gesis Leibniz-Institut für Sozialwissenschaften

Introduction to Online Data Acquisition

Meet the Experts! – GESIS online talks

Dr. Roberto Ulloa · *September 30, 2021*

Leibniz Gemeinschaft

gesis Leibniz-Institut für Sozialwissenschaften

Meet the Experts

Speaker

Dr. Roberto Ulloa

- Researcher in the team Social Analytics and Services, Department Computational Social Science
- Interested in Digital Institutions, Computer Simulations, Algorithm Auditing
- Contact: roberto.ulloa@gesis.org

gesis Leibniz-Institut für Sozialwissenschaften

Meet the Experts

Agenda

1. Why collect online data?
2. Platform selection
3. How to access online data?
4. Cases:
   - CrowdTangle
   - Comparison
   - Other platforms
   - Annotation APIs
5. Takeaways

Online Data

Web Data

Digital Behavioural Data

Social Media Data

Online Annotated Data

…

Why collect online data?

Online platforms are shaping society

Online platforms are shaping society

- mediate human communication

140 characters → 280 characters in 2017

Online platforms are shaping society

- mediate human communication
- political (ab)use

Deutscher Bundestag, 71.3K subscribers

@Bundestag, 12.4K followers

@derbundestag, 14.1K followers and 130.8K likes, except…

Broken Promises: TikTok and the German Election

Mozilla research reveals that TikTok is struggling to curb disinformation ahead of the German Federal Election 2021.

foundation.mozilla.org

https://foundation.mozilla.org/en/campaigns/tiktok-german-election-2021/

Online platforms are shaping society

- mediate human communication
- political (ab)use
- information gatekeepers

BertelsmannStiftung EN

Patients value "Dr. Google's" versatility

Whether preparing for a visit to the doctor, comparing therapies or simply engaging in online discussion with others – many people seek advice from "Dr. Google". Findings from our new study show that more than one-half of patients surveyed are satisfied with the health information they find online. But are physicians and patients making good use of the internet's potential?

https://www.bertelsmann-stiftung.de/en/topics/latest-news/2018/januar/patients-value-dr-googles-versatility

Online platforms are shaping society

- mediate human communication
- political (ab)use
- information gatekeepers
- daily algorithmic recommendations and advertising: news, products, job offers, job applicants, insurances, hotels, …

Platform selection

think about your research question…

- which population is represented?
- types of interactions that are important, e.g.: one to one or one to many, short or long
- which interaction rules are important?
- does the platform provide access to the data you need?
    - If not, can you get the data in a different way?
    - If you do, is it legal/ethical? [1]
- be careful with the way you collect the data[2,3]

[1] Watteler, O., Weller, K.,. Research Ethics and Data Protection in Social Media. Meet the Experts. https://www.youtube.com/watch?v=T6q-IcPY8GE

[2] Sen, I., Floeck, F., Weller, K., Weiss, B., & Wagner, C. (2021). TED-On: A Total Error Framework for Digital Traces of Human Behavior on Online Platforms. *ArXiv:1907.08228 [Cs]*. http://arxiv.org/abs/1907.08228

[3] Floeck, F., Sen, I. Digital Traces of Human Behavior from Online Platforms – Research Designs and Error Sources. Meet the Experts. https://www.youtube.com/watch?v=y9mVuQnXWec

How to access online data?

- Repositories
- Direct access
- Data Donations
- Web Scraping / Crawling
- Web APIs
- Web Tracking
- Automatized browsing

- Repositories
- Direct access
- Data Donations
- Web Scraping / Crawling
- **Web APIs**
- Web Tracking
- Automatized browsing

the reason is…

web scraping

- ✓ what you see is what you get
- ✗ more programming
- ✗ often violates T&C

web APIs

- ✗ what the platform provides you with
- ✓ little programming
- ✓ API itself prevents violations of T&C

