Axel Bruns und Jean Burgess

Blogforschung: Der ‚Computational Turn‘

ForscherInnen aus Sozial- und Geisteswissenschaften interessieren sich seit nunmehr einem Jahrzehnt für Blogs, Online-Tagebücher und Online-Journale. Auch wenn die Zuwachsrates der Blogosphäre seit der Blütezeit des Bloggens in den 2000ern stagniert, bleiben Blogs doch eines der bedeutendsten Genres der internetgestützten Kommunikation. Tatsächlich ist nach der Massenabwanderung zu Facebook, Twitter und anderen erst in jüngerer Zeit entstandenen Kommunikationsmitteln eine etwas kleinere, aber umso stärker etablierte Blogosphäre von engagierten und eingeschworenen Teilnehmenden übriggeblieben. Blogs werden mittlerweile als Teil einer institutionellen, persönlichen und Gruppen-Kommunikationsstrategie akzeptiert. In Stil und Inhalt liegen sie zwischen den statischeren Informationen auf konventionellen Websites und den ständig aktualisierten Facebook- und Twitter-Newsfeeds. Blogs ermöglichen es ihren AutorInnen (und deren KommentatorInnen), bestimmte Themen im Umfang von einigen hundert bis zu einigen tausend Wörtern zu durchdenken, in kürzeren Posts ins Detail zu gehen und ggf. intensiver durchdachte Texte anderswo zu publizieren. Zudem sind sie auch ein sehr flexibles Medium: Bilder, Audio-, Video- sowie andere Materialien können mühelos eingefügt werden – und natürlich auch *das* grundlegende Instrument des Bloggens: Hyperlinks.

Tatsächlich darf die Rolle des Links in Blogs und Blogposts nicht unterschätzt werden. Mit welchem Genre oder Thema auch immer sich individuelle Bloggende beschäftigen, meistens wird das eigene Blog dazu genutzt, zeitnahe Aktualisierungen und Kommentare zur Verfügung zu stellen, die dann typischerweise zu relevanten Einträgen anderer Bloggenden und ebenso zu früheren eigenen Posts verlinken. Beide Spielarten verlinken also zu weiteren Informationen über die aktuellen Themen der Bloggenden, sowie zu von ihnen als interessant oder kritikwürdig empfundenen Nachrichten und Artikeln.

Speziell dort, wo Bloggende Teil einer größeren AutorInnen-Gemeinschaft sind, die ähnliche Interessen oder Ansichten teilt, stellt das wechselseitige Schreiben und Verlinken von Posts eine asynchrone, dezentralisierte Konversation dar, die sich über Tage, Wochen und Monate entfaltet. Solche Gemeinschaften erkennt man oft an den Blogrolls, die sich meist in einer Seitenleiste auf der Blogsite befinden und die Freunde oder Favoriten der BloggerIn auflisten. Links in Blogs verhalten sich dabei durchaus anders als etwa Links in Facebook-Posts oder Tweets: während das schnelle Weiterleiten von meist nur kurz kommentierten Links auf Facebook oder Twitter eine

weitverbreitete Aktivität darstellt, die auch von den von diesen Sozialen Netzwerken bereitgestellten Posting-Funktionen ausgiebig unterstützt wird, so sind Bloggende aus technischen wie auch aus inhaltlichen Gründen meist deutlich zurückhaltender in ihrer Nutzung von Links. Die Präsenz eines Hyperlinks in einem Blogpost, oder auch die Verlinkung zu einem anderen Blog in der Blogroll, stellt daher eine bedeutendere Herausstellung des Linkrezipienten dar, als dies in anderen sozialen Medien der Fall ist. In der Blogosphäre ist der Hyperlink somit eine wichtige Währung (Farrell & Drezner, 2008).

Blogforschung ist daher aus zahlreichen Gründen interessant: Einmal kann eine qualitative Analyse eines oder mehrerer Blogs die kognitiven und kommunikativen Prozesse aufzeigen, durch die einzelne Bloggende ihre Online-Identität definieren, sich zu Blogger-KollegInnen positionieren, bestimmte Sachgebiete, Themen und Geschichten formulieren und sich auf unterschiedliche Gesichtspunkte einlassen. Blogforschung gibt auch Aufschluss darüber, wie sich solche Prozesse in verschiedenen Interessensgemeinschaften unterscheiden. Geprägt werden diese Standpunkte vielleicht durch unterschiedliche gesellschaftliche Stellungen, bestimmte Interessensschwerpunkte, den sozioökonomischen Hintergrund individueller BloggerInnen, oder durch andere externe bzw. interne Faktoren.

