BA/MA/Doktoranden-Seminar

Datenerhebung im World Wide Web
Techniken zur Gewinnung politischer und Social-Media-Daten

Donnerstag, 8.15 - 9.45/11.30 Uhr, Y326

Universität Konstanz
Fachbereich Politik- und Verwaltungswissenschaft
Sommersemester 2015

Simon Munzert
simon.munzert@uni-konstanz.de
Raum D307

Kursbeschreibung

Zur Bearbeitung einer politikwissenschaftlichen Fragestellung liegen manchmal keine adäquaten Daten vor. Neben klassischen, oft sehr aufwendigen Formen der Datenerhebung existieren eine Reihe weiterer Möglichkeiten mit etwas technischem Geschick, dafür nahezu umsonst, eigene Daten zu erheben. Das Internet bietet eine Fülle an Möglichkeiten für gewissermaßen nicht-reaktive Messungen von Verhalten politischer und anderer Akteure (z.B. Abgeordnete, Gerichte, Medien). Politische Texte (Parteiprogramme, parlamentarische Reden etc.) und Informationen aus Medien und neueren sozialen Medien können mithilfe computerbasierter Tools genutzt werden, um Daten über die Positionierung von Akteuren zu gewinnen. Ziel des Kurses ist zum einen die Vermittlung technischer Grundlagen web-basierter Datenerhebungsmethoden. Zum anderen sollen die Kursteilnehmerinnen und Kursteilnehmer selbständig Daten erheben, anhand derer sich aktuelle politische und politikwissenschaftliche Fragestellungen beantworten lassen.

Voraussetzungen

Vorkenntnisse des Statistikprogramms R sind von Vorteil, aber nicht Bedingung. Eine Fülle an hilfreichen Einführungen findet sich im Netz, z.B. hier: https://www.datacamp.com/swirl-r-tutorial.

Leistungsnachweise

Hausaufgaben, kleine Präsentation eines Anwendungs-Papiers, Seminararbeit (10-15 Seiten).

Literatur und Daten

Die Kursliteratur und verwendete Daten sowie Programm-Codes werden in ILIAS bereitgestellt. Mit Sternchen markierte Texte (*) sind als weiterführende Lektüre empfohlen. Des Weiteren liegt dem Kurs das Manuskript folgenden Buches zugrunde (im Folgenden **ADCR** abgekürzt):

Munzert, Simon, *Christian Rubba*, *Peter Meißner*, und *Dominic Nyhuis*, 2015: Automated Data Collection with R. A Practical Guide to Web Scraping and Text Mining. Chichester: John Wiley & Sons.

Kursplan

Einführung

16.04.: Organisatorisches und Überblick

VanderPlas, Jake, 2013: The Big Data Brain Drain: Why Science is in Trouble. https://jakevdp.github.io/blog/2013/10/26/big-data-brain-drain/.

23.04.: Was tun mit web-basierten Daten?

Mellon, Jonathan, 2014: Internet Search Data and Issue Salience: The Properties of Google Trends as a Measure of Issue Salience. Journal of Elections, Public Opinion and Parties 24:45–72.

30.04.: Workshop '*Einführung in R*' (ganztägig)

Interaktives R-Tutorial: https://www.datacamp.com/swirl-r-tutorial

Datenerhebung im World Wide Web

07.05.: Web Scraping mit regulären Ausdrücken, 4h

ADCR. Kapitel 'Regular Expressions and String Functions'.

* ADCR. Kapitel 'Parsing information from semistructured documents'.

14.05.: KEIN KURS

21.05.: Web Scraping mit XPath, 4h

ADCR. Kapitel 'HTML', 'XML and JSON' (ohne JSON-Abschnitt), 'XPath'.

* ADCR. Kapitel 'Mapping the Geographic Distribution of Names'.

28.05., 04.06., 11.06.: KEIN KURS

18.06.: Fortgeschrittene Scrapinganwendungen, 4h

ADCR. Kapitel 'Scraping the Web'.

