

## **Digitale Spuren: Möglichkeiten der „bedeutungsvollen“ Analyse von komplexen Online-Daten**

Die Auseinandersetzung mit den Möglichkeiten von „big data“ (Mayer-Schonberger/Cukier 2013, Kitchin 2014) hat sich zu einer wichtigen Diskussion in der Kommunikations- und Medienforschung entwickelt: Die großen, prinzipiell verfügbaren Datenmengen – bspw. durch Google oder Twitter – würden erhebliche Chancen für die Medienforschung eröffnen. Das Argument hier ist, dass die großen Datenmengen es gestatten würden, theoretische Annahmen auf ganz neue Weise zu prüfen bzw. zu falsifizieren, insbesondere weil diese Daten nicht von den Forschenden generiert werden würden, sondern gewissermaßen in der „natürlichen“ Praxis entstünden und non-responsive Untersuchungsdesigns erlaubten.

Aus Sicht der Informatik bestehen einerseits Zweifel an der Neuheit des Ansatzes, da schon immer mit den zu der jeweiligen Zeit existierenden großen Datenmengen gearbeitet wurde. Andererseits stellt die Echtzeitverarbeitung als auch die Datenvisualisierung neue Herausforderungen an die Algorithmen und die zugrundeliegenden Ontologien bzw. Datenrepräsentationen. Dabei wird als zentrale Problemlage die semantische Informationsverarbeitung identifiziert (Leskovec et al. 2014, Jagadish et al. 2014, Chen et al. 2014). Ebenso werden Vorstellung von „big data“ zunehmend in der Kommunikations- und Medienforschung kritisiert (Golder/Macy 2014). Der Hinweis hier ist, dass getrieben von dem Wunsch, große Datenmengen zu verwenden, die Entstehungskontexte und deren Einschreibung in die Algorithmen der Datengenerierung nicht hinterfragt werden bzw. die Daten selbst weitgehend „bedeutungslos“ in dem Sinne sind, dass sie kaum einen Rückbezug auf soziale bzw. kommunikative Praktiken der Bedeutungsproduktion ermöglichen (Boyd/Crawford 2012; Pasquale 2015; Lohmeier 2014). Dies hat seine Entsprechung in den bisher noch mangelhaften Ontologien aus der Informatik, die Kontextwissen nur begrenzt repräsentieren.

Beide Kritiken führen uns vor Augen, dass der Hype um „big data“ nicht vergessen machen darf, dass die Möglichkeiten einer auf (großen) Online-Datenmengen basierenden Forschung wesentlich limitierter sind, als es die Apologeten suggerieren. Dies hat nicht zuletzt etwas mit der Komplexität dieser Daten zu tun: Große Datenmengen gehen häufig mit geringen Kontextinformationen einher. Die Herausforderung besteht also darin, in der Vielfalt der Datenmengen über „bedeutungsvolle“ Daten zu verfügen, also solche, bei denen weitergehende Information zur sozialen Praxis bzw. Bedeutungsproduktion vorhanden sind.

Ausgehend von dieser Diskussion möchten wir das Konzept der „digitalen Spuren“ („digital traces“) vorschlagen. Dieses Konzept basiert auf Überlegungen der „digital methods“ (Rogers 2013) und hier vor allem dem Argument, dass online-basierte Daten es gestatten, auf bestimmte Phänomene aufmerksam zu werden, die dann durchaus auch in weitergehender Forschung beschrieben werden sollen. „Digitale Spuren“ sind dabei für uns erst einmal all diejenigen Daten, die eine Person oder Organisation generiert, wenn sie online aktiv ist. Ein Teil dieser Daten ist für die be-

treffenden Personen oder Organisationen durchaus zugänglich und offensichtlich, bspw. in der „time line“ von Facebook oder Twitter. Ein weiterer Teil der Spurendaten wird aber – insbesondere zur Generierung von Werbeinformationen (Turow 2011) – auf nicht zugängliche Weise hergestellt. Für eine „bedeutungsvolle Analyse“ bedeutet dies nun, beide Arten von Daten zusammen zu bringen um so zu einer vertiefenden Gesamtinterpretation gelangen zu können.

Unser Vortrag strukturiert sich in vier Teile: Erstens wollen wir die Diskussion um „big data“ als Kennzeichen der heutigen, komplexen Mediumgebung aufgreifen und sowohl aus Sicht der Informatik als auch aus Sicht der Kommunikations- und Medienwissenschaft kritisieren. Auf dieser Basis werden wir dann zweitens unser Konzept der „digitalen Spuren“ mit Bezug auf die generelle Diskussion um „digital methods“ vorstellen. Drittens wollen wir anhand eines empirischen Beispiels das Potenzial eines solchen Vorgehens verdeutlichen. Dabei rekurrieren wir auf unsere Studie zur Logfile-Analyse auf einer Kollaborationsplattform über sechs Monate, bei der mit Hilfe von Methoden des Data Mining Cluster identifiziert wurden (Schulz/Breiter 2013), deren Kontextualisierung in Form einer qualitativen Analyse erst möglich wurde. Abgeschlossen wird der Vortrag viertens durch ein Fazit, in dem wir die Perspektiven einer Forschung zu „digitalen Spuren“ verdeutlichen.

## **Literatur:**

Boyd, D. / Crawford, K. (2012): Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. In: *Information, communication & society*, 15 (5), S. 662-679.

Chen, C. L. Philip, & Zhang, Chun-Yang. (2014). Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data. *Information Sciences*, 295(August), 314-347.

Golder, Scott A., & Macy, Michael W. (2014). Digital Footprints: Opportunities and Challenges for Online Social Research. *Annual Review of Sociology*, 40, 129-152.

Jagadish, H. V., Gehrke, Johannes, Labrinidis, Alexandros, Papakonstantinou, Yan-nis, Patel, Jignesh M., Ramakrishnan, Raghu, & Shahabi, Cyrus. (2014). Big data and its technical challenges. *Commun. ACM*, 57(7), 86-94.

Kitchin, Rob. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.

Leskovec, Jure, Rajaraman, Anand, & Ullman, Jeffrey D. (2014). *Mining of Massive Datasets* (2nd ed.). Cambridge, MA: Cambridge University Press.

Lohmeier, C. (2014): The Researcher and the Never-Ending Field: Reconsidering Big Data and Digital Ethnography. In: Hand, M. / Hillyard, S. Emerald Group Publishing Limited, S. 75-89.

Mayer-Schonberger, V. / Cukier, K. (2013): *Big data: A revolution that will transform how we live, work and think*. New York: John Murray.

Pasquale, F. (2015): *The black box society: The secret algorithms that control money and information*. Cambridge: Harvard University Press.

Rogers, R. (2013): *Digital Methods*. Cambridge, MA: MIT Press.

Schulz, Arne Hendrik, & Breiter, Andreas. (2013). Monitoring User Patterns in School Information Systems Using Logfile Analysis. In Don Passey, Andreas Breiter & Adrie J. Visscher (Eds.), *Next Generation of IT in Educational Management* (pp. 94-103). Berlin: Springer.

Turow, J. (2011): *The daily you: How the new advertising industry is defining your identity and your worth*. New Haven and London: Yale UP.