Eine solche qualitative Forschung blickt bereits auf eine jahrzehntelange Geschichte zurück (für die wichtigsten Sammlungen, s. Gurak, et al., 2004; Bruns und Jacobs, 2006; s. auch Walker Rettberg, 2008) und ist vor kurzem auch dazu übergegangen, Blog-Praktiken in verschiedenen Kulturen zu vergleichen (Russell und Echchaibi, 2009). In anderen Studien wurden in großangelegten Umfragen die Gepflogenheiten und Motivationen von Bloggeden in bestimmten Ländern aus einer soziologischen Perspektive untersucht (e.g. Schmidt, 2009). In wiederum anderen Szenarien wurden Blogs fächerübergreifend als Werkzeuge für reflexives Lernen und Diskutieren sowohl in Pflichtschulen als auch in Hochschulen eingesetzt (Burgess, 2006).

Der „Computational Turn“

Während der letzten Jahre ist ein anderer, bedeutender Zugang zur Blogforschung entstanden – und zwar als Teil dessen, was David Berry (2010) als den „Computational Turn“ bezeichnete: die zunehmende Verfügbarkeit von Werkzeugen für die halbautomatische Erfassung und Analyse von großen Korpora von (u. a.) Web-Inhalten, und die Entwicklung von Forschungsmethoden sowohl in den Geistes- als auch in den Sozialwissenschaften, die diese neuen Möglichkeiten oft in interdisziplinären Forschungsteams nutzen. In der Blog-

forschung ist die Etablierung solcher Praktiken zweischneidig: Die frühesten dieser Forschungsprojekte (s. z. B. Adamic und Glance, 2005) sind hauptsächlich auf die in Blogseiten eingebundenen Links fokussiert. Sie nutzen diese, um Netzwerkstrukturen und die dadurch offenbarten Verkettungen darzustellen. Das geschieht mittels Web-Crawler-Software – wie dem frei verfügbaren Issue-Crawler (www.issuecrawler.net) –, die vorhandene Links aus einer Sammeliste von Webseiten identifiziert und ihnen folgt, um das Netzwerk der in der Liste angegebenen Websites zu ermitteln. Durch die Analyse der Verkettungsmuster ist es z. B. möglich, die Sites mit den meisten eingehenden Links zu identifizieren (diese sind daher als zentrale Informations- und Meinungsträger zu erkennen), jene mit den meisten ausgehenden, aber relativ wenigen eingehenden Links (also am Rande stehende Sites, die „zuschauen“), und solche, die viele Links erhalten *und* häufig zu anderen verlinken (also Sites, die in der Terminologie der Netzwerkanalyse die größte Zentralität im Netzwerk haben, und als dessen wichtigste Knotenpunkte agieren). Zudem können Netzwerke auch eine mehr oder wenige ausgeprägte Clusterbildung aufweisen, d. h. Sites in bestimmten Gruppen sind untereinander stark verlinkt, aber nicht mit Sites rivalisierender Gruppen. In ihrer Untersuchung von politischen Blogs während des US-Wahlkampfes 2004 haben z. B. Adamic und Glance (2005) herausgefunden, dass sich Bloggende der zwei politischen Lager stark mit Gleichgesinnten vernetzten, aber bedeutend seltener mit ihren politischen Gegnern.

Eine solche Analyse ist weit über das relativ enge Feld des expliziten politischen Bloggens hinaus interessant, weil sie sowohl über die Strukturierung der Blogosphäre hinausgehende Informationen liefert, als auch Auskunft über interne Strukturen kleinerer Interessensgruppen oder Identitäten gibt. In diesem Zusammenhang ist es jedoch wichtig zu betonen, dass konventionelle Web-Crawler in mehreren entscheidenden Punkten in ihren Möglichkeiten begrenzt sind: Erstens neigen sie dazu, nicht zwischen Blogs und anderen Arten von Websites zu unterscheiden, und müssen darauf trainiert werden, Links, die häufig in den analysierten Sites vorhanden, aber für die Studie selbst irrelevant sind, zu ignorieren (z. B. zu Google-Suchfunktionen oder zu Blog-Plattform-Providern wie Wordpress und Blogger). Ein unbearbeiteter, von einem Crawler generierter Link-Bestand kann deshalb eine ganze Reihe von Links enthalten, die abhängig vom Fokus der Netzwerkanalyse manuell entfernt werden müssen.

Zweitens können Crawler üblicherweise nicht zwischen verschiedenen Link-Typen innerhalb eines Blogs unterscheiden. Eine typische Blogseite enthält neben dem Blogeintrag Kopfzeilen (Headers), Fußzeilen (Footers) und Seitenleisten (Sidebars) sowie alle möglichen Kommentare der Leserschaft. Hyperlinks erfüllen in jeder dieser Sektionen sehr verschiedene diskursive Rollen. Links in Seitenleisten, Kopf-, und Fußzeilen können beispielsweise lediglich der Navigation dienen, die der Leserschaft einen einfachen Zugang

