

*Workshop des Lehrbereichs Computational Social Sciences (LMU München)
und der Forschungsgruppe digital humanities (FernUni Hagen)*

2./3. Juni 2022, online

Mensch und Maschine

Zum Vergleich von Topic Modeling und manueller Annotation in der qualitativen Sozialforschung

Topic Modeling hat sich als Heuristik zur Erschließung großer Textkorpora in den Digital Humanities (DH) und Computational Social Sciences (CSS) etabliert.¹ Es wird auch als Filter für qualitative Forschung bezeichnet.² In der qualitativen Forschung hat die Anwendung von Verfahren des maschinellen Lernens zur Analyse ein Potenzial zur Aufbereitung größerer Datenmengen etwa im Theoretical Sampling nach Grounded Theory. Indem beispielsweise durch gut trainierte Topic Models aus kompletten Sammlungen lebensgeschichtlicher Interviews thematische Zusammenhänge extrahiert werden, ist es möglich, einen ersten inhaltlichen Überblick zu erhalten.³ Darüber hinaus können bei explorativer Durchsicht der Ergebnisse unerwartete Phänomene an die Oberfläche gespült werden, die in den vielschichtigen Interviews allzu leicht verschütt gehen.⁴ Gerade für Sekundäranalysen ist das Verfahren deshalb vielversprechend, es kann aber auch auf andere Daten, beispielsweise Diskursquellen, angewendet werden. Allerdings fehlt es einerseits grundsätzlich an qualitativen Evaluationsroutinen im Machine Learning, um die automatisch generierten Ergebnisse zu validieren.⁵ Andererseits mangelt es in den DH und CSS an systematischen Studien, die computerisierte und menschliche Inhaltserschließung vergleichen.⁶ Gleichzeitig ist die Hemmschwelle zum Einstieg in die Verwendung digitaler Methoden bei qualitativ Forschenden besonders hoch.⁷

¹ Shawn Graham/Ian Milligan/Scott Weingart (2015): Exploring Big Historical Data: The Historian's Macroscope. London; Matthias Lemke/Gregor Wiedemann (2016): Text Mining in den Sozialwissenschaften: Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse. Wiesbaden.

² Gertraud Koch/Lina Franken (2020): Filtern als digitales Verfahren in der wissenssoziologischen Diskursanalyse. Potenziale und Herausforderungen der Automatisierung im Kontext der Grounded Theory. In: Samuel Breidenbach/Peter Klimczak/Christer Petersen (Hg.): Soziale Medien. Interdisziplinäre Zugänge zur Onlinekommunikation. Wiesbaden, S. 121–138.

³ Tobias Hodel/Dennis Möbus/Ina Serif (im Erscheinen): Topic Modeling im Vergleich: Aufbereitung, Umsetzung und Interpretation unterschiedlicher historischer Textkorpora. Hagen.

⁴ Dennis Möbus (2022): Holleriths Vermächtnis – ein Beitrag zur Geschichte von Frauen in der EDV. Topic Modeling als Methode digitaler Sekundäranalyse lebensgeschichtlicher Interviews. In: BIOS 33 (1).

⁵ James Dobson (2021): Interpretable Outputs. Criteria for Machine Learning in the Humanities. In: Digital Humanities Quarterly 15 (2). Online: <http://www.digitalhumanities.org/dhq/vol/15/2/000555/000555.html>.

⁶ Der Workshop ist etwa inspiriert von: Peter Andorfer (2017): Turing Test für das Topic Modeling. Von Menschen und Maschinen erstellte inhaltliche Analysen der Korrespondenz von Leo von Thun-Hohenstein im Vergleich. In: Zeitschrift für digitale Geisteswissenschaften 2. Online: https://zfdg.de/2017_002; Martin Fechner, Andreas Weiß (2017): Einsatz von Topic Modeling in den Geschichtswissenschaften: Wissensbestände des 19. Jahrhunderts. In: Zeitschrift für digitale Geisteswissenschaften 2. Online: https://zfdg.de/2017_005; Melanie Andresen, Michael Vauth, Heike Zinsmeister (2020): Modeling Ambiguity with Many Annotators and Self-Assessments. In Proceedings of the 14th Linguistic Annotation Workshop, Barcelona, S. 48–59. <https://aclanthology.org/2020.law-1.5/>.

⁷ Lina Franken (2020): Kulturwissenschaftliches digitales Arbeiten. Qualitative Forschung als ›digitale Handarbeit‹? In: Berliner Blätter – Ethnographische und ethnologische Beiträge 82, S. 107–118. Online: <https://www2.hu-berlin.de/ifeeojs/index.php/blaeetter/article/view/1069/16>.

Wie können diese neuen Verfahren qualitative Analysen bereichern? Welche Unterschiede macht es, wenn Interviewmaterialien durch ein Topic Modeling strukturiert gesichtet werden? Wie unterscheiden sich die Perspektiven auf (unbekannten) Text?

Der Workshop versucht diesen Fragen in einem von Turing inspirierten Experiment zu begegnen. Es gilt, die verschiedenen Zugänge zu vergleichen, indem in einer Gruppe zunächst händisch annotiert wird (Kategorienvergabe oder Codierung), während die andere mit einer Erschließung unbekannter Texte durch Topic Modeling startet. Danach tauschen beide Gruppen die Rollen, um die Unterschiede im Anschluss systematisch vergleichen zu können. Im Mittelpunkt steht dabei das konkrete Ausprobieren und Erlernen der methodischen Schritte sowie deren gemeinsame Reflexion.

Am Ende des Workshops haben die Teilnehmenden erste Einblicke sowohl in die qualitative Annotation mittels Grounded Theory als auch in das digitale Verfahren des Topic Modelings gewonnen. Sie können beide Methoden selbstständig anwenden und erhalten Hinweise, um das Wissen für eigene Forschungen zu vertiefen. Außerdem bekommen sie eine Einführung in die Nutzung konkreter Tools und Werkzeuge an die Hand, um die Verfahren computationell umzusetzen, insb. ein Skript zum Erstellen von Topic Models sowie ein Open Source Tool zum manuellen Annotieren. Darüber hinaus erhalten sie Einblicke in ein umfangreiches, digital erschlossenes Korpus von lebensgeschichtlichen Interviews zur Sozialgeschichte des 20. Jahrhunderts.

Zielgruppe des Workshops sind Studierende, Promovierende und Forschende aus den Geistes- Kultur- und Sozialwissenschaften, die sich für digitale Methoden interessieren, sowie Studierende, Promovierende und Forschende aus den DH und CSS, die vertiefte Einblicke in die Veränderungen von qualitativen Forschungsprozessen durch digitale Verfahren erlangen möchten.

Der Workshop findet am 2. und 3. Juni 2022 (jeweils 10 bis 17 Uhr) online statt. Anmeldungen mit einer kurzen Beschreibung der eigenen Motivation (500-1.000 Zeichen) bitte bis zum 22.04.2022 an Florian Schmid, florian.schmid@gsi.uni-muenchen.de.