umständen werden personalisierte Angebote nur im Rahmen einer einzigen, einmaligen Interaktion angeboten. Eine Definition der Grundgesamtheit muss sich in solchen Fällen auf die Realisierungsoption eines individuellen Angebots beziehen. Problematisch ist dabei, dass nur unter genauer Kenntnis der eingesetzten Interaktionsalgorithmen der Umfang aller und die Realisierungschancen einzelner Optionen abschätzbar sind.

Digitalisierung/Maschinenlesbarkeit: Online-Inhalte liegen in einer digitalen oder digitalisierten Form vor.

Quantität: Aufgrund der relativ geringen Kosten, die mit der Speicherung von Informationen im Internet verbunden sind, ist die Angebotsmenge immens groß und nimmt stetig zu. Der Umfang an sich stellt für die Stichprobenbildung kein Problem dar, geht aber oftmals mit einer erschwerten Erreichbarkeit der Grundgesamtheit einher. Neue (vgl. auch Punkt 1) oder (für einschlägige Suchmaschinen) randständige Angebote haben dadurch eine geringere oder keine Chance, bei der Auswahl von Untersuchungselementen berücksichtigt zu werden.

Alle genannten Punkte machen überdies eine *Archivierung* des Materials schwierig. Während bereits die Fülle und Dynamik der Inhalte kaum eine angemessene Offline-Speicherung möglich machen, erfordern zudem die Hypertextualität, Multimedialität bzw. -modalität einen hohen technischen Aufwand (z.B. große recherchierbare Datenarchive, anspruchsvolle Website-Spiegelsoftware, komplexe Dateiverwaltung von Bild- und Videodownloads etc.).⁵

Auswahlpraxis von Online-Inhaltsanalysen

Angesichts dieser medialen Bedingungen im Internet bietet sich eine Typologie inhaltsanalytischer Verfahren an, die zwar nicht ausdrücklich für den Online-Bereich entwickelt wurde, sich aber hierfür als günstig erwiesen hat (vgl. zum Folgenden RÖSSLER/WIRTH 2001: 284ff.; zur Synopse inhaltsanalytischer Verfahren MERTEN 1995: 119ff.). Die Typologie trennt zunächst angebots- und nutzerzentrierte Inhaltsanalysen. Dabei treten beim ersten Typus die online publizierten Medieninhalte als Angebotsoptionen in den Analysefokus, die potenziell jeder Nutzer rezipieren könnte. Der zweite Typus bezieht sich auf die vom individuellen Nutzer tatsächlich rezipierten Angebote (Angebotsrealisierungen). Beide Typen unterteilen Rössler und Wirth des Weiteren in zwei idealtypische Unterformen. So lässt sich die angebotszentrierte Inhaltsanalyse zum einen in die Sparten- oder Bereichsanalyse aufgliedern, die an Mediengenres,

5 Neben diesen Schwierigkeiten bietet das Internet aber auch eine Reihe von Vorteilen, z.B. in Bezug auf den Zugriff auf verstreut publizierte Materialien oder den erleichterten Umgang mit den bereits digitalisierten Materialien. Vgl. dazu Mann/Stewart (2000: 17ff.).

Publikationsformaten oder Anbietergruppen orientiert ist, und zum anderen in eine eher inhaltlich ausgerichtete Fokusanalyse, die »ganz konkrete Themen, Personen, Ereignisse oder auch Autoren fokussiert« (RÖSSLER/WIRTH 2001: 285)⁶. Die nutzerzentrierte Inhaltsanalyse unterteilen die Autoren erstens in die Publizitätsanalyse, die zur Eingrenzung der zu analysierenden Angebote Reichweitenkriterien anlegt (z. B. die 10 meistbesuchten Websites), welche beispielsweise anhand von Unique User, Page-Impressions oder Visits ermittelt werden. Zweitens nennen die Autoren die Selektivitätsanalyse, welche nicht die kollektive Webnutzung sondern die einzelnen Rezeptionen in den Blick nimmt, indem die konkrete Surfpraxis individueller Nutzerinnen und Nutzer anhand von servergenerierten Logfiles und/oder Videomitschnitten analysiert wird (vgl. zum Vorgehen WERNER 1998; SEIBOLD 2002).

