

BA/MA/Doktoranden-Seminar

Datenerhebung im World Wide Web

Techniken zur Gewinnung politischer und Social-Media-Daten

Dienstag, 11.45 - 13.15 Uhr, Y326

Universität Konstanz
Fachbereich Politik- und Verwaltungswissenschaft
Sommersemester 2014

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: <http://tryr.codeschool.com/>.

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

22. April 2014: Organisatorisches und Überblick

29. April 2014: Was tun mit web-basierten Daten?

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

Barberà, Pablo, 2014: Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data. Political Analysis XX:1–14.

* *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.*

06. Mai 2014: Theoretische Überlegungen zur Datenauswahl

Hox, Joop J., und Hennie R. Boeje, 2005: Encyclopedia of Social Measurement, Kapitel Data Collection, Primary vs. Secondary. San Diego, CA: Academic Press.

Greenstein, Shane, und Feng Zhu, 2012: Is Wikipedia Biased? American Economic Review 102:343–348.

* *Brown, Adam R., 2011: Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage. Political Science & Politics 44:339–343.*

13. Mai 2014: Workshop ‘Einführung in R’ (ganztägig)

Code School: <http://tryr.codeschool.com/>

ADCR. Kapitel ‘Regular Expressions and String Functions.’

Datenerhebung im World Wide Web

20. Mai 2014: Workshop ‘Technologien des World Wide Web’ (halbtägig)

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

27. Mai 2014: Workshop ‘Web Scraping mit R’ (halbtägig)

ADCR. Kapitel ‘HTTP’, ‘Scraping the Web’.

* ADCR. Kapitel ‘Collaboration Networks in the U.S. Senate’.

* ADCR. Kapitel ‘Mapping the Geographic Distribution of Names’.

03. Juni 2014: Workshop 'JSON und APIs' (halbtägig)

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

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

10. Juni 2014: Übung zu Datengewinnung mit web-basierten Tools

Methoden der quantitativen Textanalyse

17. Juni 2014: Extraktion manifester Daten aus politischen Texten

ADCR. Kapitel 'R tools for statistical text processing'.

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.

24. Juni 2014: Extraktion latenter Information aus politischen Texten

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.

* *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.*

02. Juli 2014: Medien und soziale Medien als Datenquelle

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.

Lassen, David S., und Adam R. Brown, 2011: Twitter: The Electoral Connection? Social Science Computer Review 29:419–436.

* *Tumasjan, Andranik, Timm O. Sprenger, Philipp G. Sandner, und Isabell M. Welp, 2011: Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape. Social Science Computer Review 29:402–418.*

09. Juli 2014: Übung zu Datengewinnung aus politischen Texten und sozialen Medien

16. Juli 2014: Abschlussbesprechung

15. September 2014: Abgabe der Seminararbeit