Digitale Verhaltensdaten und Webtracking

Methoden der empirischen Kommunikations- und Medienforschung

Marko Bachl
Freie Universität Berlin

Fragen zur Übung?

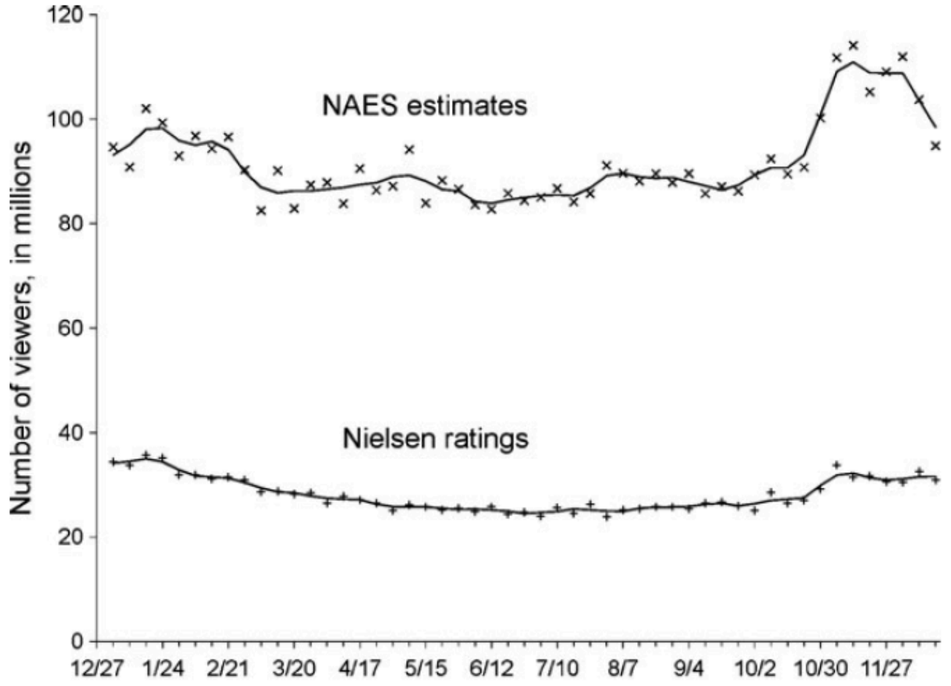
AGENDA

1. Probleme mit Selbstauskünften
2. Digitale Verhaltensdaten
3. Webtracking
4. Beispielstudie: Gesundheitsbezogene Internetsuche
5. Fazit

Probleme mit Selbstauskünften

THE IMMENSELY INFLATED NEWS AUDIENCE

- **Methode**: Aggregatsvergleich von Befragungsdaten zur TV-Nachrichtennutzung mit Nielsen TV-Meter-Daten
- **Ergebnis**: “Survey-based estimates of network news viewing were on average three times as high as Nielsen estimates in 2000 and up to eight times as high in some demographic subgroups” (S. 137)



(Prior, 2009)

ACCURACY OF SELF-REPORTED INTERNET USE

- **Methode**: Individualvergleich von Befragungsdaten zur Internetnutzung und browser-basierten Tracking-Daten
- **Ergebnis**: “Self-report measures of Internet use are rarely accurate and their convergent validity with client log files is rather weak” (S. 22)
  - Weniger Fehler für spezifische Fragen
  - Overreporting häufiger; Systematische Prädiktoren von Over- und Underreporting

Table 1. Percent differences in self-report (SR) and log data of Internet use.

Hours per week				Days per month			
	Self-report	Logs	Difference		Self-report	Logs	Difference
(a) General Internet use							
> 15 hrs	28	16	12	> daily	54	15	39
10–15 hrs	19	14	5	about daily	25	28	−3
5–10 hrs	21	32	−11	at least weekly	19	41	−22
2–5 hrs	21	31	−10	at least monthly	2	11	−9
< 2 hrs	10	8	2	once a month or less	1	4	−3
(b) Specific Internet use							
Closed social networks				Open social networks			
at least weekly	29	31	−2	at least weekly	4	1	3
at least monthly	11	16	−5	at least monthly	3	1	2
once a month or less	17	15	2	once a month or less	13	4	9
never	43	38	5	never	79	95	−16
Video platforms				Online auctions			
at least weekly	7	21	−14	at least weekly	12	8	4
at least monthly	11	16	−5	at least monthly	14	13	1
once a month or less	48	18	30	once a month or less	32	20	12
never	34	44	−10	never	42	58	−16

Note. $n = 3401$

(Scharkow, 2016)

HOW ACCURATE ARE SURVEY RESPONSES ON SOCIAL MEDIA AND POLITICS?