Während die Probleme im Umgang mit Online-Material in nutzerzentrierten Inhaltsanalysen durch den dargestellten Einsatz von technischen Aufzeichnungsverfahren lösbar erscheinen, machen sie in der angebotszentrierten Inhaltsanalyse weiterhin besondere Verfahren nötig. Für das Ziehen von Stichproben werden verschiedene Verfahren diskutiert bzw. empfohlen. Generell besteht das Problem, dass eine für die Untersuchungsfrage relevante Grundgesamtheit der Angebote kaum oder nur unter ganz bestimmten Bedingungen (z. B. durch sehr enge Auswahlzeiträume) zu generieren ist. Rössler und Wirth (2001) kommen zu dem Schluss, dass aufgrund der Nicht-Verfügbarkeit von kompletten Listen bzw. Sammlungen relevanter Webangebote Zufallsstichproben nicht gezogen werden können. Sollten entsprechende Listen dennoch off- oder online vorhanden sein, so drohen diese aufgrund der Dynamik des Internets schnell an Aktualität zu verlieren (ebenso: MCMILLAN 2000: 92). Solche Listen werden online beispielsweise aus redaktionellen (z. B. web.de), mit Suchmaschinen generierten (z. B. google.de) oder kollaborativen (z. B. Mister Wong) Web-Verzeichnissen bezogen. Falls diese trotz der Nachteile verwendet werden, empfehlen die Autoren zumindest anzugeben, wie aktuell, vollständig und glaubwürdig sie jeweils sind.

Werden Domain-Listen zur Stichprobenziehung verwendet, muss außerdem beachtet werden, dass die damit verbundene grobe Auswahl

6 Vgl. für Erstere z. B. die Studien Bruns/Marcinkowski/Schierl (1997), Dahm/Rössler (1997), Rössler/Eichhorn (1999), Bruns/Kruggel/Marcinkowski (1999), Schultz (1999), Perry/Bodkin (2000) und für Zweitere Mehta/Plaza (1997), Heddergott/Loosen (2000).

nur für Fragestellungen geeignet ist, die nicht auf tiefere Ebenen des Internets abzielen (vgl. WEARE/LIN 2000: 277). Bei thematisch stark eingegrenzten Fokusanalysen, bei denen nicht bestimmte Angebotsbereiche untersucht werden, sondern konkrete Themen, Ereignisse oder Personen, kann nach Ansicht von Rössler und Wirth (2001) auch auf die genannten Linkkataloge oder Suchmaschinen zurückgegriffen werden. Dank der thematischen Spezifität ist hier eher mit einer (fast) vollständigen Auflistung der Grundgesamtheit zu rechnen.

Mit der Nutzung von Suchmaschinen sind allerdings zahlreiche Schwierigkeiten verbunden, die sich sowohl bei der Identifizierung einer möglichen Grundgesamtheit als auch bei Stichprobenziehungen negativ auswirken. So ist für den Nutzer meist nicht ersichtlich, wie die Suchergebnisse zustande kommen (z. B. WEARE/LIN 2000; WELKER/WERNER/SCHOLZ 2005: 51f.) Mögliche Homonymien und Synonymien zu den Suchbegriffen können weitere Einschränkungen mit sich bringen. In der Regel muss zudem davon ausgegangen werden, dass auf der Auswahl von Suchmaschinen beruhende Grundgesamtheiten zugunsten stärker benutzter Bereiche des Webs verzerrt sind. Auch die Tatsache, dass Suchmaschinen nicht alle existierenden Webseiten (sofort) finden und somit indexieren bzw. die Aktualisierung von indizierten Seiten nicht immer täglich stattfindet, ist bei Fragestellungen, die neue oder kurzlebige Websites zum Inhalt haben, problematisch (vgl. GALLIKER/MÄNNEL 1999: 256). Auch veraltete, vorübergehend nicht erreichbare oder sogar tote Links, mehrere Treffer, die auf die gleiche Seite verweisen, bzw. teilweise identische Inhalte unter verschiedenen URLs schmälern die Güte der Definition der Grundgesamtheit und damit die Validität. Derartige Fehlleitungen und Dopplungen ließen sich nur durch eine manuelle Bereinigung der Auswahl vermeiden. Kaum eine Alternative ist die stichprobenartige Kontrolle, um die Größe des Fehlers einschätzen zu können, weil die Fehlergrößen nicht klar sind. Problematisch ist beim Einsatz von Suchmaschinen allerdings, dass deren Treffer nichts über die inhaltliche Relevanz einer Seite aussagen. Wenn nur nach dem Vorkommen eines Wortes gesucht wird, kann das entsprechende Wort auf der Seite auch eine untergeordnete Rolle spielen – und der Hauptinhalt der Seite eigentlich ein anderer sein. Kritisiert wird darüber hinaus, dass Suchmaschinen verfälschte Ergebnisse liefern, weil sie als Weblogs getarnte Anbieterseiten auflisten (vgl. SCHULTZE/POSTLER 2008: 377).

Im Lichte einer nutzerzentrierten Analyse und damit aus einer Wirkungsperspektive lässt sich derartigen Argumenten allerdings entgegenhalten, dass auch Nutzer beim Surfen häufig auf Ergebnisseiten von Suchmaschinen zurückgreifen (vgl. GVU 1999). Entsprechend spiegeln die Verwendung von Suchmaschinen und die daraus resultierende (unbekannte) Grundgesamtheit letztlich das normale Nutzerverhalten wider (vgl. WOLLING/KUHLMANN 2003: 150). So ist es in dieser Perspektive fraglich, ob ein Angebot, welches nicht zugänglich ist, überhaupt berücksichtigt werden sollte. Um einigermaßen erschöpfende Suchergebnisse zu erhalten, wird vorgeschlagen, verschiedene Suchmaschinen, Meta-Suchmaschinen, Linklisten und die am häufigsten genutzten Seiten des zu untersuchenden Themenfeldes miteinander zu kombinieren (vgl. BAR-ILAN 2001: 18; WEARE/YIN 2000). Ob dieses Vorgehen allerdings weiterhilft, ist mehr als fraglich, weil sich damit Fehlergrößen möglicherweise potenzieren können.