zu verschiedene Sektionen der Blogsite ermöglichen. Sie sollten normalerweise gänzlich aus einer Analyse ausgeschlossen werden. Einige Links in den Seitenleisten (z. B. in Blogrolls) wiederum können relevant sein, aber eine gänzlich andere Rolle als Links im Blogeinträgen spielen: Während Links in Blogrolls nur die Zugehörigkeit anzeigen (indem sie über die langfristigen Interessen von Bloggenden an anderen Sites informieren), sind Links in *Blogeinträgen* viel unmittelbarer diskursiv (sie beziehen sich direkt und eindeutig auf das Thema des Blogeintrags selbst und sind möglicherweise nur für diesen relevant). Tatsächlich ist es z. B. möglich, dass Bloggende ihre Verbindung mit anderen Blogs durch die Blogrolls anzeigen, aber nur selten zu aktuellen Einträgen dieser Blogs verlinken. Jede Analyse, die die Links der Blogrolls und diskursive Links gleichwertig behandelt, riskiert eine Reihe falsch positiver Resultate: Sie zeigt eine enge Zusammenhörigkeit zwischen Sites, die jeweils auf den Blogrolls der anderen auftauchen, die aber in ihren Blogeinträgen fast gar nicht miteinander „sprechen“. (Schließlich gehören natürlich alle Links in den Kommentaren eines Blogeintrags noch zu einer weiteren Kategorie – hier sind die Bloggenden selbst nicht einmal die AutorInnen dieser Links, sodass die Links weder die Zugehörigkeit der Bloggenden noch ihr diskursives Interesse an der verlinkten Site zeigen.) Eine fortschrittlichere Crawling- und Datenerfassungs-Technologie, die verlässlich zwischen diesen verschiedenen Linktypen unterscheidet, wäre daher wünschenswert. Sie wird derzeit auch von verschiedenen Projekten entwickelt, ist aber leider nicht generell verfügbar.

Drittens bieten Crawling-Prozesse der Forschung keine Möglichkeit, zeitliche Limits für die Auswahl an Blogeinträgen aus einem Korpus zu setzen, auch wenn dies für zahlreiche Zusammenhänge von entscheidender Bedeutung ist. Wie bereits beschrieben, nehmen Crawler die URL einer bestimmten Webseite, identifizieren die Links auf dieser Webseite, und wiederholen dann den Crawling-Prozess (in mehreren Durchläufen) mit jeder der neuen Seiten, die von diesen Links identifiziert wurden. Auch wenn in einer Liste von URLs Blogeinträge eines bestimmten Zeitabschnitts gesammelt werden (sagen wir, aus der letzten Woche), bedeutet dies nicht, dass in der Folge nur die Verlinkungen in den Blogs des genannten Zeitabschnitts analysiert werden. Denn wenn einer oder mehrere der Einträge auf ältere Posts verlinken, kann der vom Crawler untersuchte Link-Korpus auf ein Netzwerk von Blogeinträgen verweisen, das sich über einen Zeitraum von einigen Monaten oder sogar Jahren erstreckt. Jedoch interessiert sich das Forschungsprojekt selbst vielleicht nur für die Verkettungen zwischen den Blogs während eines laufenden Monats oder einer Woche, um herauszufinden, welche Blogs die wichtigsten Meinungsbildner in jenen Themenbereichen sind, die die Community aktuell bewegen, und wie diese Meinungsführerschaft sich über die Zeit verändert. Künftige Forschungswerkzeuge und Methodologien werden sich mit solchen Einschränkungen beschäftigen

müssen. Sie können auch zu einer Abkehr von crawling-basierten Methoden hin zu Ansätzen führen, die neue, *gerade erstellte* Blogeinträge aus einer langen Liste von bekannten Blogs inklusive ihrer Zeitstempel erfassen. Dies ermöglicht dann die Analyse einer Auswahl jener Einträge aus der großen Datensammlung, die innerhalb einer bestimmten Zeitspanne gemacht wurden (wir beschreiben ein solches System in Kirchhoff et al., 2009, und Bruns et al., 2008b).

Indem andere Mittel als (oder zusätzlich zu) Web-Crawler(n) benutzt werden, kann nicht nur das Muster der Verkettungen zwischen den Blogs untersucht, sondern es können die Bloginhalte selbst, und dies in größerem Umfang, analysiert werden. Wenn Blogeinträge erschlossen und in einer lokalen Datenbank gespeichert werden, ist die Verwendung von Werkzeugen möglich, die automatische Textanalyse anbieten. Solche Werkzeuge sind üblicherweise eine Art von Web-Content-Scraper – eine Software, die den Inhalt aus einer Liste von URLs erfasst und auswertet (normalerweise als reinen Text oder HTML). Auch hier ist es notwendig, sicherzustellen, dass der Scraper zwischen dem diskursiven Textinhalt der Blogeinträge selbst und dem zweckmäßigen oder floskelhaften Text aus Seitenleisten, Kopf- und Fußzeilen sowie den Kommentaren der Leserschaft unterscheidet – denn in fast allen Forschungsszenarios ist nur der Text der Blogeinträge selbst relevant für die Analyse. Außerdem müssen Forschende bei der Nutzung von Scrapern natürlich die rechtlichen und ethischen Implikationen abschätzen (womit wir uns weiter unten im Detail befassen werden): Beispielsweise ist zu überlegen, ob die Erfassung solcher Inhalte aufgrund der einschlägigen „fair use“- oder „fair dealing“-Vorschriften im gültigen Urheberrecht zulässig ist und ob (auch wenn das erfasste Material online öffentlich zugänglich ist) Blogposts anonymisiert oder die Identitäten der AutorInnen verschleiert werden sollten.