- **Methode**: Individualvergleich von Befragungsdaten mit API-Daten von Twitter und Facebook
- **Ergebnis**: “The good news is that self-reports are correlated with observed behavior. […] However, at the individual level there are substantial discrepancies in reporting” (S. 14)

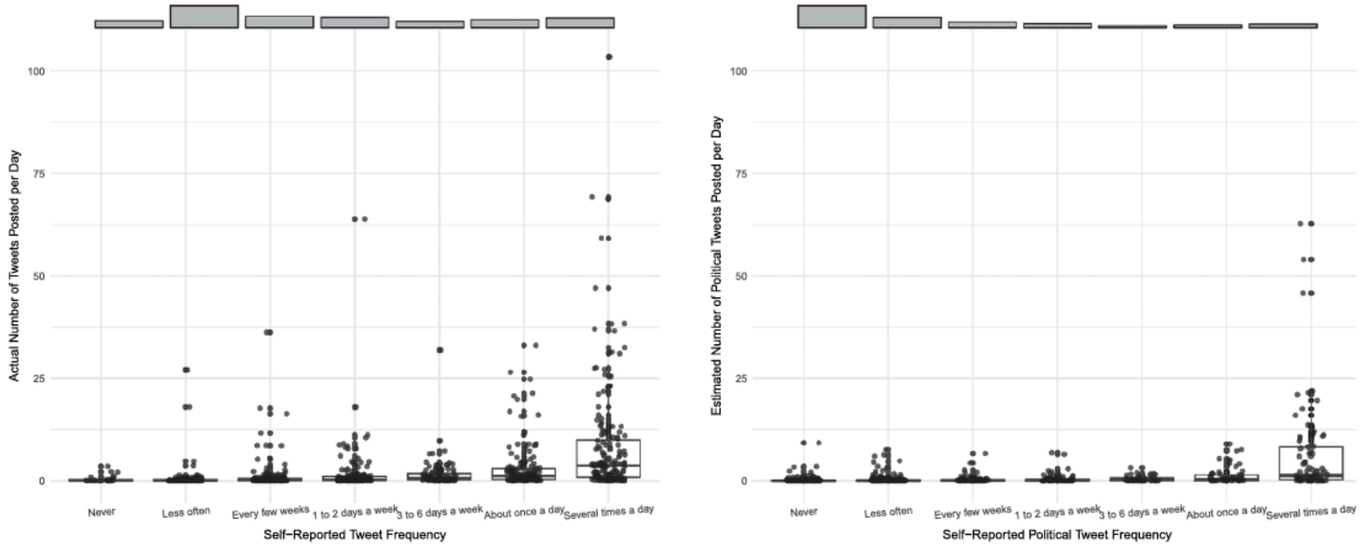


Figure 3. Number of total (left) or political (right) tweets per day posted (using linked data from respondents’ Twitter accounts) plotted against self-reported total or political tweet frequency. Tweets are categorized as “political” via the supervised learning technique discussed in Coding Social Media Posts.

(Guess et al., 2019)

PROBLEME MIT SELBSTAUSKÜNFTEN

- Menschen berichten ihr (Mediennutzungs-) Verhalten fehlerhaft
  - Unvollständige oder verzerrte Erinnerung
  - Verzerrte Angaben
  - Verständnis der Fragen

EXPOSURE TO POLITICAL CONTENT IN NEWSPAPERS

- **Methode**: Copy-Tests: Studienteilnehmende markieren in Zeitung die Artikel, die sie gelesen haben.
- **Aufwand**: ca. 75 Minuten persönliche Interviews

Exposure to newspaper content was measured by means of a copy test that was part of the interview. The interviewers presented the original newspaper issues and asked the respondents whether and to what extent they had read the articles[4] in the politics and news sections. After extensive pre-testing of several variations, exposure was measured by four categories: read whole article, read about half of the article, read only headline, and did not read the article at all. The average interview including the copy test lasted about 75 minutes.[5]

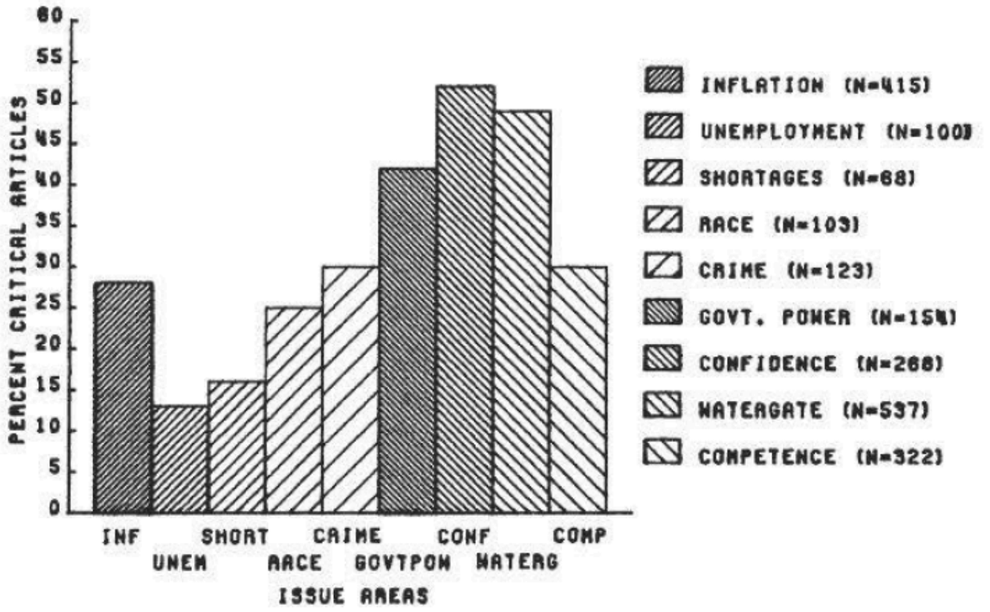
TABLE 1
Exposure to Articles in Consonant and Dissonant Cases

	Relationship between a Reader's Predisposition and the Direction of the Information (%)		
	Consonant (n=4978)	Neutral (n=10,463)	Dissonant (n=4282)
No exposure at all	43	44	44
Read at least headline	57	56	56
Read at least half of article	38	33	34*
Read whole article	24	19	20*