Insgesamt kann festgehalten werden: Die Vorstellung, dass beispielsweise mit ›Google‹ aus dem Universum der Untersuchungselemente (bestimmte symbolische Repräsentationen von Personen, Websites o.Ä.) eine wie immer geartete Zufallsauswahl getroffen werden kann, ist leider irrig. ›Google‹ arbeitet nach bestimmten Algorithmen, die dem Forscher nicht nur unbekannt sind, sondern die zudem gerade nicht auf Zufallsziehungen beruhen. Ferner ist davon auszugehen, dass ein Suchdurchlauf nicht alle Elemente einer Grundgesamtheit erfasst. Auch die Methode, sich zunächst eine Liste von ›Pseudozufallsverfahren‹ ist ungeeignet. Das liegt auch an der Tatsache, dass bei Suchmaschinen zahlreiche interne und externe Manipulationsmöglichkeiten bekannt sind, die das Suchergebnis einer Anfrage beeinflussen können (WELKER/WERNER/SCHOLZ 2005: 52).

Neben dem Einsatz von Suchmaschinen wird auch der des ›direct random sampling‹ diskutiert, das analog zum ›random digit dialing‹ automatisch URLs generiert. Anders als bei ›random digit dialing‹ kann jedoch anhand der URL die Herkunft der entsprechenden Internetseite nicht geografisch eingeschränkt werden. Zudem ist dieses Vorgehen problematisch, wenn spezifische Fragestellungen im Fokus der Analyse stehen. Denn bei zufällig ausgewählten URLs müssen diese nicht zwangsläufig Informationen zum analysierten Thema enthalten. Deshalb empfiehlt sich dieser Ansatz allenfalls für breite Populationen und sehr breite Fragestellungen. Vor diesem Hintergrund ist es nicht verwunderlich, dass die häufig im Rahmen von Online-Inhaltsanalysen eingesetzten bewussten Auswahlverfahren, etwa die Analyse typischer Fälle, als wenig proble-

matisch angesehen werden (vgl. RÖSSLER/WIRTH 2001: 290). Die Aussagekraft derartiger Analysen ist jedoch stark eingeschränkt und ermöglicht allenfalls eine (wenn auch mitunter notwendige) Momentaufnahme.

Bei der Bestimmung möglicher Auswahl- und Analyseeinheiten, nach der sich die Erhebung und Codierung des Materials richten sollen, zeigen sich medienbedingt weitere Probleme. Im Fokus der meisten bisher veröffentlichten Online-Inhaltsanalysen stehen als Analyseeinheiten entweder nur die Startseite (Toppage) oder der Gesamt-Auftritt bzw. eine Auswahl an Seiten aus verschiedenen Ebenen der Verlinkungsstruktur. Weiterhin bestimmen Rössler und Wirth (2001: 291ff.) als Codiereinheiten für einen Webauftritt insgesamt eine Reihe von Strukturparametern (Gesamtumfang der Seiten, Linkstruktur bzw. -ebenen, Nutzerführung, Netzwerkdicke bzw. -zentralität etc.) und für die Einzelseiten die Maße Bildschirmfüllung, Linkgestaltung und -funktionalität sowie die Position im Gesamtauftritt. Weitere Codiereinheiten können Merkmale des Screendesigns behandeln wie die Verteilung von Bild-Sprache-Ton-Anteilen, deren jeweilige Funktionen und Konstitution (z. B. Animation, Video, Podcasts), mögliche Frameaufteilungen und die gestalterische »Anmutung des Angebots« (ebd.: 292; Herv. im Original). Auch eingesetzte Interaktivität wird hier als Codiereinheit bestimmt, die in Untereinheiten wie mögliche Gästebücher, Downloads, Chat-Rooms, Newsgroups, Bestellformulare, Wissenstests, Befragungen etc. aufgeschlüsselt sein können. Als weitere Codiereinheit geben die Autoren die *angebotsspezifischen Inhalte* an. Die Auswahl richtet sich hier nach inhaltlichen Kriterien, die sich aus der Forschungsfragestellung ergeben. Entsprechend dieser können hierbei bestimmte Themen, Akteure, Meinungen oder Quellen ermittelt werden.