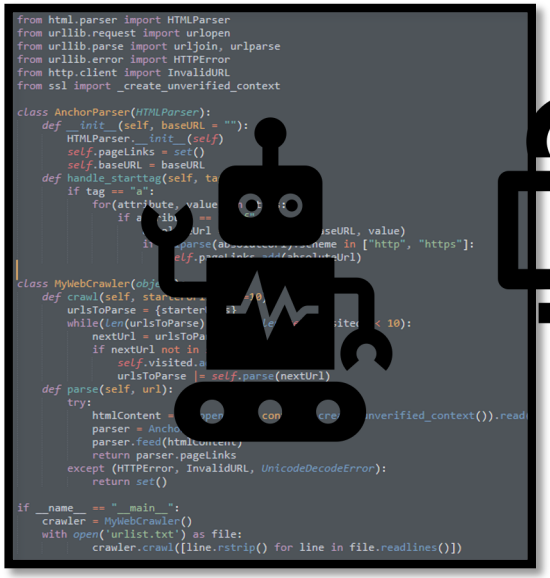
web scraping vs web crawling

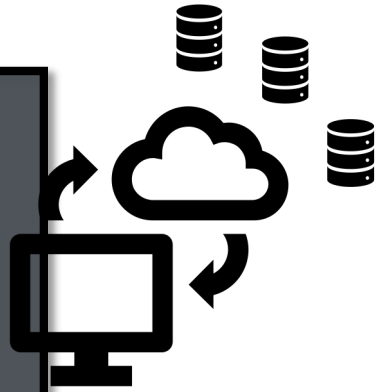
Web scraping

URL list

http://www.diw.de/en/
http://www.gesis.org/en/
http://www.giga-hamburg.de
http://www.iamo.de/
http://www.leibniz-hbi.de
http://www.ioer.de/
http://www.irs-net.de
http://www.iwh-halle.de
http://www.hsfk.de/index.php
http://en.rwi-essen.de/
https://safe-frankfurt.de/
http://www.zbw-kiel.de/

Robot (Bot)
Program - Script



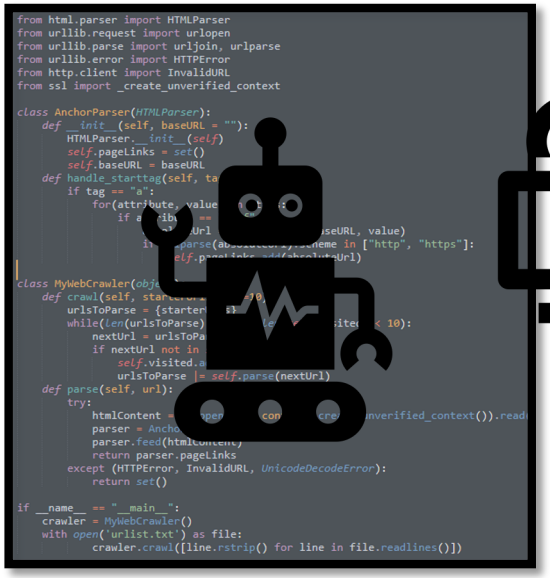


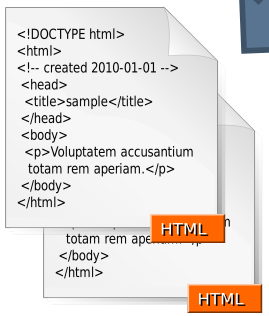
Web crawling

URL list

http://www.diw.de/en/
http://www.gesis.org/en/
http://www.giga-hamburg.de
http://www.iamo.de/
http://www.leibniz-hbi.de
http://www.ioer.de/
http://www.irs-net.de
http://www.iwh-halle.de
http://www.hsfk.de/index.php
http://en.rwi-essen.de/
https://safe-frankfurt.de/
http://www.zbw-kiel.de/

