

Textanalyse für Sozialwissenschaften

Masterseminar
Frühjahrssemester 2023

Dozent: Dr. Ueli Reber (ueli.reber@eawag.ch)

Zeit: Montag, 12:15–14:00

ECTS: 6

Raum: 004 (Fabrikstr. 2e)

Kurzbeschreibung

Verschiedene Formen der Textanalyse gehören schon lange zum methodischen Repertoire der Sozialwissenschaften. Aufgrund der hohen Verfügbarkeit von Text und neuen computergestützten Verfahren ist die Textanalyse heute jedoch so populär wie nie. In diesem Methodenseminar beschäftigen wir uns vertieft mit der Analyse von Texten und behandeln dabei nicht nur praktische Fragen des «Wie», sondern auch theoretische des «Weshalb» und «Wofür». Im Zentrum stehen dabei konkrete Forschungsfragen, für deren Bearbeitung wir quantitativ-textanalytische Lösungen entwickeln und die es uns auch erlauben, die Vorteile und Grenzen der Herangehensweise zu reflektieren. Der Fokus der Veranstaltung liegt auf computergestützten Analyseverfahren, wobei jedoch auch manuelle Ansätze angesprochen werden.

Lernziele

Die Studierenden...

- kennen die zentralen Konzepte, Vor- und Nachteile sowie Grenzen der quantitativen Textanalyse.
- können einfache textanalytische Forschungsdesigns entwickeln.
- können sich Texte aus dem Internet beschaffen (Web scraping, APIs) und für die Analyse aufbereiten.
- beherrschen und verstehen verschiedene überwachte und unüberwachte Verfahren der computergestützten Textanalyse (Informationsextraktion, Textklassifikation).
- können die Resultate von Textanalysen kommunizieren.

Didaktisches Konzept

Das Seminar gliedert sich in zwei Phasen: (1) eine individuellen, die *Sitzungsvorbereitung*, und (2) eine kollektive, die *Sitzung*. Bei der Sitzungsvorbereitung steht die selbständige Erarbeitung der Inhalte im Vordergrund, also das Lesen der Texte, das Vertrautmachen mit der Methode/Technik sowie das Lösen von Übungsaufgaben. Die Sitzung dient dann der Diskussion der Texte und Übungsaufgaben, dem Klären von Verständnisfragen und der gemeinsamen Arbeit an Übungen und dem eigenen Gruppenprojekt.

Leistungsnachweis

Der Leistungsnachweis besteht in der Durchführung einer eigenen Textanalyse von der Forschungsfrage und dem -design über die Textbeschaffung, -aufbereitung und -analyse bis hin zur «Publikation» in Form einer schriftlichen Arbeit. Es handelt sich dabei um ein Gruppenprojekt in Gruppen à zwei bis drei Personen.

20.02.2023 #1 Einführung

Vorbereitung

- Podcast: [«Tech and Politics: Text analysis»](#)

Sitzung

- Einstieg in die Textanalyse für Sozialwissenschaften
- Übersicht Seminar
- Gruppenprojekt: Kennenlernen

27.02.2023 #2 Konzeptionelle Grundlagen von Textanalyse

Vorbereitung

- Lektüre: Kapitel 1 «History» (S. 10-23) und Kapitel 2 «Conceptual Foundation» (S. 24-50) aus Krippendorff, K. (2019). *Content analysis: an introduction to its methodology*. Sage.

Sitzung

- Diskussion: konzeptionelle Grundlagen
- Gruppenprojekt: Gruppenbildung, Forschungsfrage

06.03.2023 #3 Texte beschaffen 1

Vorbereitung

- Lektüre: Kapitel 6 «Sampling» (S. 115-127) aus Krippendorff, K. (2019). *Content analysis: an introduction to its methodology*. Sage.
- Lektüre: Kapitel 4 «Selecting Documents» (S. 41-47) aus Grimmer et al. (2022). *Text as data: a new framework for machine learning and the social sciences*. Princeton University Press.
- Skript: Bestehende Textkorpora (1.1.), Textdatenbanken (1.2)

Sitzung

- Diskussion und Gruppenprojekt/Übung: Sampling, Textdatenbanken

13.03.2023 *Ausfall*

20.03.2023 #4 Texte beschaffen 2

Vorbereitung

- Gruppenprojekt: Forschungsfrage formulieren
- Lektüre: Freelon, D. (2018). Computational research in the post-API age. *Political Communication*, 35(4), 665-668.