Wenn für diese Fragen Lösungen gefunden werden können, kann die Erzeugung eines solchen Textbestandes – zusätzlich zu den Informationen über Verkettungsmuster, die von Scrapern oder ähnlichen Werkzeugen erzeugt werden können – weitere Forschungsmöglichkeiten eröffnen. Halbautomatisierte Textanalyse-Werkzeuge können Ranglisten der am meisten benutzten Schlüsselwörter innerhalb eines bestimmten Blogs, aus einer Gruppe von Blogs, oder während eines bestimmten Zeitraumes erstellen. Diese liefern Informationen über Kernthemen der Bloggenden (haben z. B. Bloggende, die zu verschiedenen Gruppen in einem Link-Netzwerk gehören, verschiedene Themenvorlieben?) sowie über einen Wechsel der allgemeinen Interessen im Laufe der Zeit (z. B. als Antwort auf externe Stimuli wie Nachrichten oder aktuelle Ereignisse). Textanalyse-Software wie Leximancer bietet auch die Möglichkeit, Muster wie Kookkurrenz zwischen individuellen Schlüsselwörtern festzustellen: Auf diese Weise lassen sich Gruppen von Schlüsselbegriffen identifizieren, die gemeinhin nebeneinander verwendet werden. Diese Begriffsgruppen zeichnen somit ein

besseres Bild von Themenbesetzungen innerhalb eines einzelnen Blogs (oder einer Gruppe von Blogs, oder während einer bestimmten Zeitperiode), als es eine bloße Rangliste der am häufigsten verwendeten Schlüsselwörter tun kann (cf. Bruns, et al., 2008a). Sie vermitteln sogar einen ersten Einblick davon, wie spezielle Themen eingebettet sind: Kookkurrenz von Namen bestimmter Personen oder Organisationen mit emotional besetzten Themen („schlau“, „unehrlich“, „sexy“, „stark“, ...) können auf die generelle Wahrnehmung der untersuchten Bloggenden hinweisen.

Kombination von quantitativen und qualitativen Ansätzen

Solche Beispiele erklären auch, warum ein rein rechnerischer, quantitativer Ansatz bei der Blogforschung selten ausreichend ist. Netzwerkvisualisierungen von Links und inhaltliche Analyse der Schlüsselwörter können auch dort, wo hochentwickelte Werkzeuge für die Erfassung großer Datenmengen benutzt werden, im Großen und Ganzen nur Annäherungswerte an die Aktivitäten der Bloggenden und Blog-UserInnen liefern und sollten nie als gänzlich schlüssige Analysen betrachtet werden. Obwohl Clustering-Tendenzen bei Link-Netzwerken und Schlüsselbegriffen ein scheinbar überzeugendes Bild von Beziehungen und Verbindungen zeichnen können, ist es, erstens, doch notwendig, mindestens durch stichprobenartige qualitative Analyse die engeren Zusammenhänge zwischen Knoten im Netzwerk zu überprüfen und diese zu interpretieren. In Link-Netzwerken kann beispielsweise häufiges und wechselseitiges Verlinken zwischen zwei Blogs das Zeichen einer engen Beziehung sein – zwischen zwei Bloggenden, die eine herzliche und inhaltsreiche Konversation pflegen –, aber es könnte auch auf ein erbittertes, langgezogenes Wortgefecht hindeuten. In ähnlicher Weise können das gleichzeitige Aufscheinen des Namens eines Politikers und von Schlüsselwörtern wie z. B. „schlau“ und „klug“ im Textmaterial ein Zeichen für einen hohen Zustimmungsgrad sein – aber andererseits könnten solche Ausdrücke auch ironisch benutzt worden sein und so genau das Gegenteil bedeuten. Ein genaues Bild wird man nur erhalten, wenn Forschende wenigstens einen Teil des erfassten Materials zur Verifikation ihrer Interpretationen *in situ* untersuchen.

Zweitens, und das ist von größerer Bedeutung, birgt die Benutzung von automatisierten Werkzeugen wenigstens für einen Teil der Analyse – auch wenn es grundsätzlich notwendig ist, die potentiell sehr großen Datensätze, die der computerunterstützte Ansatz erzeugen kann, automatisch zu verarbeiten – die Gefahr, dass Forschende diese Werkzeuge wie eine Black Box behandeln,