Robot (Bot)
Program - Script



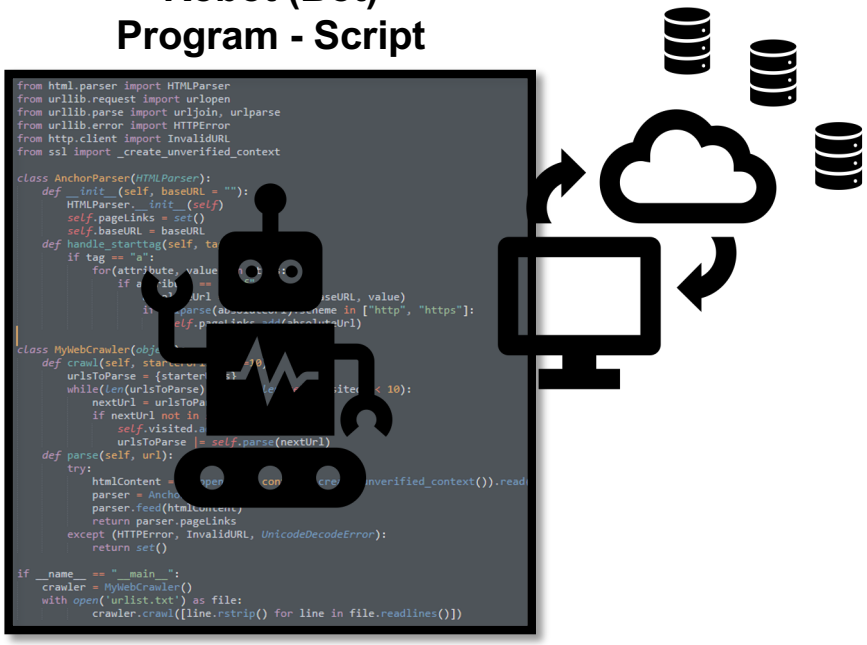


Web crawling

URL list

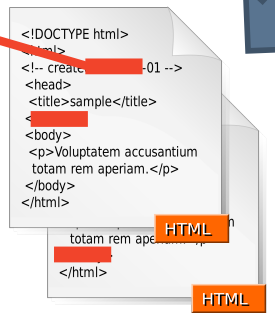
http://www.diw.de/en/
http://www.gesis.org/en/
http://www.giga-hamburg.de
http://www.iamo.de/
http://www.leibniz-hbi.de
http://www.ioer.de/
http://www.irs-net.de
http://www.iwh-halle.de
http://www.hsfk.de/index.php
http://en.rwi-essen.de/
https://safe-frankfurt.de/
http://www.zbw-kiel.de/

Robot (Bot)
Program - Script



```
from html.parser import HTMLParser
from urllib.request import urlopen
from urllib.parse import urljoin, urlparse
from urllib.error import HTTPError
from http.client import InvalidURL
from ssl import _create_unverified_context

class AnchorParser(HTMLParser):
    def __init__(self, baseURL = ""):
        HTMLParser.__init__(self)
        self.pageLinks = set()
        self.baseURL = baseURL
    def handle_starttag(self, ta
        if tag == "a":
            for(attribute, value
                if a
                    Url
                    parse(a
                    self.p
class MyWebCrawler(obje
    def crawl(self, start
        urlsToParse = {starter
        while(len(urlsToParse)
            nextUrl = urlsToPa
            if nextUrl not in
                self.visited.a
                urlsToParse |= self.parse(nextUrl)
    def parse(self, url):
        try:
            htmlContent =
            parser = Ancho
            parser.feed(htmlCon
            return parser.pageLinks
        except (HTTPError, InvalidURL, UnicodeDecodeError):
            return set()

if __name__ == "__main__":
    crawler = MyWebCrawler()
    with open('urlist.txt') as file:
        crawler.crawl([line.rstrip() for line in file.readlines()])
```

https://www.gesis.org/en/research

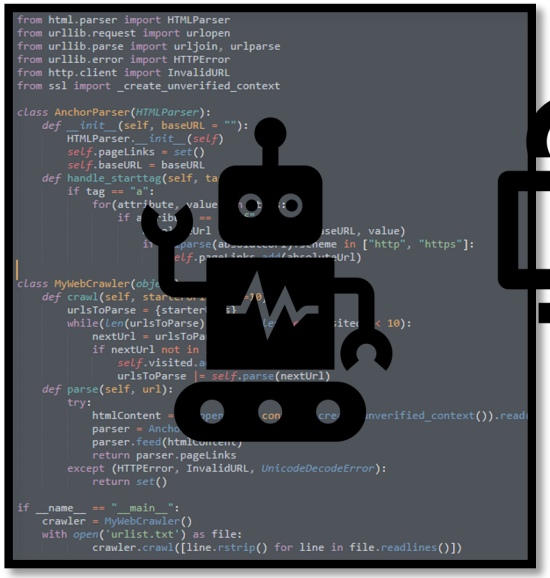


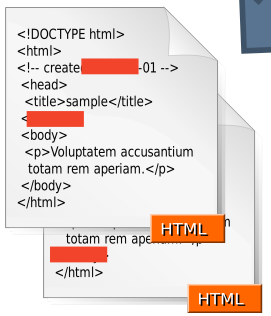
Web crawling

URL list

http://www.diw.de/en/
http://www.gesis.org/en/
http://www.giga-hamburg.de
http://www.iamo.de/
http://www.leibniz-hbi.de
http://www.ioer.de/
http://www.irs-net.de
http://www.iwh-halle.de
http://www.hsfk.de/index.php
http://en.rwi-essen.de/
https://safe-frankfurt.de/
http://www.zbw-kiel.de/
https://www.gesis.org/en/research
…