(Donsbach, 1991)

LINKAGE ANALYSIS

- **Problem**: "'Directional,' evaluative media research has been confounded by the frequent use of, at best, indirect measures of media impact which require large inferential leaps. It has suffered from the methodological problems inherent in a **failure to distinguish between reliance on a medium and exposure to a message, or between exposure in general and exposure to particular message content**" (S. 68)
- **Methode**: Linkage Analysis: Inhaltsanalyse der genutzten Tageszeitungen und Zuweisung der Inhalte im Befragungsdatensatz; ANES **1974**



Source: Center for Political Studies Media Content Analysis Study, 1974; available through the University of Michigan, ICPSR. Not to be cited without full bibliographical reference to the present article.

Figure 1. Degree of Criticism by Issue Area

(Miller et al., 1979)

PROBLEME MIT SELBSTAUSKÜNFTEN

- Menschen berichten ihr (Mediennutzungs-) Verhalten fehlerhaft
  - Unvollständige oder verzerrte Erinnerung
  - Verzerrte Angaben
  - Verständnis der Fragen
- Menschen können unmöglich alle genutzten Inhalte berichten
  - Erhebungsaufwand außerhalb kontrollierter Rezeptionsstudien
  - Nur grobe Annäherung bei Linkage Analysis
- Selbstauskünfte sind fehlerhafte, verzerrte und in vielerlei Hinsicht ungenaue Messungen von Mediennutzung und Rezeption von Inhalten
- Aber Selbstauskünfte bleiben eine wichtige Methode zur Erfassung von Mediennutzung (Bachl, 2025)

Fragen?

Digitale Verhaltensdaten

Basiert auf Wagner et al. (2025)

DEFINITION

> We call these research data “digital behavioral data” (DBD). DBD encompass digital observations of human or algorithmic behavior. DBD are generated (1) through interactions and content production online (e.g., on platforms such as Google, Facebook or websites on the World Wide Web) or (2) by software or sensors for recording specific processes(e.g., smartphones, RFID sensors, satellites, street view cameras, or web tracking).

(Wagner et al., 2025, S. 2)

KOMMUNIKATIONSWISSENSCHAFTLICHE BEISPIELE

- Beiträge und Interaktionen (z.B. “Likes”) auf Social-Media-Plattformen
- E-Mail- oder Chat-Kommunikation
- Verläufe der Nutzung von elektronischen Medien (z.B. Browser History, zuletzt gesehene Inhalte auf Plattformen)
- Cookies und Server Logs
- Standortverlauf von mobilen Geräten
- Aufzeichnung von Gerätinteraktionen (z.B. Bildschirmzeit, Mauszeiger und Klicks)
- Aufzeichnung von Blickverläufen (Eyetracking)

EIGENSCHAFTEN

- volume, voluminous, big
- veracity, unobtrusive, reliable, nonreactive
- velocity, always-on, real-time
- variety, unstructured, incomplete, drifting, dirty
- nonrepresentative, unstructured, multimodal
- algorithmically confounded

(Wagner et al., 2025, S. 3–4)

SYSTEMATISIERUNG

	Found data non-reactive, no control or interruption of data-generation process	**Designed data** researchers control and influence the data-generation process
Platform-centered data collection without user collaboration	(1) ▪ APIs ▪ web scraping ▪ platform cooperations ▪ buying platform data	(2) ▪ experiments on platforms (e.g., via platform cooperations)
User-centered data collection with user collaboration	(3) ▪ data donations via participant self-selection	(4) ▪ data donations via systematic recruitment of participants ▪ online experiments (e.g., via crowdsourcing platforms or artificial online environments) ▪ research software (e.g., web tracking or smartphone apps) ▪ integration into panel survey infrastructures

(Wagner et al., 2025, S. 6)

Fragen?

Webtracking

Basiert auf Mangold & Stier (2025)