deren innere Funktionalität nicht verstanden werden muss, solange sie „richtig aussehende“ Resultate produziert. Tatsächlich könnte der „Computational Turn“ in einem erheblichen Ausmaß auch als Entwicklung hin zu mehr Daten-Visualisierung beschrieben werden – aber jeder auch nur beiläufige Blick auf die Visualisierungsforschung zeigt eine Vielzahl von Ansätzen, die Rohdaten in Graphen verwandeln können. Entscheidungen, die während der Datenvisualisierung getroffen wurden, sind weit davon entfernt, nur ein neutraler Schritt im Analyseprozess zu sein, und können stattdessen einen entscheidenden Einfluss auf die Interpretation der Forschungsergebnisse haben. Sie können Forschende möglicherweise dazu verführen, Schlüsse aus ihren Studienobjekten zu ziehen, die einem genaueren Blick auf die Ursprungsdaten nicht standhalten. Wenn beispielsweise graphische Abbildungen von Link-Netzwerken oder der Kookkurrenz von Schlüsselwörtern erzeugt werden, müssen WissenschaftlerInnen fragen, aufgrund welcher mathematischer oder netzwerktheoretischer Basis einzelne Knoten im Netzwerk in bestimmten Abständen voneinander oder in speziellen Gruppierungen zueinander gesetzt werden. Dies ist insbesondere auch ein Aufruf, interdisziplinäre Forschungsteams zu bilden, in denen Fertigkeiten in Kultur- und Kommunikationswissenschaften, Soziologie und Netzwerk-Mapping vereint werden.

Schließlich muss betont werden, dass die über die beschriebenen Methoden gewonnenen Daten – unabhängig davon, wie ausgeklügelt die Forschungswerkzeuge sind – immer nur annäherungsweise die tatsächlichen Aktivitäten von Bloggenden bzw. deren Leserschaft abbilden. Während Blogposts eindeutig als mehr oder weniger öffentliche Meinungsäußerungen der Bloggenden beabsichtigt sind, können die dahinter stehenden Individuen nicht nur als durch die jeweils verkündeten Meinungen konstituiert angesehen werden. Ihre Blog-Inhalte repräsentieren nur die – oder, genauer, *eine* – öffentliche Facette der AutorIn, die privat oder in einem anderen Kontext ganz anders auftritt. In ähnlicher Weise sind wir beim Analysieren eines Link-Netzwerks zwischen Blogs daran interessiert, ein Gespür für den ungefähren Verkehr an Lesern zu bekommen, der uns im Gegenzug eine Vorstellung über den Einfluss von bestimmten Bloggenden und deren Ansichten auf die weitere Community (von LeserInnen und AutorInnen) in der Blogosphäre gibt. In Ermangelung eines Zugangs zu den oft hochbegehrten (und sehr teuren) Netzverkehrsdaten, die über Online-Marktanalyse-Firmen wie Alexa (alexa.com/) oder Hitwise (hitwise.com/) beschafft werden, können ForscherInnen Verlinkungsmuster als einen vernünftigen und einigermaßen genauen Indikator für den wahrscheinlichen Netzverkehr nutzen: Die Annahme einer Korrelation zwischen eingehenden Links und eingehendem Netzverkehr ist zulässig, weil einerseits viele eingehende Links von anderen Blogs zu einer Site es wahrscheinlicher machen, dass Blog-LeserInnen die Site finden, und weil andererseits mehr

eingehende Links auch die Gesamtwertung der Site in Suchmaschinen wie Google erhöhen. Davon unabhängig können wir weiter erwarten, dass die höhere Sichtbarkeit, die aus einer größeren Besucherzahl resultiert, diesen Sites auch mehr Dominanz über und Einfluss auf die Meinungsbildung der Leserschaft gibt. Das ist natürlich wiederum eine Annahme, die erst dann genauer überprüft werden kann, wenn Forschende dazu bereit sind, einen Exkurs in das Feld der Kognitionswissenschaften zu unternehmen, wo die Frage der Auswirkungen der Medien auf die Leserschaft weiterhin heiß umstritten bleibt.

Auswirkungen des „Computational Turn“

Unter dem Einfluss unterschiedlicher fachlicher Schwerpunkte und Forschungsagenden ist das neu entstehende Feld in der Online-Forschung, welches auf diese computergestützte Ergänzung (oder für manche Beobachter, Ersetzung) konventioneller sozialwissenschaftlicher Forschungsmethoden fußt, als „Web Science“ beschrieben worden (Hendler et al., 2008). Dieser Paradigmenwechsel geht mit ähnlichen Bewegungen in den „Digital Humanities“ einher, wo die Verfügbarkeit von unüberschaubaren digitalen Datenmengen nicht nur die spezifischen Methoden verändert, sondern auch die Methodologie von bestimmten Unterdisziplinen als Ganzes (Manovich, 2007; Moretti, 2005) neu ausrichtet. Manovich (2005) prägt den Ausdruck „Cultural Analytics“, um die Art und Weise zu beschreiben, in der computerunterstützte Ansätze nicht nur die Methoden, sondern in gewisser Weise auch das Studienobjekt der geisteswissenschaftlichen Forschung selbst verwandeln. Während wir in diesem Aufsatz nicht den Platz haben, die ganze Bandbreite von erkenntnistheoretischen und methodologischen Ansätzen aufzuzeigen, die mit dem „Computational Turn“ assoziiert werden, sind WissenschaftlerInnen, die ein Interesse an den hier verfügbaren Möglichkeiten haben, angehalten, die in verschiedenen Disziplinen vorhandenen Aspekte und Zugänge kennen zu lernen und zu verstehen, wie sie sie auf ihre eigene Arbeit anwenden können. Dies muss auch in dem Wissen geschehen, dass sowohl Methoden und Methodologien als auch disziplinäre Rahmenbedingungen in diesem neuen Forschungsbereich immer noch höchst wandlungsfähig sind und bleiben.