Robot (Bot)
Program - Script





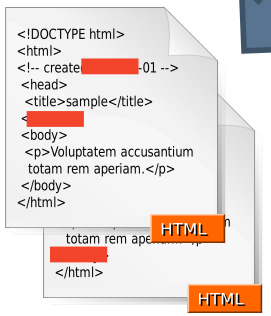
Web crawling

URL list

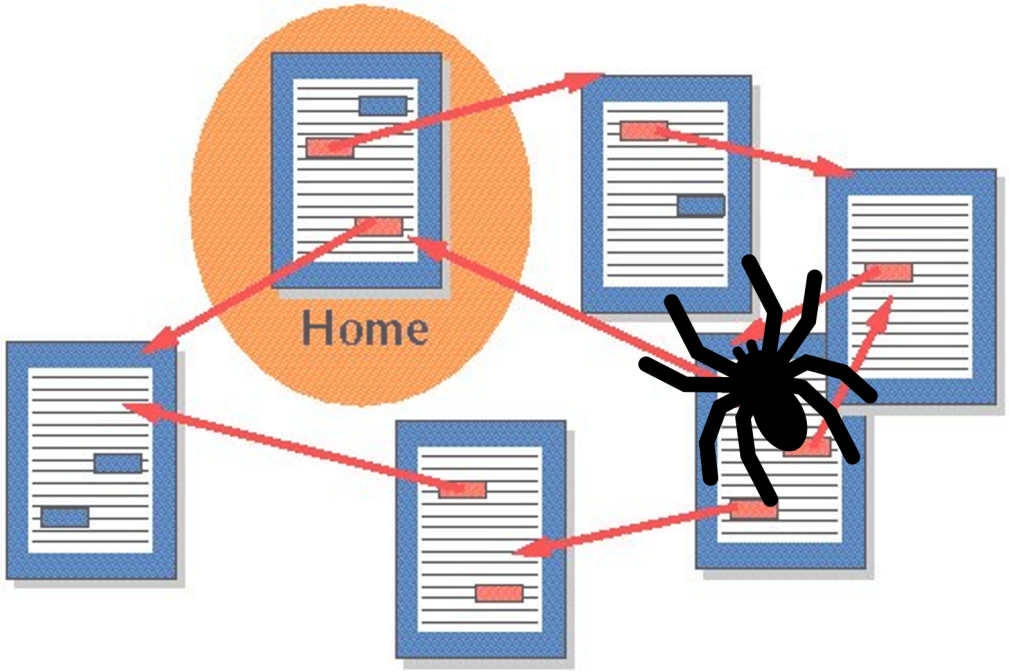
http://www.diw.de/en/
http://www.gesis.org/en/
http://www.giga-hamburg.de
http://www.iamo.de/
http://www.leibniz-hbi.de
http://www.ioer.de/
http://www.irs-net.de
http://www.iwh-halle.de
http://www.hsfk.de/index.php
http://en.rwi-essen.de/
https://safe-frankfurt.de/
http://www.zbw-kiel.de/
https://www.gesis.org/en/research
…

web crawler (spider)

```
from html.parser import HTMLParser
from urllib.request import urlopen
from urllib.parse import urljoin, urlparse
from urllib.error import HTTPError
from http.client import InvalidURL
from ssl import _create_unverified_context

class AnchorParser(HTMLParser):
    def __init__(self, baseURL = ""):
        HTMLParser.__init__(self)
        self.pageLinks = set()
        self.baseURL = baseURL
    def handle_starttag(self, [redacted], attrs):
        if tag == "a":
            for([redacted], [redacted] in attrs:
                if [redacted] == "href":
                    [redacted] value)
                    [redacted] ["http", "https"]:
                        [redacted]

class MyWebCrawler(object):
    def crawl(self, [redacted]
        urlsToParse = [redacted]
        while([redacted] < 10):
            nextUrl = [redacted]
            [redacted]
                [redacted]
                [redacted]
    def parse(self, url):
        try:
            htmlContent = urlopen(url, context=_create_unverified_context()).[redacted]
            parser = AnchorParser(url)
            parser.feed(htmlContent)
            return parser.pageLinks
        except (HTTPError, InvalidURL, UnicodeDecodeError):
            return set()

if __name__ == "__main__":
    crawler = MyWebCrawler()
    with open('urlist.txt') as file:
            crawler.crawl([line.rstrip() for line in file.readlines()])
```



