**GESIS Workshop**

Grunddaten-Blatt  
**zur Ankündigung und Planung sowie zur Vertragsvorbereitung**

|  |  |
| --- | --- |
| **Basisinformationen** | |
| Name des Workshops (deutsch) | Automatic sampling and analysis of YouTube data |
| Name des Workshops (englisch) | Automatic sampling and analysis of YouTube data |
| Datum | 10. – 11.02.2020 |
| Voraussichtliche Dauer in Tagen | 2 |

|  |  |
| --- | --- |
| Ort: | ☐Mannheim ☒Köln |

|  |  |
| --- | --- |
| Sprache | ☐ Deutsch ☒ Englisch |
| Hotelreservierung | ☐ Nein ☒ Ja, von 09.02.2020 bis 11.02.2020 (für eine Person) |
| Themenbereiche | ☐Forschungsansätze und –designs  ☐Daten- und Informationsrecherche  ☐Studienplanung  ☒Datenerhebung und –verarbeitung  ☒Datenaufbereitung und –analyse  ☐Datenregistierung und –archivierung, Datenmanagement |
| Methode | ☒Quantitativ ☐Qualitativ ☐Mixed |
| Zielgruppe/Niveau (möglichst wenig restriktiv): | ☒ Studierende  ☒ Master-Studierende  ☒ Promovierende  ☒ Post-Docs  ☒ ProfessorInnen  ☒ WissenschaftlerInnen  ☒ Interessierte Nicht-WissenschaftlerInnen  ☐ Interessierte Allgemein |
| Vorkenntnisse | ☒ Einsteiger/Anfänger  ☒ Fortgeschrittene  ☐ Experten  ☐ keine |

|  |  |  |
| --- | --- | --- |
| **Information zu Dozentinnen und Dozenten:** | | |
| DozentIn 1 (HauptansprechpartnerIn für GESIS) | Name: Julian Kohne | |
| E-Mail: julian.kohne@gesis.org | |
| Telefon: 0221 47694 – 222 | |
| Adresse 1: GESIS – Leibniz-Institut für Sozialwissenschaften, Unter Sachsenhausen 6-8, 50667 Köln | |
| Adresse 2: (für Vertrag, falls abweichend von Adresse 1) | |
| Bei Workshop anwesend:  ☒ über die ganze Dauer  ☐ nur teilweise | Kurzbiografie (bis zu 100 Wörter):  Julian Kohne studied Psychology and Social & Organizational Psychology at the University of Groningen. He joined the Computational Social Science Department at GESIS in 2017 where he is coordinating the developments of GESIS in the area of digital behavioral data. His research interests concern using digital data for quantifying human behavior, especially interpersonal relationships through text mining and the simulation of social systems. | |
| DozentIn 2  trifft nicht zu | Name: Johannes Breuer | |
| E-mail: johannes.breuer@gesis.org | |
| Telefon: 0221 47694-471 |  |
|  | |
|  | |
| Adresse 1: GESIS – Leibniz-Institut für Sozialwissenschaften, Unter Sachsenhausen 6-8, 50667 Köln | |
|  | Adresse 2: (für Vertrag, falls abweichend von Adresse 1) | |
| Bei Workshop anwesend:  ☒ über die ganze Dauer  ☐ nur teilweise | Kurzbiografie (bis zu 100 Wörter):  Johannes Breuer works as a senior researcher in the team Data Linking & Data Security at the GESIS Data Archive. He received his Ph.D. in psychology from the University of Cologne in 2013. Before joining GESIS, he worked in several research projects investigating the use and effects of digital media at the universities of Cologne, Hohenheim, and Münster, and the Leibniz-Institute für Wissensmedien (Knowledge Media Research Center). His other research interests include the methods of media (effects) research, data management, and open science. | |
| DozentIn 3  trifft nicht zu | Name: M. Rohangis Mohseni | |
|  | E-mail: rohangis.mohseni@tu-ilmenau.de | |
|  | Telefon: 03677 69-4739 | |
|  |  | |
|  |  | |
|  | Adresse 1:  Technische Universität Ilmenau  IfMK (Institut für Medien und Kommunikationswissenschaft)  Ernst Abbe Zentrum für Forschung und Transfer (EAZ)  Ehrenbergstr. 29  D-98693 Ilmenau | |
|  | Adresse 2: (für Vertrag, falls abweichend von Adresse 1) | |
| Bei Workshop anwesend:  ☒ über die ganze Dauer  ☐ nur teilweise | Kurzbiografie (bis zu 100 Wörter):  M. Rohangis Mohseni is a postdoctoral researcher of the Media Psychology and Media Design research group at TU Ilmenau in Germany. He received his Ph.D. in psychology from the University Osnabrück in 2013 and is currently working on a habilitation on the topic of sexist online hate speech. His research interests include electronic media effects and moral behavior. | |
| Ggf. Kooperationspartner |  | |