Fazit und Anwendungsbeispiele

Die Notwendigkeit, mit Stichproben zu arbeiten, und deren methodologische Grundlagen und Gütekriterien unterscheiden sich bei der Durchführung von Online-Inhaltsanalysen nicht grundsätzlich von anderen Untersuchungsgegenständen der empirischen Sozialforschung. Bei der Stichprobenziehung kann der Forscher daher auch bei Online-Inhalten zunächst auf das allgemeine, methodische und handwerkliche Inventar von Auswahlverfahren zurückgreifen. Auf der anderen Seite weist der

Untersuchungsgegenstand, wie dargelegt, zahlreiche online-spezifische Merkmale auf, welche die Auswahlverfahren vor Herausforderungen stellt. Hierzu zählen insbesondere das Problem der unbekanntenen oder zumindest nicht in Gänze erreichbaren Grundgesamtheit und das Problem der Abgrenzung der Auswahleinheiten. Zur Lösung der damit verbundenen Probleme liegt kein Patentrezept vor. Einige bereits häufig eingesetzte Lösungsansätze haben wir vorgestellt und kritisch diskutiert, da sie alle auch spezifische Probleme aufweisen. Das adäquate Vorgehen bei der Stichprobenziehung von Online-Inhalten sehen wir daher nicht in der Anwendung eines spezifischen Prozederes, sondern in einem *offengelegten* und für die beschriebenen Probleme *sensibilisierten* Forschungsprozess.

Um dennoch zum einen unsere bisherigen Ausführungen zu illustrieren und zum anderen dem anwendungsorientierten Leser einen schnelleren Zugriff zu ermöglichen, möchten wir zum Schluss fünf beispielhafte Fragestellungen, ihre Anforderungen an ein Auswahlverfahren und Möglichkeiten zu deren Umsetzung vorstellen. Die jeweils vorgestellten Lösungen sind nicht in jedem Fall einfach auf andere Fragestellungen übertragbar, demonstrieren aber dennoch das prinzipielle Vorgehen und deuten das Spektrum an Möglichkeiten, aber auch Problemen an.

Beispiel 1 [echte Zufallsauswahl]: Wie wurde die Bundeskanzlerin Angela Merkel 2007 im Forum auf *Spiegel-Online* beurteilt? Die Grundgesamtheit lässt sich demnach eindeutig definieren als alle Beiträge im Forum auf *Spiegel-Online* im Jahre 2007 inkl. aller Unterforen. Auch die Auswahleinheiten sind als einzelne Beiträge klar voneinander abgrenzbar. Als Forscher ist man weiterhin in der vorteilhaften Situation, auch rückwirkend auf alle Beiträge zugreifen zu können. Die Auswahl kann so unter Rückgriff auf die vollständige Grundgesamtheit erfolgen. In der Regel werden Gründe der Forschungsökonomie gegen eine (prinzipiell mögliche) Vollerhebung sprechen. Im vorliegenden Beispiel bietet sich eine echte, einfache Zufallsauswahl aus allen Beiträgen an. Dies kann z. B. durch die Auswahl jedes n-ten Beitrages in einer chronologisch oder zufällig⁷ sortierten Liste aller Beiträge erfolgen.

Beispiel 2 [mehrstufige, zufällige Auswahl]: Wie berichteten *FAZ*, *Net*, *Spiegel.de* und *Handelsblatt.de* 2008 über die Finanz- und Wirt-

7 Die Anzahl der Backlinks steht in einem proportionalen Verhältnis zur Wahrscheinlichkeit, dass ein Blog von Nutzern gefunden wird.

schaftskrise? Auch in diesem Beispiel ist die Grundgesamtheit prinzipiell bekannt und erreichbar: alle Beiträge über die Finanz- und Wirtschaftskrise der drei Online-Medien im Jahr 2008. Als schwierig könnte sich die thematische Einschränkung auf die Berichterstattung über die Finanz- und Wirtschaftskrise erweisen. Diese Schwierigkeiten der Definition des Zugriffskriteriums sind zwar nicht online-spezifisch, allerdings zeigen sich hier online-spezifische Lösungsmöglichkeiten. So kann mithilfe einer Liste einschlägiger Begriffe und der Suchfunktion der Archive dieser Online-Angebote ein Zugriff auf relevante Beiträge erfolgen. Dabei geht man aber davon aus, dass alle zur Grundgesamtheit gehörenden Beiträge einen dieser Begriffe enthalten und Beiträge, welche einen solchen Begriff enthalten, auch immer zur Grundgesamtheit gehören. Voraussetzung für die Validität dieses Vorgehens ist somit eine vollständige und eindeutige Liste von Schlüsselbegriffen (zur grundlegenden Problematik solchen Vorgehens vgl. Scharnow in diesem Band). Weiterhin setzt so ein Vorgehen voraus, dass alle Beiträge archiviert und dem Forscher zugänglich sind. Sollte dies nicht der Fall sein, ist eine retrospektive Untersuchung kaum möglich, und es muss prospektiv eine Archivierung des Materials erfolgen. Da sich auch in diesem Beispiel eine Vollerhebung als kaum möglich oder sinnvoll erweist, muss eine Auswahl getroffen werden. Obwohl Online-Medien nur bedingt dem Publikationsrhythmus der klassischen Medien, insbesondere der Printmedien, unterliegen, weisen die Elemente der Grundgesamtheit starke saisonale Schwankungen auf. Daher bietet sich hier analog zum Vorgehen bei Offline-Medien die Verwendung einer künstlichen Woche an (vgl. FRÜH 2007: 117ff.). Dabei wird über einen bestimmten Zeitraum (bspw. eine zufällig ausgewählte Woche) nur ein Wochentag im rotierenden Verfahren ausgewählt. Von jedem dieser in die künstliche(n) Woche(n) aufgenommenen Tage werden dann alle Meldungen berücksichtigt.