Sichtbar wird jedenfalls, dass unabhängig von der künftigen Benennung des Forschungsfeldes der Bereich der Forschungsethik mit einiger Dringlichkeit in Angriff genommen werden muss. Die hier verwendeten computerunterstützten Forschungswerkzeuge sind leistungsfähig, und sollten nicht gedankenlos genutzt werden. Immerhin haben wir in den letzten Jahren schon umfangreiche Debatten über die öffentliche Freigabe von großen Datenmengen

und/oder über die breit angelegte Erfassung von persönlichen Daten – von Suchmaschinen-Anfragen (van Wel und Royakker, 2004), Facebook-Daten (Zimmer, 2010, *But the data*) und Twitter-Updates (Zimmer, 2010, *Is it Ethical*) – gesehen, die mit computergestützten Analysewerkzeugen, ähnlich den hier beschriebenen, erstellt wurden.

Die Macht dieser neuen Werkzeuge verschärft dramatisch die bestehenden Schwierigkeiten, etablierte forschungsethische Abläufe auf Online-Forschung, bei der es um Menschen geht, zu übertragen. Ethische Fragen entstehen in jedem Abschnitt der Online-Forschung, der sich mit von Usern geschaffenen Inhalten und Kommunikationsbeiträgen beschäftigt. Bei computerunterstützten Ansätzen zur Blogforschung betreffen sie zwei große, miteinander verbundene Problemfelder, von denen jedes Auswirkungen darauf hat, wie wir verschiedene im Forschungsprozess getroffene Entscheidungen als ethisch begreifen können. Erstens, sind Blogs „Publikationen“, oder eher mit persönlicher Kommunikation verwandt? Zweitens, sollten individuelle Bloggende deshalb als AutorInnen betrachtet oder als Forschungs-„Subjekte“ behandelt werden?

Um dieses Dilemma zu illustrieren: Computerunterstützte Blog-Forschung kann das Erfassen, Sammeln und Analysieren der tatsächlichen Inhalte einer großen Anzahl (hunderter, oder sogar tausender) Blogs mit einbeziehen. Abhängig von der schlussendlichen Form der wissenschaftlichen Analyse wird eine Entscheidung getroffen werden müssen, ob diese Inhalte wörtlich zitiert werden können (etwa um Befunde zu veranschaulichen), und wenn ja, ob sie ihren AutorInnen zugeschrieben oder die Zuschreibung anonymisiert werden soll. Um das zu beantworten, müssen Forschende im Hinblick auf die Sammlung und spätere öffentliche Zurverfügungstellung der Daten auf Fragen über Privatsphäre sowie schriftstellerische Selbstbestimmung (z. B. in Hinsicht auf Fragen des Urheberrechts) eingehen. Dazu gibt es allerdings keine einfachen und direkten Antworten, und sogar in Kleinprojekten hat sich dies bisher als kontrovers erwiesen (siehe eine ausführliche Diskussion in Bruckman, 2002). Eine Möglichkeit wäre, dass Blogs ganz einfach Publikationen sind, und deshalb ihr Inhalt zitiert und im Rahmen des Urheberrechts wiederverwendet werden kann; tatsächlich kann eine solche Wiederverwendung von Blog-Material dem kreativen Werk der Blog-AutorInnen zusätzliche Beachtung verschaffen. Aus einer Perspektive aber, die die Vielfalt der eigentlichen Praktiken und Diskussionen der Bloggenden selbst berücksichtigt, könnte man zu dem Schluss kommen, dass zwar einige Blogs, speziell aus dem Bereich der Politik und des Journalismus, eindeutig „öffentlicher“ Natur sind, andere aber als private Medien für ein begrenztes gedachtes Publikum funktionieren – eine persönliche, personenbezogene Öffentlichkeit, wie Schmidt (2011) sie beschreibt –, so dass eine von der ursprünglichen Intention abweichende Verbreitung dieser Bloginhalte die Privatsphäre oder Handlungsfreiheit der AutorInnen verletzt.

