

BA/MA/Doktoranden-Seminar

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

Dienstag, 11.45 - 13.15 Uhr, Y326

Sommersemester 2014

Universität Konstanz

Fachbereich Politik- und Verwaltungswissenschaft

Simon Munzert

simon.munzert@uni-konstanz.de

Raum D307 (Sprechstunde nach Vereinbarung)

Kursbeschreibung

Zur Bearbeitung einer politikwissenschaftlichen Fragestellung liegen manchmal keine adäquaten Daten vor. Neben klassischen, oft sehr aufwendigen Formen der Datenerhebung (z.B. standardisierte oder nicht-standardisierte Befragung, teilnehmende oder nicht-teilnehmende Beobachtung) 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 Einführungen findet sich im Netz, z.B. hier: <http://tryr.codeschool.com/>.

Leistungsnachweise

Regelmäßige Teilnahme, Hausaufgaben, drei Blog-Artikel mit eigener Datensammlung und Auswertung zu aktueller politischer Fragestellung. Abgabe 1. Blogartikel: 24. Juni 2014, Abgabe 2. Blogartikel: 29. Juli 2014, Abgabe 3. Blogartikel: 15. September 2014

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, 2014: Automated Data Collection with R. A Practical Guide to Web Scraping and Text Mining for the Social Sciences. Hoboken, NJ: John Wiley & Sons. Manuskript.

Vorläufiger Kursplan

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 XX:XX.

Barberà, Pablo, 2013: Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data.

* *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.*

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

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

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

ADCR. Kapitel ‘Regular Expressions and String Functions’.

* *Adler, Joseph, 2010: R in a Nutshell. Sebastopol, CA: O’Reilly Media, 1. Auflage. Deutsche Übersetzung von Jörg Beyer.*

* ADCR. Kapitel ‘Parsing Information from Semi-Structured Documents’.

13. Mai 2014: Workshop ‘Technologien des World Wide Web’ (ganztägig)

ADCR. Kapitel ‘HTML’, ‘XML and JSON’, ‘XPath’ und ‘HTTP’.

20. Mai 2014: Workshop “Web Scraping mit R” (ganztägig)

ADCR. Kapitel ‘Scraping the Web’.

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

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

* ADCR. Kapitel ‘Predicting the 2014 Academy Awards using Twitter’.

27. Mai 2014: JSON und APIs

ADCR. Kapitel ‘XML and JSON’ und ‘Scraping the Web’.

03. Juni 2014: Übungssitzung

10. Juni 2014: Scraping von AJAX-basierten Webseiten

ADCR. Kapitel 'AJAX' und 'Scraping the Web, Abschnitt 9.1.9'.

17. Juni 2014: Wie verfasse ich einen datenbasierten journalistischen Text?

Nyhan, Brendan, und John Sides, 2011: How Political Science Can Help Journalism (and Still Let Journalists Be Journalists). The Forum 9:Article 2.

Stray, Jonathan, 2012: Data, uncertainty, and specialization: What journalism can learn from FiveThirtyEight's election coverage. <http://www.niemanlab.org/2012/10/data-uncertainty-and-specialization-what-journalism-can-learn-from-fivethirtyeight-election-coverage/>

17. Juni 2014: Besprechung zur Verfassung der Blogartikel

24. Juni 2014: Datenvisualisierung mit R

Gelman, Andrew, und Antony Unwin, 2013: Infovis and statistical graphics: different goals, different looks. Journal of Computational and Graphical Statistics 22:2–28.

Wickham, Hadley, 2010: A Layered Grammar of Graphics. Journal of Computational and Graphical Statistics 19:3–28.

Alcocer, Pedro, 2010: Advance plotting in R with ggplot2. <http://www.pealco.net/downloads/ws2010-ggplot2.pdf>

01. Juli 2014: Abgabe 1. Blogartikel

01. Juli 2014: Scraping Workflow und Projekt-Management

ADCR. Kapitel 'Managing Data Projects'.

08. - 22. Juli 2014: Projektarbeit

29. Juli 2014: Abgabe 2. Blogartikel

15. September 2014: Abgabe 3. Blogartikel