Beispiel 3 [mehrstufige und nicht zufällige Auswahl]: Wie stellen sich die Nutzer des deutschen studiVZ in ihren Profilen dar? Auch in diesem dritten Beispiel ist die Grundgesamtheit klar definiert, allerdings nicht komplett erreichbar. Die einzige Möglichkeit, eine Liste mit allen Profilen (die Grundgesamtheit) zu erzeugen, ist die Nutzung der Suchfunktion von studivz. Da jedes Nutzer-Profil die Informationen Bundesland, Stadt, Hochschule und Geschlecht des Nutzers enthält (weitere Angaben sind freiwillig), kann die Suche auf Kombinationen dieser Merkmale ein-

geschränkt werden. Dennoch reicht die Trefferliste – die nur bis zu 300 Profile auflistet – nicht aus, um alle Profile aufzuführen. So sind bspw. mehr als 300 männliche Studierende der FU Berlin in studivz registriert. Somit bleibt dem Forscher nur die Möglichkeit, ein nicht-zufälliges Verfahren einzusetzen oder die zufällige Auswahl dem System von studivz zu überlassen. Mit letzterem ist gemeint, dass unter Nutzung der Suchfunktion quasi-zufällige Ziehungen von Profilen erfolgen. Beispielsweise kann eine Suchanfrage nach Profilen einer Stadt oder Hochschule durch einen zufällig erzeugten Suchstring für den Profilnamen ergänzt werden. Dann werden nur noch Profilnamen aufgeführt, welche diesen String enthalten. Aus dieser Liste kann dann z. B. jedes n-te Profil ausgewählt werden. Fraglich bleibt bei diesem Vorgehen die Güte der Zufallsauswahl, da weder der genaue Suchalgorithmus von studivz noch die Verteilung der verwendeten Suchstrings in der Grundgesamtheit bekannt sind. Ein naheliegendes nicht-zufälliges Verfahren stellt die Verwendung der weiter oben beschriebenen Trefferlisten dar. Dabei können aus den (aufgrund der Begrenzung auf 300 Treffer) nicht vollständigen Listen per Zufall Profile (bspw. jedes n-te) ausgewählt werden. Da aber die nicht gelisteten Profile keine Chance zur Berücksichtigung in der Stichprobe aufweisen, handelt es sich dennoch um keine Zufallsstichprobe. Um den damit einhergehenden Fehler abschätzen und offenlegen zu können, müsste der Algorithmus bekannt sein, nach dem studivz seine Suchtreffer sortiert resp. welche Treffer ausgeblendet werden. Beide Verfahren sind mit einer weiteren Unsicherheit behaftet: Die Grundlage der Auswahl sind u. a. Angaben, welche von Nutzern gemacht wurden (Geschlecht, Hochschule, Bundesland) und nicht der Realität entsprechen müssen.

Beispiel 4 [systematische Auswahl; Konzentrationsprinzip]: Welche Themen werden in der deutschen Blogosphäre behandelt? Die Grundgesamtheit dieser Fragestellung – die ›deutsche Blogosphäre‹ – ist sowohl definitorisch als auch physisch schwer zu fassen. Im Rahmen der Definition ist zunächst die Entscheidung zu treffen, was alles als ›Blog‹ betrachtet werden kann. Zählen alle Internetangebote dazu, welche sich als solche bezeichnen, welche von der Blog-Suchmaschine Technorati gelistet werden, welche auf einer sogenannten ›Blog-Software‹ basieren oder nur solche, welche durch Autorenschaft, Stil und Periodik Merkmale eines Blogs aufweisen? Die zweite Frage innerhalb der definitorischen Bestimmung der Grundgesamtheit betrifft das Attri-