25.06.: JSON und APIs, 4h

ADCR. Kapitel 'XML and JSON' (nur JSON-Abschnitt) und 'Scraping the Web', Abschnitte 9.1.10 und 9.1.11.

* ADCR. Kapitel 'Predicting the 2014 Academy Awards using Twitter'.

Methoden der quantitativen Textanalyse

02.07.: Verarbeitung von Textdaten, 4h

ADCR. Kapitel 'R tools for statistical text processing', Abschnitte 10.1 und 10.2.

Grimmer, Justin, und *Brandon M. Stewart*, 2013: Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis 21:267–297.

09.07.: Extraktion latenter Information aus politischen Texten, 4h

ADCR. Kapitel 'R tools for statistical text processing', Abschnitte 10.3 und 10.4.

Laver, Michael, *Kenneth Benoit*, und *John Garry*, 2003: Extracting Policy Positions from Political Texts Using Words as Data. The American Political Science Review 97:311–331.

16.07.: KEIN KURS

Präsentationspapiere

Barberá, Pablo, 2015: Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data. Political Analysis 23:76–91.

Barberá, Pablo, 2014: How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S. Unpublished Manuscript.

Barberá, Pablo, und *Gonzalo Rivero*, 2014: Understanding the Political Representativeness of Twitter users. Social Science Computer Review 1–18.

Chadefaux, Thomas, 2014: Early warning signals for war in the news. Journal of Peace Research 51:5–18.

Enos, Ryan D., und *Anthony Fowler*, 2014: The Effects of Large-Scale Campaigns on Voter Turnout: Evidence from 400 Million Voter Contacts. Unpublished Manuscript.

Gayo-Avello, Daniel, 2013: A Meta-Analysis of State-of-the-Art Electoral Prediction From Twitter Data. Social Science Computer Review 31:649–679.

Gill, Michael, und *Arthur Spirling*, 2015: Estimating the Severity of WikiLeaks U.S. Diplomatic Cables Disclosure. Political Analysis 23:299–305.

Gohdes, Anita R., 2015: Pulling the plug: Network disruptions and violence in civil conflict. Journal of Peace Research 52:1–16.

Hassanpour, Navid, 2013: Tracking the Semantics of Politics: A Case for Online Data Research in Political Science. PS: Political Science & Politics 46:299–306.

King, Gary, *Jennifer Pan*, und *Margaret E. Roberts*, 2013: How Censorship in China Allows Government Criticism but Silences Collective Expression. American Political Science Review 107:326–343.

Rød, Espen Geelmuyden, und *Nils B. Weidmann*, 2015: Empowering activists or autocrats? The Internet in authoritarian regimes. Journal of Peace Research 52:1–14.

Sagi, Eyal, und *Morteza Dehghani*, 2014: Measuring Moral Rhetoric in Text. Social Science Computer Review 32:132–144.

Shaw, Aaron, und *Bejamin Mako Hill*, 2014: Laboratories of Oligarchy? How The Iron Law Extends to Peer Production. Journal of Communication 64:215–238.

Slapin, Jonathan B., und *Sven-Oliver Proksch*, 2008: A Scaling Model for Estimating Time-Series Party Positions from Texts. American Journal of Political Science 52:705–722.

Street, Alex, *Thomas A. Murray*, *John Blitzer*, und *Rajan S. Patel*, 2015: Estimating Voter Registration Deadline Effects with Web Search Data. Political Analysis 1–2.

Wu, Shaomei, *Jake M. Hofman*, *Winter A. Mason*, und *Duncan J. Watts*, 2011: Who Says What to Whom on Twitter. In: Proceedings of the 20th International Conference on World Wide Web, WWW '11, 705–714. New York, NY, USA: ACM.

Zeitzoff, Thomas, 2011: Using Social Media to Measure Conflict Dynamics: An Application to the 2008 - 2009 Gaza Conflict. Journal of Conflict Resolution 55:938–969.

Zeitzoff, Thomas, *John Kelly*, und *Gilad Lotan*, 2015: Using social media to measure foreign policy dynamics: An empirical analysis of the Iranian–Israeli confrontation (2012-13). Journal of Peace Research 52:1–16.