

CfP: Text Mining in der Politikwissenschaft

Anwendung von Text Mining-Tools auf demokratiebezogene Fragestellungen im Kontext der BR Deutschland (1949–2012)

Das Projekt „ePol – Postdemokratie und Neoliberalismus“ untersucht die Entwicklung der Begründungsmuster von Politik im medialen Diskurs der vergangenen sechs Jahrzehnte. Dazu wird ein Korpus von ca. 3,5 Millionen Zeitungsartikeln mit Hilfe von Text Mining Verfahren untersucht. Methodisch steht dabei die enge Integration von qualitativen und quantitativen Textanalyseperspektiven im Vordergrund. Zur Methodenqualifikation des Einsatzes von Text Mining in der Politikwissenschaft laden wir mit diesem Call (Nachwuchs-)WissenschaftlerInnen ein, sich ihren demokratiebezogenen Fragestellungen mit den im Rahmen des ePol-Projekts entwickelten Verfahren zu nähern. Geplant ist eine Einführung in projekteigene Analyseoberfläche sowie die Publikation eines Sammelbandes mit den Ergebnissen aus den Analysen der TeilnehmerInnen.

Das Beispiel „ePol - Postdemokratie und Neoliberalismus“

Ökonomisierungsprozesse haben die Politik westlicher Demokratien in den letzten Jahren maßgeblich geprägt: Zahlreiche Studien zu Deregulierungsprozessen, beispielsweise der Finanzmärkte, zu Privatisierungen ehemals öffentlicher Aufgaben und zur Zunahme von Public Privat Partnerships haben deren Konsequenzen in verschiedensten Politikfeldern aufzeigen können. Zu wenig untersucht wurde bislang allerdings, welche Rolle Ökonomisierung im Entscheidungsfindungsprozess in der politischen Öffentlichkeit spielt — und dies, obwohl in repräsentativen Demokratien, in denen Politik gegenüber den Bürgern als begründungspflichtig gilt, anzunehmen ist, dass entsprechende Entwicklungen sich nicht nur in den policies, sondern auch in der öffentlichen Kommunikation niederschlagen.

In der Politischen Theorie werden die Konsequenzen einer neoliberalen Ökonomisierung unter dem Stichwort „Postdemokratie“ gegenwärtig intensiv diskutiert. Unser Forschungsprojekt hat es sich zum Ziel gesetzt, die von der Postdemokratiedebatte formulierte Diagnose einer zunehmenden Ökonomisierung der politischen Öffentlichkeit für die Bundesrepublik erstmals umfassend empirisch zu analysieren. Hierzu wird auf die zentrale These des Diskurses zurückgegriffen, wonach der Prozess der Postdemokratisierung durch die zunehmende Bedeutung der Leitideen des Neoliberalismus in der politischen Sphäre (d.h. eine Ökonomisierung des Politischen) forciert wird. Die Relevanzverschiebung dieser Leitideen für den Zeitraum von 1949 bis 2011 wird anhand einer Diskursanalyse untersucht, bei der neben qualitativen Ausprägungen vor allem auch quantitative Verteilungen und Veränderungen politischer Argumentationsmuster analysiert werden.

Forschungsdesign

In einem innovativen Forschungsdesign kommen zur Aufarbeitung des umfangreichen Quellenmaterials Verfahren des Text Mining zum Einsatz. Ziel des ePol-Projekts ist es einerseits, etablierte Verfahren und Tools wie Topic Modelle, Term Extraction sowie Sentiment- und Kookkurenzanalysen für die sozialwissenschaftliche Auswertung qualitativer Daten fruchtbar zu machen. Darüber hinaus werden für die spezifische Fragestellung neue (halb-)automatische Text Mining Verfahren entwickelt, mit dem Ziel, Veränderungen politischer Argumentationsmuster und Begründungszusammenhänge in großen, zeitindizierten Korpora zu identifizieren und der politikwissenschaftlichen Analyse zugänglich zu machen. Durch die intensive Kooperation von Politischer Theorie und Informatik verfolgt ePol zusätzlich zu den Erkenntnisinteressen der am Projekt beteiligten Einzeldisziplinen auch metadisziplinäre Ziele. Angestrebt wird ein entscheidender Beitrag zur Nutzbarmachung von „Big Data“ für die empirische Sozialforschung mit qualitativen Daten.

Datenbestand

Die oben genannten Verfahren werden auf einen Datenkorpus angewendet, der für den Zeitraum von 1949 bis 2012 insgesamt ca. 3,5 Millionen Zeitungsartikel enthält, die aus der Süddeutschen Zeitung, der tageszeitung, der Zeit und der Frankfurter Allgemeinen Zeitung stammen.

Zeitung	abgedeckter Zeitraum	Anzahl Ausgaben	Anzahl Artikel
Frankfurter Allgemeine Zeitung (FAZ)	1949 – 2011	Stichprobe	200.389
Süddeutsche Zeitung (SZ)	1992 – 2011	6027	1.505.714
die tageszeitung (taz)	1986 – 2012	7821	1.391.981
Die Zeit	1946 – 2012	3841	397.729
		Gesamt	3.495.822

Tab.1: Datenbestand.