Es ist unmöglich, die Antworten auf diese Fragen im Voraus zu finden, weil Bloggen kulturell komplex und vielfältig ist und sich die Antworten selbst im Fall von einzelnen Blogs im Laufe der Zeit oder aus vielfältigen anderen Umständen ändern können. Im Zusammenhang mit Großprojekten, die auf automatisierter Datenerfassung und -analyse beruhen, ist zudem die sonst gängige Praxis des Einholens einer Nutzungs- und Analyseerlaubnis von allen Untersuchungsteilnehmenden zur Umgehung der ethischen Problematik in der Regel impraktikabel. Aufgrund dieser Komplexität sind die etablierten sozialwissenschaftlichen Forschungsabläufe selten geeignet, solchen Themen adäquat zu begegnen, weshalb es wichtig ist, dass angehende Blog-forscherInnen sich gründlich mit den einschlägigen Diskussionen und den aktuellen Meinungen zur bestehenden Best Practice im Bereich der Internet-Forschungsethik auseinandersetzen. Nur das kann einen gut begründeten Standpunkt garantieren, wenn Fälle vor einen ethischen Untersuchungsausschuss oder eine Ethikkommission gebracht werden. Es gibt einige maßgebliche Online-Forschungs-Handbücher, die diese sich entfaltenden Debatten zusammenfassen (s. z. B. Buchanan, 2004; Burnett et al., 2010). Zusätzlich bietet die Association of Internet Researchers (aoir.org/), die sich auch mit dem Thema Ethik in der Internetforschung beschäftigt, ein nützliches Handbuch über derzeit generell einvernehmlich gesehene Bereiche und über fortlaufende Debatten in der Internet-Forschungsgemeinschaft (Ess et al., 2002).

Eine andere Herausforderung für die an den meisten Universitäten üblichen ethischen Prüfvorgänge ist, dass es speziell für großangelegte Forschungsprojekte in einem neu entstehenden Forschungsfeld nicht immer möglich ist, im Planungsstadium alle ethischen Fragen vorzusehen und anzusprechen – geschweige denn, Lösungen für Fragen zu finden, die im Laufe des Projektes entstehen werden –, da das Forschungsobjekt selbst ständig im Wandel begriffen ist. Forschung, die die hier beschriebenen großangelegten computerunterstützten Methoden nutzt, wird nicht nur unmittelbar zwischen den komplexen ethischen Wägbarkeiten navigieren müssen, sondern trägt auch mit jeder methodologischen und daraus folgenden technologischen Entscheidung dazu bei, diese ethischen Problemkomplexe weiter auszuformen. Deshalb ist es wichtig, dass ein reflexiver und offener Ansatz, der von den Debatten rund um die Internetforschungsethik geprägt ist und auf sie eingeht, Eingang in die während des Forschungsprozesses erwachsende Methodologie findet.

Eine englischsprachige Version dieses Artikels erschien zuerst unter dem Titel „Doing Blog Research“ in: Research Methods & Methodologies in Education (Thousand Oaks, Calif.: Sage, 2012), herausgegeben von James Arthur et al.

Weitere empfohlene Lektüre

Jill Walker Rettberg: *Blogging*. Cambridge 2008.

Als eine der ersten Akademikerinnen, die Blogs und Blogging untersuchte, bietet Walker eine grundlegende Einführung zur Geschichte der Form, und erläutert aktuelle Themen, Fragen und Forschungsmöglichkeiten.

Richard Rogers: *The End of the Virtual: Digital Methods*. Antrittsvorlesung, University of Amsterdam, 08.05.2009 [www.govcom.org/publications/full_list/oratie_Rogers_2009_preprint.pdf], eingesehen 15.12. 2010.

Rogers, der Gründer von Govcom.org, einer unabhängigen Organisation, die das Forschungswerkzeug IssueCrawler zur Verfügung stellt, beschreibt den ‚Computational Turn‘ in den Internetforschungsmethoden und erläutert die sich daraus ergebenden Herausforderungen.

Nancy K. Baym: *Personal Connections in the Digital Age*. Cambridge 2010.

Indem sie soziale und mobile Medien in der weiteren Medienumwelt ansiedelt, deren Teil sie sind, demonstriert Baym die wachsende Wichtigkeit von nichtmassenmedialen Formen für das tägliche Leben, Beziehungen und Kommunikation. Sie zeigt deutlich auf, warum weitere Forschung in diesem Feld notwendig und wichtig ist.

Nigel Shadbolt/Tim Berners-Lee: *Web science emerges*, in: *Scientific American*, Oct. 2008, S. 32–37.

Mitverfasst von Tim Berners-Lee, dem Erfinder des World Wide Web, erklärt dieser kurze Artikel die Ideen hinter der „Web Science“-Initiative für eine quantitative, daten-getriebene Erforschung der Nutzung des Webs. Er enthält auch nützliche weitere Referenzen.

Fragen zur weiteren Ermittlung

- Welches analytische Rahmenwerk existiert, um die Daten, die jetzt von Blogs und anderen Social-Media-Plattformen in großem Ausmaß erfasst werden können, zu hinterfragen und auszuwerten?
- Welche Unterschiede oder Ähnlichkeiten existieren zwischen bestimmten (thematisch, demographisch oder anders definierten) Abschnitten innerhalb der gesamten Blogosphäre, und/oder zwischen verschiedenen Blogosphären, die sich etwa durch ethnische Zugehörigkeit, Sprache, Geographie oder Nationalität differenzieren?