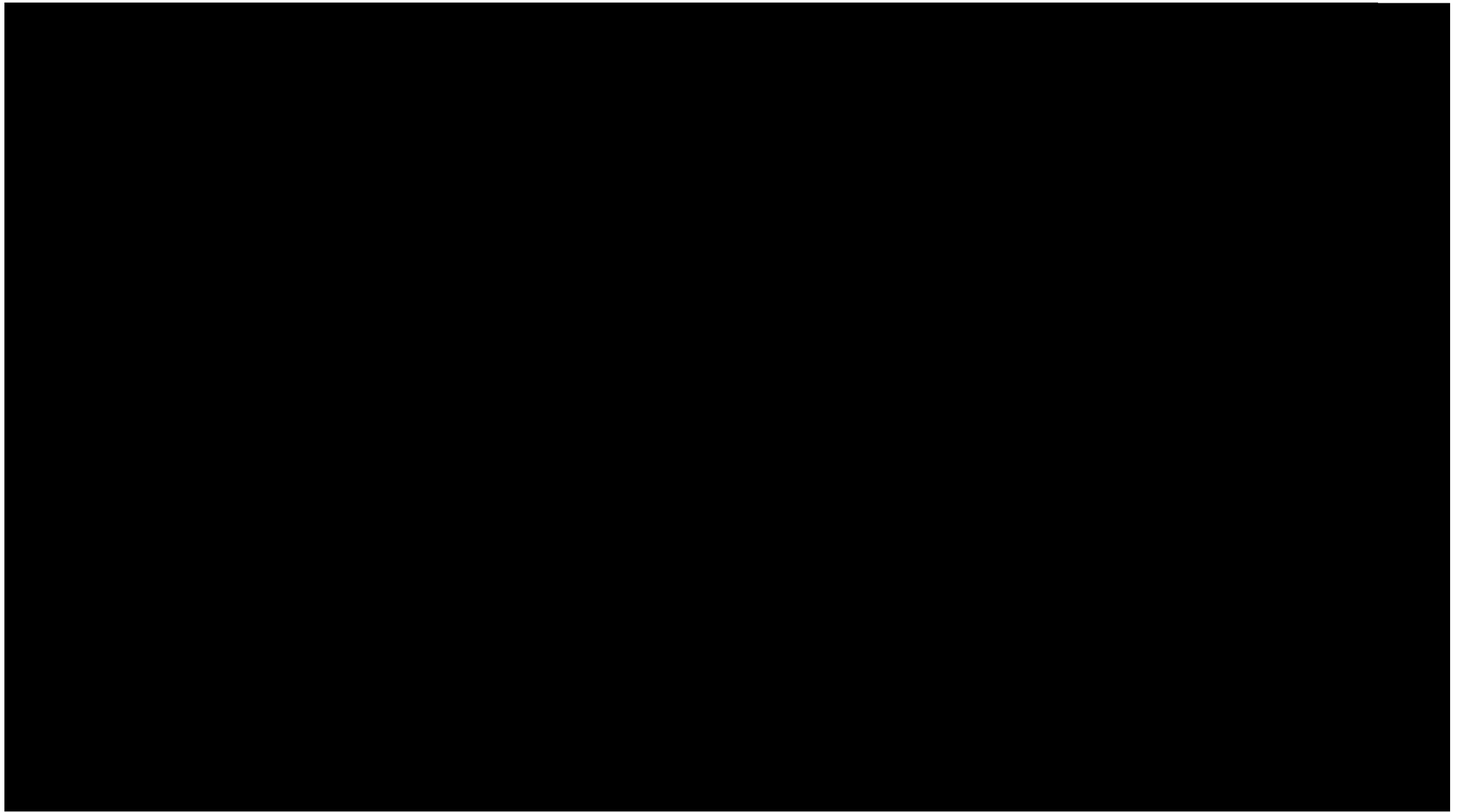
DEFINITION

> Web tracking refers specifically to the collection of detailed behavioral data through research software – typically browser extensions or similar tools – that participants voluntarily install on their devices. These tools record users' web activity in real time, including the websites they visit, the content they engage with, and their browsing behavior over time.

- Designed & user-centered digital behavioral data

(Mangold & Stier, 2025, S. 1–2)

BEISPIEL



Quelle: GESIS

VORTEILE UND HERAUSFORDERUNGEN

Vorteile

- Objektive Verhaltensmessung
- Hoch aufgelöste Längsschnittdaten
- Plattformübergreifend (solange im Browser)
- Erfassung der angezeigten Inhalte möglich

Herausforderungen

- Browser- und geräteübergreifende Messung, In-App-Messung
- Softwareentwicklung und Wartung
- Rekrutierung ausgewogener Stichproben
- Komplexe Datenaufbereitung und -auswertung

Fragen?

Beispielstudie: Gesundheitsbezogene Internetsuche

GESUNDHEITSBEZOGENE INTERNETSUCHE

Search Engine Use for Health-Related Purposes: Behavioral Data on Online Health Information-Seeking in Germany

Marko Bachl[a*], Elena Link[b*], Frank Mangold[c], and Sebastian Stier[c,d]