but ›deutsch‹. Hier ist zu entscheiden, ob die Sprache des Blogs oder das Hosting auf der deutschen Top-Level-Domain ›.de‹ Definitionskriterium ist. Diese Fragen sind aber ausschließlich auf der Basis der Fragestellung und vor deren theoretischem Hintergrund zu klären. Die methodischen Entscheidungen betreffen hingegen die Erreichbarkeit der Grundgesamtheit. Das Problem besteht darin, dass keine vollständige Liste deutscher Blogs existiert und auch aufgrund der komplexeren Definition von ›deutsche Blogs‹ – wie auch immer sie genau ausfällt – nur schwer mithilfe von Suchmaschinen recherchierbar ist. Hinzu kommt, dass der Rückgriff auf Suchmaschinen – wie bereits mehrfach beschrieben – stets (potenziell) unvollständig bleibt. Daher ist weder physisch noch symbolisch eine Liste aller Elemente der Grundgesamtheit vorhanden, noch kann eine solche erstellt werden. Wir schlagen daher in diesem Fall eine systematische Stichprobe nach dem Konzentrationsprinzip vor. Dazu können wir auf ein Ranking der deutschsprachigen Blogs nach den Relevanzkriterien der Blog-Suchmaschine Technorati, wie sie bspw. unter www.deutscheblogcharts.de gelistet wird, zurückgreifen. Hier werden die einhundert deutschen Blogs mit den meisten Backlinks aufgeführt. Die Backlinks sind Verweise (Links) von anderen Seiten im Internet auf ein Blog und dienen als Indikator für dessen Linkpopularität. Grundlage der Stichprobe können nun diese 100 Blogs oder eine systematische (z. B. Top Ten) oder zufällige Auswahl daraus sein. Es muss aber betont werden, dass diese Stichprobe eben nicht repräsentativ ist, da das Auswahlprinzip ›Anzahl der Backlinks‹ nicht berechenbaren Verzerrungen unterliegen kann. Hier ist bspw. an (im Extremfall) Manipulationen durch die Blog-Betreiber – welche an einem hohen Ranking interessiert sind – zu denken. Ebenfalls sind neue Blogs bei diesem Auswahlverfahren systematisch benachteiligt – unabhängig davon, wie repräsentativ oder typisch sie für die Blogosphäre sind. Nichtsdestotrotz ist diese Form der Auswahl bei gegebener nutzerzentrierter Fragestellung und den üblichen forschungsökonomischen Beschränkungen die bestmögliche, effizienteste Lösung. Dagegen ist aber zu bedenken, dass bei anderen, insbesondere angebotszentrierten Fragestellungen gerade auch die kleinen Blogs von Interesse sind. Diese würden beim beschriebenen Verfahren überhaupt nicht berücksichtigt, sodass hier andere Lösungen gesucht werden müssen. Dies ist im nächsten Beispiel ähnlich.

Beispiel 5 [Auswahl bei unbekannter Grundgesamtheit]: Wie wird im Internet über die Universität Leipzig berichtet und geschrieben? Die

Grundgesamtheit für diese Fragestellung besteht aus allen online veröffentlichten Aussagen über die Universität Leipzig. Die für die Stichprobenziehung entscheidende Schwierigkeit in diesem Beispiel ist die unbekannte, nicht erreichbare Grundgesamtheit. Eine Möglichkeit, dieses Problem zu umgehen (aber nicht zu lösen), wäre, auch hier eine systematische Auswahl vorzunehmen und bspw. nur drei ausgewählte Studierenden-Portale für die Untersuchung auszuwählen. Eine andere Strategie bestünde im Verfahren des theoretischen Samplings (vgl. dazu Meier/Pentzold in diesem Band). Wir schlagen in diesem Fall aber folgende Strategie vor: Das Hauptaugenmerk des Auswahlprozesses liegt auf der Beschaffung möglichst umfassender Informationen über die Grundgesamtheit. Dazu sind verschiedene Such- und Metasuchmaschinen und verschiedene Suchanfragen zu kombinieren (zu diesem Verfahren vgl. bspw. SCHWEIGER/MARKMILLER [im Druck]). Die so gewonnenen Trefferlisten müssen anschließend konsolidiert werden. Zum einen müssen doppelte bzw. mehrfache Nennungen derselben Online-Quelle eliminiert, zum anderen müssen Seiten identifiziert und aus der Liste entfernt werden, welche fälschlicherweise als Suchtreffer aufgeführt sind. Insbesondere die Verwendung mehrerer Suchanfragen in Form unterschiedlicher Kombinationen verschiedener Suchbegriffe (bspw. Uni, Universität, Hochschule, university etc.) wird eine hohe Rate an nicht zur Grundgesamtheit gehörenden Treffern erzeugen. Zum Teil sind solche Fehltreffer jedoch nur in einem aufwendigen, nicht-maschinellen Verfahren zu identifizieren.

Diese so gewonnene Liste an Adressen mit Aussagen zur Universität Leipzig stellt für das weitere Auswahlverfahren die bestmögliche *Annäherung* an eine vollständige Liste der Grundgesamtheit dar. Für das weitere Vorgehen stehen demnach wieder prinzipiell alle Auswahlmöglichkeiten zur Verfügung. Dazu gehört auch eine Zufallsauswahl von bspw. jedem zehnten Eintrag der Adressliste, die wir in unserem Beispiel vorschlagen. Wichtig ist anzumerken, dass es sich hierbei nicht um eine vollständige Liste aller Elemente der Grundgesamtheit handelt und je nach Gründlichkeit der Konsolidierung auch Elemente doppelt auftauchen. Daher kann man in solchen Fällen nicht von einer echten Zufallsauswahl sprechen, die als Voraussetzung für die Repräsentativität der Auswahl gewährleistet sein müsste. Bestenfalls kann man die vom Forscher kaum sicher verifizierbare, gleiche Datenqualität erreichen.