<https://doi.org/10.1080/10584609.2018.1477506>

- Skript: APIs (1.3.)

Sitzung

- Diskussion und Gruppenprojekt/Übung: APIs

27.03.2023 #5 Text beschaffen 3

Vorbereitung

- Skript: Web scraping (1.4.)

Sitzung

- Diskussion und Gruppenprojekt/Übung: Web scraping

03.04.2023 #6 Texte aufbereiten und bereinigen

Vorbereitung

- Lektüre: Kapitel 5 «Bag of Words» (S. 48-59) aus Grimmer et al. (2022). *Text as data: a new framework for machine learning and the social sciences*. Princeton University Press.
- Skript: Texte aufbereiten und bereinigen (2.)

Sitzung

- Diskussion und Gruppenprojekt/Übung: Texte aufbereiten und bereinigen

10.04.2023 Ausfall (Ostermontag)

17.04.2023 #7 Statistische Analysen

Vorbereitung

- Gruppenprojekt: Textkorpus finalisieren
- Skript: Statistische Analysen (3.1.)

Sitzung

- Diskussion und Gruppenprojekt/Übung: Statistische Analysen, Versionsverwaltung

24.04.2023 #8 Informationsextraktion

Vorbereitung

- Lektüre: Kapitel 9 «Representations from Language Sequences» (S. 90-98) aus Grimmer et al. (2022). *Text as data: a new framework for machine learning and the social sciences*. Princeton University Press.
- Skript: Informationsextraktion (3.2.)

Sitzung

- Diskussion und Gruppenprojekt/Übung: Informationsextraktion, gute Grafiken

01.05.2023 #9 Textklassifizierung 1: Dictionaries

Vorbereitung

- Lektüre: Boumans, J. W., & Trilling, D. (2016). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital Journalism*, 4(1), 8–23.
<https://doi.org/10.1080/21670811.2015.1096598>

- Skript: Dictionaries (3.3.)

Sitzung

- Diskussion und Gruppenprojekt/Übung: Textklassifikation, Dictionaries

08.05.2023 #10 Textklassifizierung 2: Topic Models

Vorbereitung

- Lektüre: Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2-3), 93-118. <https://doi.org/10.1080/19312458.2018.1430754>

- Skript: Topic Models (3.4.)

Sitzung

- Diskussion und Gruppenprojekt/Übung: Topic Models

15.05.2023 #11 Textklassifizierung 3: Überwachte Klassifikation

Vorbereitung

- Lektüre: Kapitel 17-20 (S. 184-217) aus Grimmer et al. (2022). *Text as data: a new framework for machine learning and the social sciences*. Princeton University Press.
- Skript: Überwachte Klassifikation (3.5.)

Sitzung

- Diskussion und Gruppenprojekt/Übung: Überwachte Klassifikation

22.05.2023 12# Textklassifizierung 4: Transformer

Vorbereitung

- Lektüre: Kapitel 1 «Hello Transformers» bis zur Überschrift «The Hugging Face Ecosystem» aus Tunstall et al. (2022). *Natural Language Processing with Transformers, Revised Edition*. O'Reilly.
<https://www.oreilly.com/library/view/natural-language-processing/9781098136789/ch01.html>
- Natural Language Processing with Transformers, Revised Edition

- Skript: Transformer (3.6.)

Sitzung

- Diskussion und Gruppenprojekt/Übung: Transformer
- Rückblick und Abschluss Seminar

29.05.2023 Ausfall (Pfingstmontag)