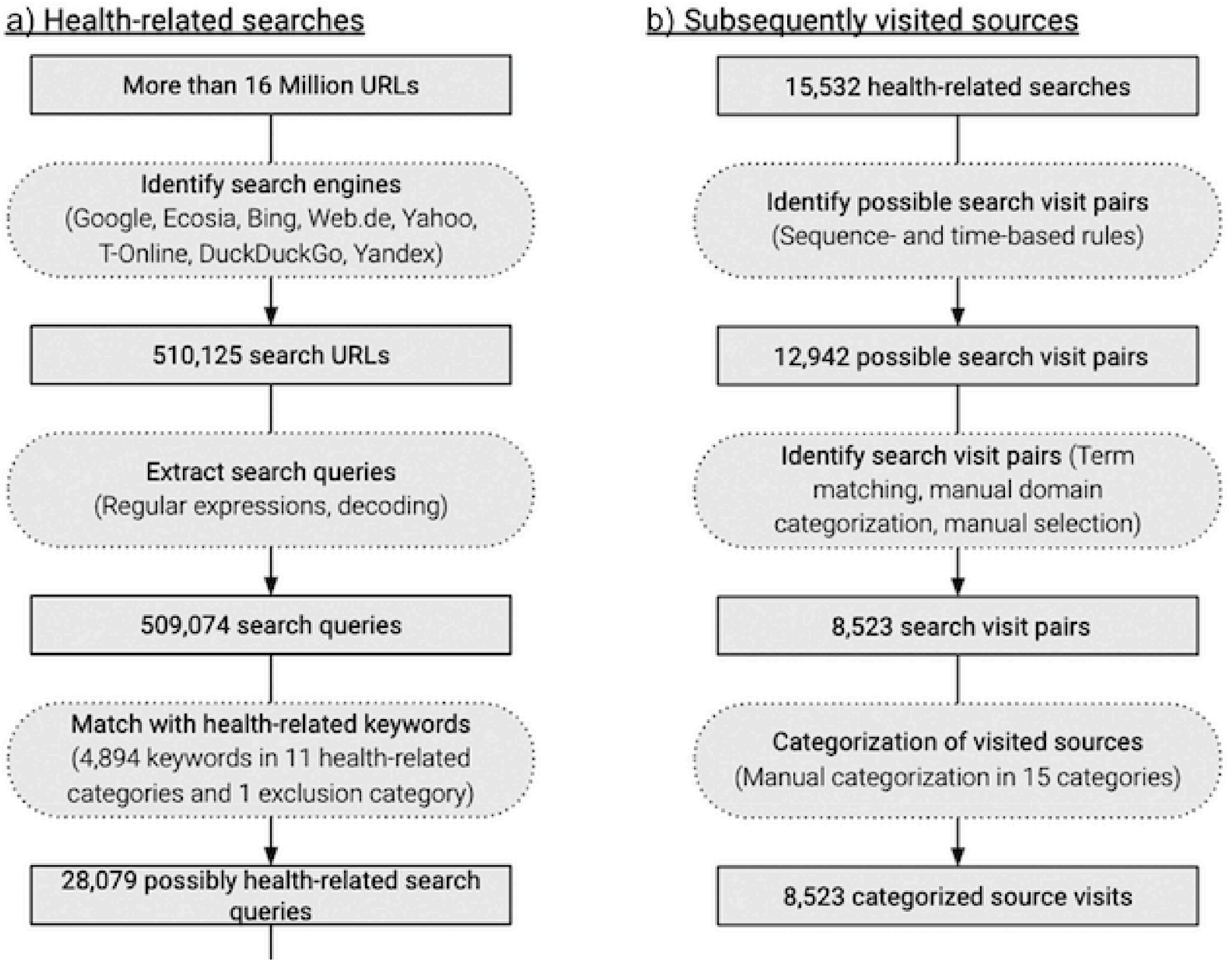
[a]Institute for Media and Communication Studies, Department of Political and Social Sciences, Freie Universität Berlin; [b]Department of Communication, Johannes Gutenberg University Mainz; [c]Department Computational Social Science, GESIS – Leibniz Institute for the Social Sciences; [d]School of Social Sciences, University of Mannheim

ABSTRACT

Internet searches for health-related purposes are common, with search engines like Google being the most popular starting point. However, results on the popularity of health information-seeking behaviors are based on self-report data, often criticized for suffering from incomplete recall, overreporting, and low reliability. Therefore, the current study builds on user-centric tracking of Internet use to reveal how individuals actually behave online. We conducted a secondary analysis of passively recorded Internet use logs to examine the prevalence of health-related search engine use, the types of health information searched for, and the sources visited after the searches. The analysis revealed two key findings. 1) We largely support earlier survey-based findings on the prevalence of online health information seeking with search engines and the relatively minor differences in information-seeking behaviors between socio-demographic groups. 2) We provide a more granular picture of the process of HISB using search engines by identifying different selection patterns depending on the scope of the searches.

(Bachl et al., 2024), Material: https://osf.io/4hnfv/

AUFBEREITEN DER WEBTRACKING-DATEN


a) Health-related searches
More than 16 Million URLs
Identify search engines
(Google, Ecosia, Bing, Web.de, Yahoo, T-Online, DuckDuckGo, Yandex)
510,125 search URLs
Extract search queries
(Regular expressions, decoding)
509,074 search queries
Match with health-related keywords
(4,894 keywords in 11 health-related categories and 1 exclusion category)
28,079 possibly health-related search queries
b) Subsequently visited sources
15,532 health-related searches
Identify possible search visit pairs
(Sequence- and time-based rules)
12,942 possible search visit pairs
Identify search visit pairs (Term matching, manual domain categorization, manual selection)
8,523 search visit pairs
Categorization of visited sources
(Manual categorization in 15 categories)
8,523 categorized source visits

EXTRAKTION DER SUCHTERME

Pakete laden und Beispielsuchen anlegen

```
1 library(urltools)
2 library(tidyverse)
3 urls <- c(
4   "https://userpage.fu-berlin.de/bachlm83/mame_publ
5   "https://www.google.com/search?q=was+tun+bei+hust
6   "https://www.google.de/maps/place/Institut+f%C3%B
7   "https://www.google.com/search?q=was+ist+das+ifpu
8 )
9 urls
```

```
[1] "https://userpage.fu-
berlin.de/bachlm83/mame_public/"
[2] "https://www.google.com/search?
q=was+tun+bei+husten%3F&oq=was+tun+bei+husten%3F&gs_lcr
8"
[3]
"https://www.google.de/maps/place/Institut+f%C3%BCr+Pub
+und+Kommunikationswissenschaft/@52.4601017,13.305423,1
entry=ttu&g_ep=EgoyMDI2MDEwNi4wIKXMDSoASAFQAw%3D%3D"
[4] "https://www.google.com/search?
q=was+ist+das+ifpuk&oq=was+ist+das+ifpuk&gs_lcrp"
```

