

Kapitel 7 – Quantitative Software zur Textanalyse

Margrit Schreier

Frühe Formen quantitativer Software waren darauf ausgerichtet, die inhaltsanalytische Auswertung manifester Textbedeutung zu automatisieren (z. B. General Inquirer, Textpack, Intext, LIWC; im Überblick Alexa & Zuell, o.J.). Je nach Fragestellung wird ein sog. Dictionary erstellt. Dieses Dictionary erfüllt die Rolle des Kategoriensystems: Für jede Kategorie werden (im Sinne einer Kategorienexplikation) Suchbegriffe definiert. Bei einer Analyse von Leitartikeln deutschsprachiger Tageszeitungen wäre etwa eine Kategorie „Wirtschaft“ vorstellbar; diese würde Suchbegriffe enthalten wie: Wirtschaft, Finanz*, Kapital*, Börse usw. Das Programm durchsucht die zu analysierenden Texte nach diesen Suchbegriffen; für jeden „Fund“ wird dem Textteil automatisch die Kategorie zugewiesen, die durch den Suchbegriff spezifiziert ist. Programme dieses Typs realisieren also eine „Arbeitsteilung“ zwischen Mensch und Computer: Der Mensch muss das Dictionary spezifizieren; die eigentliche Codierung wird vom Computer übernommen (kann aber ihrerseits immer nur so gut sein wie das Dictionary). Diese Programme stellen gerade bei der Analyse großer Textmengen eine erhebliche Arbeitserleichterung dar, sind aber, wie erwähnt, in der Anwendung weitgehend auf die Analyse manifester Textbedeutungen beschränkt.

Im Zusammenhang mit der Verfügbarkeit großer Datenmengen („Big Data“) aus digitalen Medien hat in den vergangenen Jahren das sog. **Text Mining** zunehmend an Bedeutung gewonnen. Es handelt sich dabei um die automatisierte Analyse großer Mengen verbaler Daten auf der Grundlage von Natural Language Processing (NLP), um Informationen und Informationsmuster sichtbar zu machen, die nicht auf den ersten Blick erkennbar sind (Zhai & Massung, 2016, Kap. 1). Zu den bekanntesten Formen des Text Mining zählen die Analyse von Schlüsselworten in ihrem jeweiligen Kontext (KWIC: Keywords in Context); das Topic Mining, wobei aus dem Material Themen extrahiert und dargestellt werden; die Sentiment Analysis zur Bestimmung der emotionalen Valenzen von Äußerungen; die Netzwerkanalyse; sowie die Vorhersage von Verhaltensweisen auf der Grundlage von Äußerungen (Predictive Analysis). Von den früheren Formen quantitativer Software unterscheidet sich das Text Mining vor allem dadurch, dass hier eine Vielzahl von Variablen gleichzeitig analysiert und untereinander in Zusammenhang gesetzt werden (der sog. Feature Space bzw. Merkmalsraum). Einerseits handelt es sich bei den entsprechenden Programmen um ausgesprochen effiziente Werkzeuge zur automatisierten Textanalyse.

Andererseits hat natürliche Sprache Eigenschaften, denen eine rein automatisierte Bedeutungsanalyse nicht oder nicht hinreichend gerecht werden kann, wie beispielsweise semantische Mehrdeutigkeit, die Kontextabhängigkeit von Bedeutung oder implizite Bedeutungsdimensionen. Hier bietet sich eine Kombination quantitativer Analyse größerer Textmengen und eine vertiefte Analyse mittels qualitativer Auswertungsverfahren an.

Literatur

Alexa, M. & Zuell, C. (o.J.). *A review of software for text analysis* (ZUMA Nachrichten spezial, Bd. 5). GESIS/ZUMA.

Zhai, C. & Massung, S. (2016). *Text data management and analysis: A practical introduction to information retrieval and text mining*. Association for Computing Machinery and Morgan & Claypool Publishers.