- Was ist die Dynamik der Blogaktivitäten: welche jahreszeitlichen Muster weist Blogging-Aktivität auf, und wie funktioniert das Kommen und Gehen von Popularität in der Blogosphäre?
- Wie verhalten sich Blogs zu anderen massenmedialen oder Social-Media-Öffentlichkeiten: wie bewegen sich Informationen durch die weitere Mediumwelt, und wie wird Bloggen von Entwicklungen in jüngeren Social-Media-Plattformen beeinflusst?
- Welche forschungsethischen Grundregeln bestimmen das wissenschaftliche Erfassen, Analysieren und anderweitige Befassen mit Inhalten, die, obschon öffentlich erhältlich, vielleicht nur für eine kleine persönliche Audienz bestimmt waren?

Referenzen

L. Adamic/N. Glance: The political blogosphere and the 2004 U.S. election: Divided they blog. Proceedings of the 3rd International Workshop on Link Discovery 2005 [www.blogpulse.com/papers/2005/AdamicGlanceBlogWWW.pdf], eingesehen 19.01.2009.

D. M. Berry (Hrsg.): The Computational Turn: The Digital Humanities and New Technology. London 2010.

A. Bruckman: Studying the amateur artist. A perspective on disguising data collected in human subjects research on the Internet, in: Ethics and Information Technology, 4, (3), 2002, S. 217–231.

A. Bruns/J. Jacobs (Hrsg.): Uses of Blogs. New York 2006.

A. Bruns/J. Wilson/B. Saunders u.a.: Australia's Political Blogosphere in the Aftermath of 2007 Federal Election. Paper presented at the AoIR 2008 conference. Copenhagen, 18 Oct. 2008 [snurb.info/files/aoir2008/Australia%27%20Political%20Blogosphere%20in%20the%20Aftermath%20of%20the%202007%20Federal%20Election%20(AoIR%202008).pdf], eingesehen 07.04. 2010.

A. Bruns/J. Wilson/B. Saunders u.a.: Locating the Australian Blogosphere: Towards a New Research Methodology. Paper presented at the ISEA 2008 conference, Singapore, 25 July – 3 August 2008 [snurb.info/files/Locating%20the%20Australian%20Blogosphere%20(final%20%20long).pdf], eingesehen 07. 04. 2010.

- E. Buchanan (Hrsg.): Readings in virtual research ethics: issues and controversies. London 2004.
- R. Burnett/M. Consalvo/C. Ess: The Handbook of Internet Studies. London 2010.
- J. Burgess: Blogging to learn, learning to blog, in A. Bruns und J. Jacobs (Hrsg.): Uses of Blogs. New York, S. 105–114.
- C. Ess und AoIR ethics working group: Ethics Guide [aoir.org/documents/ethics-guide/], eingesehen 27.05.2010.
- H. Farrell/D.W.Drezner: The Power and Politics of Blogs, in: Public Choice, 134, 2008, S. 15–30.
- L.J. Gurak u. a. (Hrsg.): Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs, 2004 [blog.lib.umn.edu/blogosphere/], eingesehen 27.05.2010.
- J. Hendler u. a.: Web Science: An Interdisciplinary Approach to Understanding the Web, in: Communications of the ACM, 51/7, 2008, S. 60–69.
- L. Kirchhoff u. a.: Monitoring the Australian Blogosphere through the 2007 Australian Federal Election. Paper presented at ANZCA 2009, Brisbane, 8 July 2009 [snurb.info/files/talks2009/Monitoring%20the%20Australian%20Blogosphere%20through%20the%202007%20Australian%20Federal%20Election.pdf], eingesehen 07.04.2010.
- L. Manovich: Cultural analytics, 2007 [www.manovich.net/cultural_analytics.pdf], eingesehen 24.04.2010.
- F. Moretti: Graphs, Maps, Trees: Abstract Models for a Literary History. London 2005.
- A. Russell/N. Echchaibi: International Blogging: Identity, Politics, and Networked Publics. New York 2009..
- J. Schmidt: Weblogs: *Eine kommunikationssoziologische Studie*. Konstanz 2009.
- J.-H. Schmidt: Das neue Netz: Merkmale, Praktiken und Folgen des Web 2.0. 2nd ed. Konstanz: 2011.
- J. van Dijck: Composing the Self: Of Diaries and Lifelogs, in: Fibreculture Journal, 3, 2004. [www.journal.fibreculture.org/issue3/issue3_vandijck.html].
- L. van Wel/L. Royakker: Ethical Issues in Web Data Mining, in: Ethics and Information Technology, 6, 2004, S. 129–140.
- J. Walker Rettberg: Blogging. Malden, MA 2008.
- M. Zimmer, M.: „But the data is already public“: on the ethics of research in Facebook, in: Ethics and Information Technology, 4, (3), 2010, S. 205–216.

M. Zimmer: Is it Ethical to Harvest Public Twitter Accounts Without Consent?, 2010 [michaelzimmer.org/2010/02/12/is-it-ethical-to-harvest-public-twitter-accounts-without-consent/].