URLs, die “google.” enthalten

```
1 google_urls <- urls |>
2   str_subset(fixed("google."))
3 google_urls
```

```
[1] "https://www.google.com/search?
q=was+tun+bei+husten%3F&oq=was+tun+bei+husten%3F&gs_lcr
8"
[2]
"https://www.google.de/maps/place/Institut+f%C3%BCr+Pub
+und+Kommunikationswissenschaft/@52.4601017,13.305423,1
entry=ttu&g_ep=EgoyMDI2MDEwNi4wIKXMDSoASAFQAw%3D%3D"
[3] "https://www.google.com/search?
q=was+ist+das+ifpuk&oq=was+ist+das+ifpuk&gs_lcrp"
```

URLs, die “search?” enthalten

```
1 google_search_urls <- google_urls |>
2   str_subset(fixed("search?"))
3 google_search_urls
```

```
[1] "https://www.google.com/search?
q=was+tun+bei+husten%3F&oq=was+tun+bei+husten%3F&gs_lcr
8"
[2] "https://www.google.com/search?
q=was+ist+das+ifpuk&oq=was+ist+das+ifpuk&gs_lcrp"
```

Suchanfrage extrahieren

```
1 search_queries <- google_search_urls |>
2   str_extract(regex("(?<=q\\=).*?(?=&)"))
3 search_queries
```

```
[1] "was+tun+bei+husten%3F" "was+ist+das+ifpuk"
```

Suchanfrage decodieren

```
1 search_queries_decoded <- search_queries |>
2   url_decode()
3 search_queries_decoded
```

```
[1] "was tun bei husten?" "was ist das ifpuk"
```

POTENTIELL GESUNDHEITSBEZOGENE SUCHANFRAGEN

- 11 gesundheitsbezogene Diktonäre und ein Ausschlussdiktionär
  - Scraping relevanter Listen, z.B. netDoktor Krankheiten von A bis Z oder Wikipedia Kategorie Arzneistoffe
  - Ergänzung aus eigener Expertise, Modifizieren und Löschen, damit als Diktionärseintrag geeignet
  - Induktive Ergänzung
- Optimiert für *Recall*: Möglichst viele potentiell relevante Suchanfragen finden, dabei *false positives* in Kauf nehmen
- Abgleichen mit Suchanfragen zur Identifikation potentiell gesundheitsbezogener Suchanfragen

BEISPIELDIKTIONÄRE

Krankheiten, Krankheitserreger und Symptome

```
*allergie, *anämie,
*ausfluss, *ausschlag,
*bakterien, *blutung,
*demenz*, *dysfunktion,
*dyspepsie, *dystrophie,
*entzünd*, *erkrankung*,
*geschwollen*, *geschwür*,
*hepatitis, *herpes,
*husten*, *hypertenie,
*hypoplasie, *hypotonie,
*infarkt, *infektion*,
*insuffizienz, *karzinom,
*krankheit*, *krebs
```

Medikamente und Wirkstoffe

```
*medikament*, *tablette*,
1-Butanol, 1-
Deoxygalactonojirimycin,
18-Methoxycoronaridin, 2-
Thiouracil, 3-Methyl-1-
phenyl-5-pyrazolon, 4-
Methylumbelliferon, 6-
Thioguanin, Abacavir,
Abamectin, Abarelix,
Abciximab, Abemaciclib,
Abirateron, Abrocitinib,
Acamprosat, Acarbose, ACE-
Hemmer, Acebutolol
```

Medizinische Fachgebiete, Fachkräfte, Institutionen

```
*apotheke, *arzt*, *ärzt*,
*chirurg*, *gynäkol*,
*klinik*, *krankenhaus*,
*krankenkasse*,
*krankenversicherung*,
*neurolog*, *onkolog*,
*orthopäd*, *pneumolog*,
*zahnzusatzversicherung*,
allergol*, ambulante
pflege, angiolog*, aok,
bkk, dermatolog*,
diabetolog*, doc morris,
docmorris, doctolib, dr
```

Ausschluss

```
*schrank, augenbraue*,
augenfarbe,
augenfleckstachelaal,
augenstängel*, ausbildung,
baum, bauplan, bauskizze,
bdsm, beruf*, blätter,
chinchilla, clipart, die
ärzte, die jungen ärzte,
dildo, dr alban, dr doom,
dr dre, dr gachet, dr
house, dr mabuse, dr
moreau, dr movie, dr
music, dr no, dr seuss, dr
```

GESUNDHEITSBEZOGENE SUCHANFRAGEN

- Manuelle Prüfung der identifizierten Suchanfragen (OSF)

SUCHE-BESUCH-PAARE

search	domain	url
bmi	fitforfun.de	www.fitforfun.de/abnehmen/schlankmacher/bmi-rechner_aid_3821.html
fahrradunfall kieferbruch	nwzonline.de	www.nwzonline.de/gesundheit/ein-fall-fuer-den-chirurgen_a_9,3,2876442063.ht
psychotherapie fürth	vhs-fuerth.de	vhs-fuerth.de/programm/gesundheit-ernaehrung.html/kurs/477-C-191-42009/t/
asbach autismus	wikipedia.org	de.wikipedia.org/wiki/Asperger-Syndrom
wozu kreatinwert für mrt	hilferuf.de	www.hilferuf.de/forum/gesundheit/193950-wichtig-mrt-auch-ohne-keratinwert.h
dr. höllerhagen	cylex.de	web2.cylex.de/medizin/neurochirurgie/Hachenburg
hautarzt berlin kreuzberg	google.com	www.google.com/maps/uv?hl=en&pb=[redacted]//geo1.ggpht.com/cb
brustverkleinerung	bodenseeklinik.de	www.bodenseeklinik.de/bruststraffung-verkleinerung
heilpraktiker hormone fürth	lisa-montag.de	lisa-montag.de
tripper test	iwwit.de	www.iwwit.de/geschlechtskrankheiten
krankenversicherungsbeitrag rentner	deutsche-rentenversicherung.de	www.deutsche-rentenversicherung.de/Allgemein/de/Navigation/2_Rente_Reha/01_Rente/04_in
in den usa urlaub operation	mawista.com	www.mawista.com/auslandskrankenversicherung/usa/
pille zoely kaufen online	treated.com	de.treated.com/checkout?orderid=f1a40a2c-ce95-4d50-832f-673f893da223
Cuprum aceticum schilddrüse	heel.de	www.heel.de/media/de/downloads_pdf/heel_de_1/gebrauchsinformationen_bp

