

Gutachten zur Masterarbeit von: Lukas Heinzmann

Titel: Der Stand der Handschriftenerkennung im ABD Kontext

Name des Gutachters: Tobias Hodel

Vertreter / Vertreterin der Studienleitung:

Die im Rahmen des MAS ALIS der Universitäten Bern und Lausanne entstandene Masterarbeit beschäftigt sich mit der Anwendung von Topic Modeling anhand eines Quellenkorpus, bestehend aus Tagebucheinträge des Einsiedler Paters Joseph Dietrich (1645-1704). Die Arbeit thematisiert damit einen Bereich der Information Extraktion, die in den Computerwissenschaften seit zwanzig und in den digitalen Geisteswissenschaften seit zehn Jahren diskutiert wird. Trotz dieser bereits ausgiebigen Diskussionen vermag die Arbeit neue Akzente zu setzen und besticht durch konzeptionelle Schärfe und analytische Klarheit.

Ausgangspunkt ist ein klar definiertes Korpus, das bereits ein Subkorpus einer umfangreichen Edition bildet. Die Wetterbeobachtungen von Pater Dietrich stellen ein Forschungsergebnis dar, das in der Arbeit in einem *distant reading* Verfahren analysiert wird. Wie Heinzmann selber darlegt, bilden solche und ähnliche grosse Quellenmassen eine Herausforderung für den ABD-Sektor, der in den kommenden Jahren an Wichtigkeit gewinnen wird.

Obwohl das Quellenkorpus durch den Autor im Rahmen einer Dissertation bearbeitet wird, handelt es sich hierbei um eine vollständig unabhängige Arbeit, die zwar auf dasselbe Quellenkorpus abstützt, insgesamt aber nur minim von den Vorarbeiten durch die im Entstehen begriffene Promotionsarbeit profitiert.

Aufbau

Die Arbeit ist klar aufgebaut und leistet nach einer konzisen Einleitung, die die Methode ein erstes Mal skizziert, vertiefte Einsichten in die unterschiedlichen Prozessschritte von *topic modeling* (preprocessing, Modellierungsprozesse und postprocessing). Innerhalb dieses Kapitels wird nicht nur auf die technischen Nuancen und Herausforderungen eingegangen, sondern eine unglaublich tiefe, beeindruckende Literaturreview vorgelegt. Der Rundumblick bleibt dabei

nicht bei den unterschiedlichen Einsatzszenarien und Perspektiven stehen, sondern zeigt auf, inwiefern übereinstimmende und widersprüchliche Schlüsse aus und für die Methode gezogen werden. Ausgehend von diesen Einblicken wird das eigene Korpus analysiert und Schlüsse daraus gezogen. Die tiefgreifenden Einsichten, die aus dem Korpus mittels *topic modeling* gewonnen werden können, müssen im nächsten Abschnitt noch eingehender gewürdigt werden. Ein kurzes Fazit mit Ausblick rundet die Arbeit ab und ein enorm hilfreicher Anhang (zentral für die Nachvollziehbarkeit) führt zu einem aussergewöhnlich überlegt aufgebautem Dokument.

Anwendungs- und Praxisorientierung

Sowohl der theoretische Hintergrund des *topic modeling* als auch die eigenständige Umsetzung der Arbeit mit einem Korpus führt zu einer optimalen Verknüpfung von Anwendungsfragen mit methodenkritischen Bedenken. Dabei wird aufgezeigt, dass mit sinnvoll aufbereiteten Korpora tiefgreifende Schlüsse gezogen werden, die sich insbesondere auf den Inhalt, aber auch auf orthografische Eigenheiten Dietrichs sowie Anpassungen darin beziehen.

Die Ausführungen zu den gewonnenen Topics wird so zur Auslegeordnung, die nicht nur die Fähigkeiten der Methode und daraus gezogene Einsichtsgewinne vorzeigt, sondern auch Fragen der Segmentierung von Dokumenten analysiert (und teilweise problematische Aussagen des Gutachters in wissenschaftlichen Publikationen offenlegt).

Der Umgang mit dem Material mittels digitaler Analysemethoden entspricht dabei dem *state-of-the-art*, die unbedingt eine Publikation der Arbeit nach sich ziehen sollte. Insbesondere die Verknüpfung mit dem Konzept des *scalable reading* überzeugt vollends, wobei auch das *stacking* der Methoden, hinzu einem «*closer*» *reading* mittels den Voyant Tools ausgesprochen kreativ ist und gleichzeitig zu neuen Einsichten verhilft.

Theoretische Fundierung

Die Literaturreview ist wie bereits oben angesprochen von höchster Qualität und bringt Argumentationsnetze zum Vorschein, die mir persönlich nicht bekannt waren. Dank der Nutzung von sog. Notebooks (via *observablehq*) wird die Theorie mit Visualisierungsformen verknüpft, die schliesslich zu einer sinnvollen Narration der Nutzung von Topic Modeling im Forschungs- und ABD-Bereich zusammengefügt wird.

Sprachliche Kohärenz

Die Sprache ist sehr gewählt, korrekt und exakt. Dennoch ist der Text gut lesbar und hält wissenschaftlichen Ansprüchen stand. Fachbegriffe werden gezielt und korrekt eingesetzt.

Formales

Die Arbeit ist sorgfältig aufgebaut und Formalia werden regelkonform eingehalten. Es gibt keine Beanstandungen.

Gesamtbeurteilung

Es handelt sich um eine hervorragende Arbeit, die eine Methode des *distant reading (topic modeling)* in ihrer Anwendung fachgerecht situiert und einem Korpus neue Bedeutungen entlockt. Die Mischung zwischen Anwendung, methodischen Abwägungen und Einsichten in die (mathematische) Theorie des Ansatzes vermögen zu überzeugen und bestätigen, dass bereits mit mittelgrossen Korpora interessante Einsichten gewonnen werden können.

Wie bereits oben angedeutet, wird eine Publikation dringend empfohlen.

.....

Der unterzeichnende Gutachter beantragt hiermit der Programmleitung, die Masterarbeit mit der Note 6,0 zu bewerten.

Ort, Datum

Bern, 8.09.2020

Unterschrift des Gutachters / der Gutachterin:



.....