|  |
| --- |
| **Workshop Inhalt (für die Ankündigung):** |
| **Beschreibung der Inhalte:** *(150-250 Wörter; bitte nennen Sie auch explizit Software Pakete und/oder Datensätze/Studien, die im Workshop Anwendung finden, da dies eine für potentielle TeilnehmerInnen relevante Information ist)*  YouTube is the largest and most popular video platform on the internet. The producers and users of YouTube content generate huge amounts of data. These data are also of interest to researchers (in the social sciences as well as other disciplines) for studying different aspects of online media use and communication. Accessing and working with these data, however, can be challenging. In this workshop, we will first discuss the potential of YouTube data for research in the social sciences, and then introduce participants to different tools and methods for sampling and analyzing data from YouTube. We will then demonstrate and compare several tools for collecting YouTube data. Our focus for the main part of the workshop will be on using the tuber package for R to collect data via the YouTube API and wrangling and analyzing the data in R (using various packages). Regarding the type of data, we will focus on user comments but also will also (briefly) look into other YouTube data, such as video statistics and subtitles. For the comments, we will show how to clean/process them in R, how to deal with emojis, and how to do some basic forms of automated text analysis (e.g., word frequencies, sentiment analysis). While we believe that YouTube data has great potential for research in the social sciences (and other disciplines), we will also discuss the unique challenges and limitations of using this data. |
| **Keywords / max. 5 mit Komma getrennt (deutsch):** YouTube, R, computational methods, digital trace data |
| **Keywords / max. 5 mit Komma getrennt (englisch):** YouTube, R, computational methods, digital trace data |
| **Zielgruppe:** *Für wen ist die Veranstaltung primär gedacht, d.h. welche Interessen sollten bei Teilnehmenden vorliegen?*  The workshop is aimed at people who are interested in using YouTube data for their research. |
| **Lernziele:** *Was können die Teilnehmenden erwarten am Ende der Veranstaltung zu wissen bzw. zu können? Bitte möglichst konkret in 2-3 Sätzen.*  Participants will learn how they can use YouTube data for their research. They will get to know several tools for collecting YouTube data and learn about their advantages and disadvantages. At the end of the workshop, participants should be able to automatically collect YouTube data, process/clean it, and do some basic (exploratory) analyses of user comments. |
| **Voraussetzungen:** *Was sollen die TeilnehmerInnen vorher bereits wissen bzw. können, um diese Lernziele auch erreichen zu können? Bitte möglichst konkret in 2-3 Sätzen*.  Participants should at least have some basic knowledge of R and, ideally, also the tidyverse. Basic R knowledge can, for example, be acquired through the swirl (Learn R, in R) course “R Programming” (see <https://swirlstats.com/>), the DataCamp online course “Introduction to R” (<https://www.datacamp.com/courses/free-introduction-to-r>) or the RStudio Primer “Programming basics“ (<https://rstudio.cloud/learn/primers/1.2>), all of which are available for free. There also are many free online introductions to the tidyverse: For example, this blog post by Martin Frigaard (<http://www.storybench.org/getting-started-with-tidyverse-in-r/>) or this webinar by Thomas Mock (<https://resources.rstudio.com/webinars/a-gentle-introduction-to-tidy-statistics-in-r>). |
| **Literatur:**  Thelwall, M. (2017). Social media analytics for YouTube comments: Potential and limitations. *International Journal of Social Research Methodology*, *21*(3), 303–316. <https://doi.org/10.1080/13645579.2017.1381821> |

|  |
| --- |
| **Agenda (mindestens Anfangs- und Endzeiten!)** |

**NB: This is a first draft and likely to change**

|  |  |
| --- | --- |
| **Day 1** | |
| 10:00 – 11:00 | Introduction: Why is YouTube data interesting for research? |
| 11:00 – 11:30 | Coffee Break |
| 11:30 – 12:30 | The YouTube API |
| 12:30 – 13:30 | Lunch |
| 13:30 – 14:30 | Tools for the automatic sampling of YouTube data |
| 14:30 – 15:30 | Collecting data with the tuber package for R |
| 15:30 – 16:00 | Coffee break |
| 16:00 – 17:30 | Processing and cleaning user comments (in R) |
| 18:00 | Get together |
| **Day 2** | |
| 09:00 – 10:30 | Basic text analysis of user comments |
| 10:30 – 11:00 | Coffee Break |
| 11:00 – 12:00 | Sentiment analysis of user comments |
| 12:00 – 13:00 | Lunch |
| 13:00 – 14:00 | Excursus: Retrieving video subtitles |
| 14:00 – 14:30 | Coffee break |
| 14:30 – 16:00 | Practice session, questions, and outlook |

|  |  |  |
| --- | --- | --- |
| **Technische Ausstattung:** |  | |
| PC/Laptop für DozentInnen? | ☒ja  ☐nein  ☐noch nicht bekannt | |
| PCs/Laptops für TeilnehmerInnen? | ☒ja  ☐nein  ☐noch nicht bekannt | |
| Software Anforderungen: | …für DozentInnnen PC/Laptop: | …für TeilnehmerInnen PCs/Laptops: |
| *(Standardausstattung: Whiteboard, Beamer, MS Office, pdf-Reader, Internet Browser)* | ☐keine  ☒R mit folgenden Packages: devtools, tuber, tidyverse, tm, quanteda, qdapRegex, syuzhet, lexicon, subtools, stm  ☐SPSS  ☐Stata mit folgenden ados:  ☐Unipark  ☐Lisrel  ☐Mplus  ☐MaxQDA  Andere: | ☐keine  ☒R mit folgenden Packages: devtools, tuber, tidyverse, tm, quanteda, qdapRegex, syuzhet, lexicon, subtools, stm  ☐SPSS  ☐Stata mit folgenden ados:  ☐Unipark  ☐Lisrel  ☐Mplus  ☐MaxQDA  Andere: |
| Weitere Ausstattung gewünscht: |  | |
| Internetzugang (WLAN) | ☐nein  ☐ja, für Lehrkraft  ☒ja, für Lehrkraft und Teilnehmende |  |
| Besteht Interesse, Materialien (z.Bsp Datensätze, Skripte) über ein geschütztes Online-Portal oder gemeinsames Laufwerk an die WorkshopteilnehmerInnen zu verteilen? | | ☒ja  ☐nein  ☐noch nicht bekannt |