search	domain	url
zahnwurzelbehandlung fruckempfindlich	nie-wieder-zahnschmerzen.de	www.nie-wieder-zahnschmerzen.de/schmerzen-nach-wurzelbehandlung/
ohr schwill zu	yahoo.com	de.yahoo.com/

GESUNDHEITSBEZOGENE BESUCHE

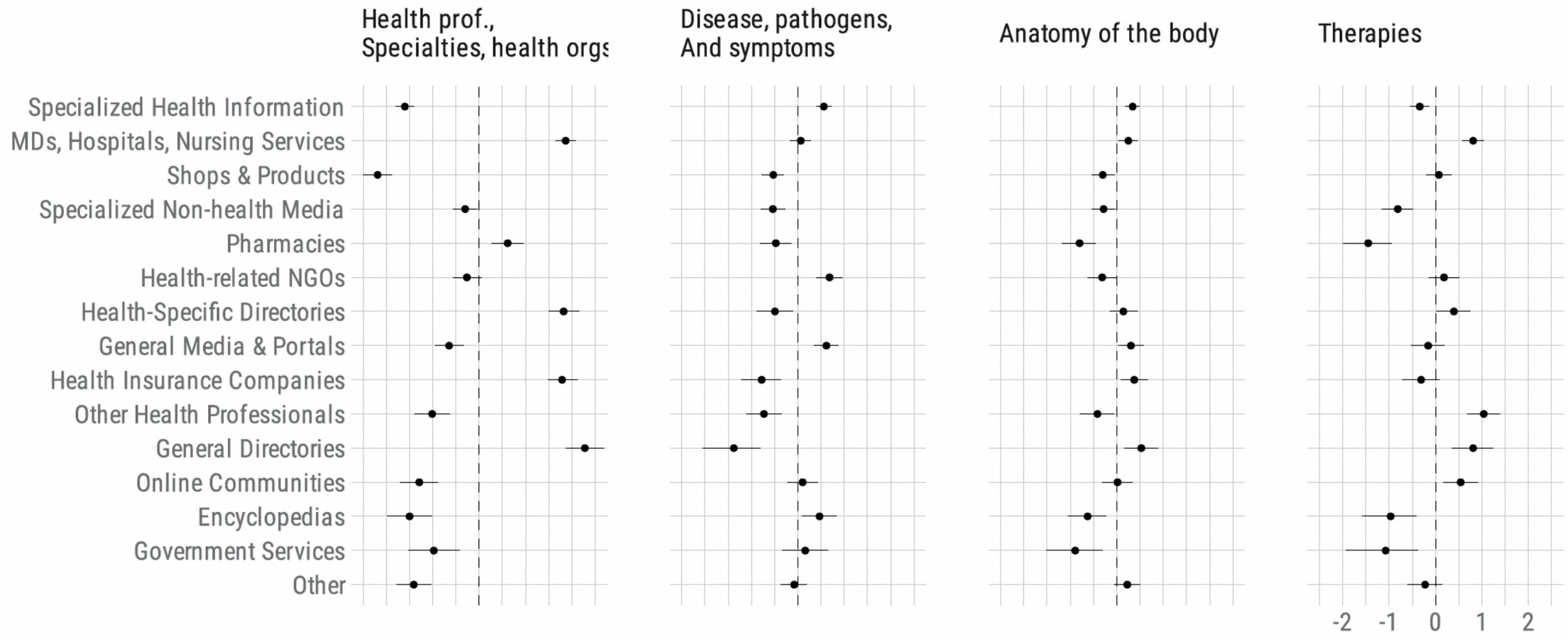
- Abgleich von Suchbegriffen und URLs
- Vorkodierung einschlägig gesundheitsbezogener URLs
- Manuelle Prüfung und Kategorisierung (OSF)

KATEGORISIERTE SUCHE-BESUCH-PAARE

category	search	domain	url
Specialized Health Information	bmi	fitforfun.de	www.fitforfun.de/abnehmen/schlankmacher/bmi-rechner_aid_
General Media & Portals	fahrradunfall kieferbruch	nwzonline.de	www.nwzonline.de/gesundheit/ein-fall-fuer-den-chirurgen_a_9
Specialized Media & Portals	psychotherapie fürth	vhs-fuerth.de	vhs-fuerth.de/programm/gesundheit-ernaehrung.html/kurs/47
Encyclopedias	asbach autismus	wikipedia.org	de.wikipedia.org/wiki/Asperger-Syndrom
Peer Advice	wozu kreatinwert für mrt	hilferuf.de	www.hilferuf.de/forum/gesundheit/193950-wichtig-mrt-auch-o
Health-Specific Directories	dr. höllerhagen	cylex.de	web2.cylex.de/medizin/neurochirurgie/Hachenburg
General Directories	hautarzt berlin kreuzberg	google.com	www.google.com/maps/uv?hl=en&pb=[redacted]//ge
Medical Doctors, Hospitals, Nursing Services	brustverkleinerung	bodenseeklinik.de	www.bodenseeklinik.de/bruststraffung-verkleinerung