Literatur

- ADM ARBEITSKREIS DEUTSCHER MARKT- UND SOZIALFORSCHUNGSINSTITUTE E.V.; AG.MA ARBEITSGEMEINSCHAFT MEDIA-ANALYSE E.V. (Hrsg.): *Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis*. Opladen [Leske + Budrich] 1999
- BAR-ILAN, J.: Data collection methods on the web for informetric purposes – a review and analysis. In: *Scientometrics*, 50(1), 2001, S. 7-32
- BERELSON, B.: *Content Analysis in Communication Research*. New York [Hafner] 1952
- BRUNS, T.; M. KRUGGEL; F. MARCINKOWSKI: *Die Kommunikationspolitik deutscher Hochschulen im Internet*. pro online papiere, Nr. 1/1999 [online] <http://pro-online.uni-duisburg.de/papiere/index.html>
- DAHM, H.; P. RÖSSLER: Marktplatz der Sensationen? Deutsche Anbieter von Online-Shopping im Test. In: *Media Spektrum*, 3/1997, S. 32-36
- FRAAS, C.; C. PENTZOLD: Online-Diskurse. Theoretische Prämissen, methodische Anforderungen und analytische Befunde. In: WARNKE, I.; J. SPITZMÜLLER (Hrsg.): *Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene*. Berlin/New York [de Gruyter] 2008, S. 291-326
- FRETWURST, B; V. GEHRAU; R. WEBER: Notwendige Angaben zu Auswahlverfahren. Theoretische Überlegungen und eine empirische Auswertung der Dokumentationspraxis in der KW. In: GEHRAU, V.; B. FRETWURST; B. KRAUSE; G. DASCHMANN (Hrsg.): *Auswahlverfahren in der Kommunikationswissenschaft*. Köln [Herbert von Halem] 2005, S. 33-51
- FRÜH, W.: *Inhaltsanalyse. Theorie und Praxis*. 6Konstanz [UVK] 2007
- GALLIKER, M.; O. MÄNNEL: Suchmaschinen als Datenerhebungsinstrument von WWW-Inhaltsanalysen: Möglichkeiten und Grenzen. In: BATINIC B.; A. WERNER; L. GRÄF; W. BANDILLA (Hrsg.): *Online Research. Methoden, Anwendungen und Ergebnisse*. Göttingen [Hogrefe] 1999, S. 245-261
- GEHRAU, V.; B. FRETWURST: Auswahlverfahren in der Kommunikationswissenschaft. Eine Untersuchung aktueller Veröffentlichungen über empirische Studien in der Kommunikationswissenschaft. In: GEHRAU, V.; B. FRETWURST; B. KRAUSE; G. DASCHMANN (Hrsg.): *Auswahlverfahren in der Kommunikationswissenschaft*. Köln [Herbert von Halem] 2005, S. 13-31
- [GVU] GRAPHICS, VISUALIZATION AND USABILITY (Hrsg.): *GVU's www user surveys*. Graphics, Visualization & Usability Center at Georgia Tech. 1999 http://www.cc.gatech.edu/gvu/user_surveys/ [Zugriff am 24.07.08]

- HEDDERGOTT, K.; W. LOOSEN: Ins Netz gegangen? Eine Inhaltsanalyse im Umfeld der Fußball-WM 1998 im World Wide Web. In: BROSIUS, H.-B. (Hrsg.): *Kommunikation über Grenzen und Kulturen*. Schriftenreihe der DGPK, Bd. 27. Konstanz [UVK] 2000, S. 209-224
- KOPS, M.: *Auswahlverfahren in der Inhaltsanalyse*. Kölner Beiträge zur Sozialforschung und angewandten Soziologie. Bd. 21. Meisenheim am Glan [Hain] 1977
- KOPS, M.: Auswahlverfahren in der Inhaltsanalyse: Die bewusste Auswahl als mögliche Alternative zur Wahrscheinlichkeitsauswahl. In: MOCHMANN, E. (Hrsg.): *Computerstrategien für die Kommunikationsanalyse*. Frankfurt/M. [Campus] 1980, S. 47-67
- MANN, C.; F. STEWART: *Internet Communication and Qualitative Research. A Handbook for Researching Online*. London, Thousand Oaks, New Delhi [Sage] 2000
- MCMILLAN, S.J.: The Microscope and the Moving Target: The Challenge of Applying Content Analysis to the World Wide Web. In: *Journalism & Mass Communication Quarterly*, 77(1), 2000, S. 80-98
- MEHTA, M.D.; D. PLAZA: Content Analysis of Pornographic Images Available on the Internet. In: *The Information Society*, 13(2), 1997, S. 153-161
- MEIER, S.: *(Bild-)Diskurs im Netz. Konzept und Methode für eine semiotische Diskursanalyse im World Wide Web*. Köln [Herbert von Halem] 2008
- MERTEN, K.: *Inhaltsanalyse. Einführung in Theorie, Methode und Praxis*. 2., verb. Aufl. Opladen [Westdeutscher Verlag] 1995
- PERRY, M.; C. BODKIN: Content Analysis of Fortune 100 company Web sites. In: *Corporate Communications*, 5(2), 2000, S. 87-96
- RIFFE, D.; A. FREITAG: A Content Analysis of Content Analyses. Twenty-Five Years of Journalism Quarterly. In: *Journalism and Mass Communication Quarterly*, 74(3), 1997, S. 515-524
- RÖSSLER, P.: Content Analysis in Online Communication: A Challenge for Traditional Methodology. In: REIPS, U.-D.; M. BOSNJAK (Hrsg.): *Dimensions of Internet Science*. Lengerich et al. [Pabst] 2001, S. 291-307
- RÖSSLER, P.: *Inhaltsanalyse*. Konstanz [UVK] 2005
- RÖSSLER, P.: Standardisierte Inhaltsanalysen im WorldWideWeb. Überlegungen zur Anwendung der Methode am Beispiel einer Studie zu Online-Shopping-Angeboten. In: BECK, K.; G. VOWE (Hrsg.): *Computernetze – ein Medium öffentlicher Kommunikation?* Berlin [Spiess] 1997, S. 245-267