Analysetool

Für das ePol-Projekt wurde eine webbasierte Benutzeroberfläche entwickelt, die sowohl quantitative (textstatistische) als auch qualitative Auswertungsschritte unterstützt. Dabei sind die Anwendungszwecke freilich nicht auf die Fragestellungen aus dem ePol-Projekt beschränkt. Die Weboberfläche wurde mit dem Ziel entwickelt, eine *einfach bedienbare* Anwendungsoberfläche für komplexe textstatische und algorithmische Auswertungsverfahren auf sehr großen Textkollektionen zu schaffen. Darüber hinaus integriert sie einzelne Text Mining Verfahren miteinander, um so die Verknüpfung isolierter Einzelergebnisse zu komplexen, multiperspektivischen Analysen zu ermöglichen. Zum Funktionsumfang gehören:

- Volltextsuche auf dem Gesamtdatenbestand
- Erstellen von Sub-Kollektionen aus dem Gesamtdatenbestand
- Frequenzanalysen
- Kookkurrenzanalysen
- Topic Modelle
- Keyword-Extraktion
- Manuelle Codierung einzelner Dokumente mit hierarchischen Categoriesystemen
- Automatische Text-Klassifikation auf Basis manueller Annotationen als Trainingsbeispiele
- Visualisierung von Analyseergebnissen

Im Rahmen eines Workshops werden den TeilnehmerInnen die algorithmischen Grundlagen sowie die Anwendung der Verfahren erläutert mit dem Ziel selbstständig Analysen auf den ePol-Daten durchzuführen.

Zielgruppe und Verfahren

Der Call richtet sich an ForscherInnen, die zu politikwissenschaftlichen und dabei insbesondere demokratiebezogenen Fragestellungen im Zusammenhang mit für die BR Deutschland relevanten Themen arbeiten. Zudem müssen die Fragestellungen so beschaffen sein, dass eine systematische Auswertung der im Datenbestand von ePol vorhandenen Texte einen signifikanten Erkenntnisgewinn erwarten lässt. Dies kann unter anderem durch die Einbeziehung einer zeitlichen Entwicklungsperspektive in die Analyse geschehen. Statistikgrundkenntnisse sowie Vorkenntnisse im Bereich korpuslinguistischer Analysen sind hilfreich, bilden jedoch keine zwingende Voraussetzung zur Teilnahme.

Positiv evaluierte Interessenbekundungen werden zu einem 4-tägigen Workshop im August 2014 nach Hamburg eingeladen. Hierfür können einmalig die Hin- und Rückfahrtkosten (Bahnfahrt, 2. Klasse) sowie die Verpflegungskosten während des Workshops übernommen werden. Übernachtungen können auf Anfrage im Gästehaus der Universität organisiert werden, die Kosten hierfür können anteilig übernommen werden. Ziel des Workshops ist es, die jeweiligen Projektideen zu diskutieren sowie in die Arbeitstechniken mit der Analyseoberfläche einzuführen.¹ Die TeilnehmerInnen sollen so befähigt werden, in der Folge ihre Fragestellungen eigenständig zu bearbeiten und die Ergebnisse der Arbeit bis zum 31.10.2014 in einem Artikel schriftlich darzustellen. Die Artikel werden nach einem nochmaligen Reviewverfahren in einem Sammelband publiziert.

InteressentInnen reichen bis zum 31.03.2014 unter **matthias.lemke@hsu-hh.de** eine Interessenbekundung im Volumen von maximal 2 DIN-A4 Seiten ein, aus der folgende Informationen hervorgehen:

¹ **Rechtlicher Hinweis:** Mit der Teilnahme am Workshop erklären sich die BewerberInnen dazu bereit, als Projektpartner des Verbundprojekts „Postdemokratie und Neoliberalismus“ zu fungieren. Das Projekt „Postdemokratie und Neoliberalismus“ hat für seine Analysen die Nutzung urheberrechtlich geschützter Daten durch Zeitungsverlage lizenzieren lassen. Die Nutzung der Daten ist ausschließlich im Rahmen dieses Forschungsprojekts gestattet. Eine Weitergabe an Dritte für Zwecke außerhalb des Projekts, insbesondere von gebündelten Datenbeständen findet nicht statt. Der Zugriff auf die Daten und die Auswertung findet ausschließlich über eine Weboberfläche statt, die eine textstatische Auswertung sowie qualitative Begutachtung von Belegstellen in den Daten erlaubt.

- **Fragestellung/Themengebiet:** Kurze Erläuterung der Fragestellung/des Themengebiets inklusive der Angabe, ob es sich um die Bearbeitung im Rahmen einer Qualifikationsarbeit (BA-, MA-Arbeit, Dissertation, Habilitation) handelt;
- **Anschlussfähigkeit Text Mining:** Erläuterung, inwiefern die Bearbeitung der Fragestellung unter Zuhilfenahme von Text Mining Verfahren auf Basis des oben vorgestellten Korpus erfolversprechend ist;
- **Vorerfahrungen:** Angaben darüber, inwieweit bereits Vorkenntnisse im Umgang mit Text Mining bzw. lexikometrischen oder korpuslinguistischen Verfahren bestehen, und wenn ja, welche;
- **Zur Person:** Kontaktdaten und aktuelle Position.

Zeitplan

10.02.2014	Veröffentlichung des Calls
31.03.2014	Deadline für Einreichung der Interessenbekundung
30.04.2014	Bekanntgabe der Ergebnisse der Auswahl
01.-04.9.2014	Workshop in Hamburg
30.11.2014	Abgabe der Artikel
15.12.2014	Rückmeldung der Reviewergebnisse an AutorInnen und Überarbeitung
Januar 2014	Finalisierung der Publikation (Publikation im VS-Verlag, Reihe <i>Kritische Studien zur Demokratie</i>)

Rückfragen/Kontakt

Rückfragen im Zusammenhang mit diesem Call for Papers richten Sie bitte an:

Dr. Matthias Lemke

Helmut-Schmidt-Universität / Universität der Bundeswehr Hamburg

Institut für Politikwissenschaft

(Tel) +49/(0)40/6541-3379 – (Mail) matthias.lemke@hsu-hh.de