category	search	domain	url
Other Health Professionals	heilpraktiker hormone fürth	lisa-montag.de	lisa-montag.de
Health-related NGO	tripper test	iwwit.de	www.iwwit.de/geschlechtskrankheiten
Government Services	krankenversicherungsbeitrag rentner	deutsche-rentenversicherung.de	www.deutsche-rentenversicherung.de/Allgemein/de/Navigation/2_Rente_Reha
Health Insurance Companies	in den usa urlaub operation	mawista.com	www.mawista.com/auslandskrankenversicherung/usa/
Pharmacies	pille zoely kaufen online	treated.com	de.treated.com/checkout?orderid=f1a40a2c-ce95-4d50-832f-673
Shops & Products	Cuprum aceticum schilddrüse	heel.de	www.heel.de/media/de/downloads_pdf/heel_de_1/gebrauchsi
Other	zahnwurzelbehandlung fruckempfindlich	nie-wieder-zahnschmerzen.de	www.nie-wieder-zahnschmerzen.de/schmerzen-nach-wurzelbe

ERGEBNISBEISPIEL



(Bachl et al., 2024)

Fragen?

Fazit

FAZIT

- Digitale Verhaltensdaten haben großes Potenzial für die Kommunikationswissenschaft, da sie kommunikationsbezogene Handlungen (z.B. Selektion, Rezeption, Interaktion, Kreation) in hoher Auflösung erfassen können.
- Je nach Datenzugang (Plattform- oder User-centered data collection) und Design der Messung (found oder designed data) sind die Daten bezüglich verschiedener Merkmale unterschiedlich aussagekräftig.
- Wertvolle Ergänzung der sozialwissenschaftlichen Standarddatentypen (v.a. von Selbstauskünften)

MEHR WEBTRACKING?

Workshop: *Web Tracking: From Raw Data to Analysis*

Zeit und Ort: 12. & 13. Februar, ca. 9-16 Uhr, Garystr. 55

Dozent: Dr. Frank Mangold (GESIS)

Anmeldung: 23. Januar. Noch einzelne Plätze verfügbar.

Inhalt: The Internet has profoundly reshaped everyday life and media use. Traditional self-reports reach their limits when capturing the diversity and dynamics of digital behavior. Web tracking offers a behaviorally grounded approach—especially in combination with panel surveys. At the same time, such data pose theoretical, infrastructural, ethical, and legal challenges. This workshop provides a practice-oriented introduction to preprocessing and analyzing web-tracking data, highlights typical decision points (“researcher degrees of freedom”) and best practices, and—together with participants—develops suitable research designs in Computational Communication Research. Alongside conceptual and methodological input, hands-on exercises with example datasets in R are a central focus.

Alle Informationen und Link zur Anmeldung

Fragen?

HAUSAUFGABE

Lesen Sie diese beiden GESIS Guides to Digital Behavioral Data:

- What is Digital Behavioral Data? (Wagner et al., 2025)
- Overview of Working with Web Tracking Data (Mangold & Stier, 2025)

Nächste Einheit

Kausalinferenz

Danke

Marko Bachl

marko.bachl@fu-berlin.de

LITERATUR

Bachl, M. (2025). Self-reported media exposure. In A. Nai, M. Grömping, & D. Wirz (Hrsg.), *Elgar Encyclopedia of Political Communication* (S. 465–468). Edward Elgar Publishing. https://doi.org/10.31219/osf.io/m76y8

Bachl, M., Link, E., Mangold, F., & Stier, S. (2024). Search engine use for health-related purposes: Behavioral data on online health information-seeking in Germany. *Health Communication*, *39*(8), 1651–1664. https://doi.org/10/gtg7gv

Donsbach, W. (1991). Exposure to political content in newspapers: The impact of cognitive dissonance on readers' selectivity. *European Journal of Communication*, *6*(2), 155–186. https://doi.org/10/bj5kt8

Guess, A., Munger, K., Nagler, J., & Tucker, J. (2019). How accurate are survey responses on social media and politics? *Political Communication*, *36*(2), 241–258. https://doi.org/10/gfs6mz

Mangold, F., & Stier, S. (2025). *Overview of working with web tracking data (GESIS guides to digital behavioral data, 23)*. GESIS - Leibniz-Institute for the Social Sciences. https://doi.org/10.60762/GGDBD25023.1.0

Miller, A. H., Goldenberg, E. N., & Erbring, L. (1979). Type-set politics: Impact of newspapers on public confidence. *American Political Science Review*, *73*(01), 67–84. https://doi.org/10/bnw5fw

Prior, M. (2009). The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly*, *73*(1), 130–143. https://doi.org/10/dx3vbs

Scharkow, M. (2016). The accuracy of self-reported Internet use — A validation study using client log data. *Communication Methods and Measures*, *10*(1), 13–27. https://doi.org/10/gdvnnw

Wagner, C., Stier, S., & Zens, M. (2025). *What is digital behavioral data? (GESIS guides to digital behavioral data, 1)*.