- RÖSSLER, P.; W. EICHHORN: WebCanal – ein Instrument zur Beschreibung von Inhalten im World Wide Web. In: BATINIC, B. et al. (Hrsg.): *Online Research. Methoden, Anwendungen und*
- RÖSSLER, P.; W. WIRTH: Inhaltsanalysen im World Wide Web. In: WIRTH, W.; E. LAUF (Hrsg.): *Inhaltsanalysen. Perspektiven, Probleme, Potentiale*. Köln [Herbert von Halem] 2001, S. 280-302
- SCHULTZ, T.: Interactive Options in Online Journalism: A Content Analysis of 100 U.S. Newspapers. In: *Journal of Computer-Mediated Communication*, 5(1), 1999 [online]
- SCHULTZE, M.; A. POSTLER: Online-Trend-Monitoring bei der EnBW: Mit dem Ohr am Kunden. In: A. ZERFASS; M. WELKER; J. SCHMIDT (Hrsg.): *Kommunikation, Partizipation und Wirkungen im Social Web. Band 2: Strategien und Anwendungen: Perspektiven für Wirtschaft, Politik und Publizistik*. Köln [Herbert von Halem] 2008, S. 370-382
- SCHWEIGER, W.; I. MARKMILLER: WebScreen – ein Verfahren zur Messung der öffentlichen Meinung im Internet. In: JANDURA, O.; M. MAURER; J. WOELKE (Hrsg.): *Forschungsmethoden für die Markt- und Organisationskommunikation*. Köln [Herbert von Halem] [im Druck]
- SEIBOLD, B.: Die flüchtigen Web-Informationen eingefangen. Lösungsansätze für die Online-Inhaltsanalyse bei dynamischen Inhalten im Internet. In: *Publizistik*, 47(1), 2002, S. 45-56
- WEARE, C.; L. WAN-YING: Content analysis of the world wide web. Opportunities and challenges. In: *Social Science Computer Review*, 18, 2000, S. 272-292
- WELKER, M.; A. WERNER; J. SCHOLZ: *Online-Research*. Heidelberg [dpunkt Verlag] 2005
- WERNER, A.: Online-Medien. Theoriebereicherung durch neue Forschungsmethoden. In: RÖSSLER, P. (Hrsg.): *Online-Kommunikation. Beiträge zu Nutzung und Wirkung*. Opladen [Westdeutscher Verlag] 1998, S. 227-241
- WOLLING, J.: Normalzeit vs. Spezialzeit. Besondere Ereignisse als Problem der Stichprobenziehung bei Inhaltsanalysen von Medienangeboten. In: GEHRAU, V.; B. FRETWURST; B. KRAUSE; G. DASCHMANN (Hrsg.): *Auswahlverfahren in der Kommunikationswissenschaft*. Köln [Herbert von Halem] 2005, S. 138-157
- WOLLING, J.; C. KUHLMANN: Das Internet als Gegenstand und Instrument der empirischen Kommunikationsforschung. In: LÖFFELHOLZ, M.; T. QUANDT (Hrsg.): *Die neue Kommunikationswissenschaft. Theorien, Themen*

und Berufsfelder im Internet-Zeitalter. Eine Einführung. Wiesbaden [Westdeutscher Verlag] 2003, S. 131-161

ZÜLL, CORNELIA: Identifying Societal Events by Computer-Assisted Text Analysis. In: WELKER, M. et al. (Hrsg.): *11th General Online Research Conference GOR 09. Proceedings.* Wien 2009, S. 148-149