crawl the web



API

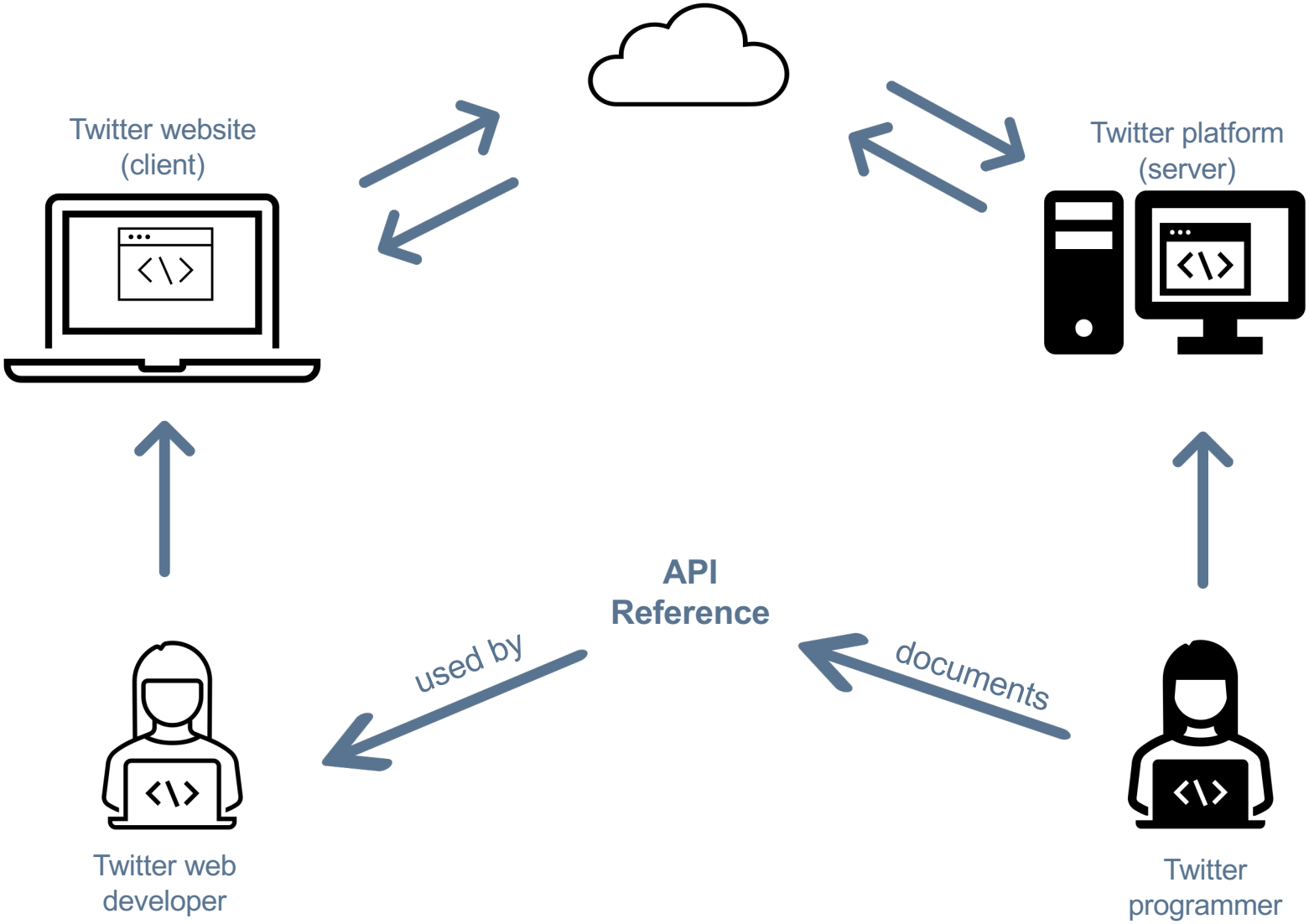
Application Programming Interface

gesis Leibniz-Institut für Sozialwissenschaften

Meet the Experts

Web API

Web Application Programming Interface


Twitter website
(client)
Twitter platform
(server)
API
Reference
used by
documents
Twitter web
developer
Twitter
programmer

The purpose of APIs is
to **coordinate communication between programs**,
not to provide data to researchers

- query system (parameters)
- programming (API scraping)
- data is not be tabulated, e.g. Excel, instead JSON/XML
- results are going to come in chunks (100 “rows”)
- limits: requests per minute

they are getting **more common**,
even **for providing data**

- query system (parameters)
- programming (API scraping)
- data is not be tabulated, e.g. Excel, instead JSON/XML
- results are going to come in chunks (100 “rows”)
- limits: requests per minute

- query system (parameters)
- ~~programming (API scraping)~~
- ~~data is not be tabulated, e.g. Excel, instead JSON/XML~~
- ~~results are going to come in chunks (100 “rows”)~~
- limits: ~~requests per minute~~ MB per minute

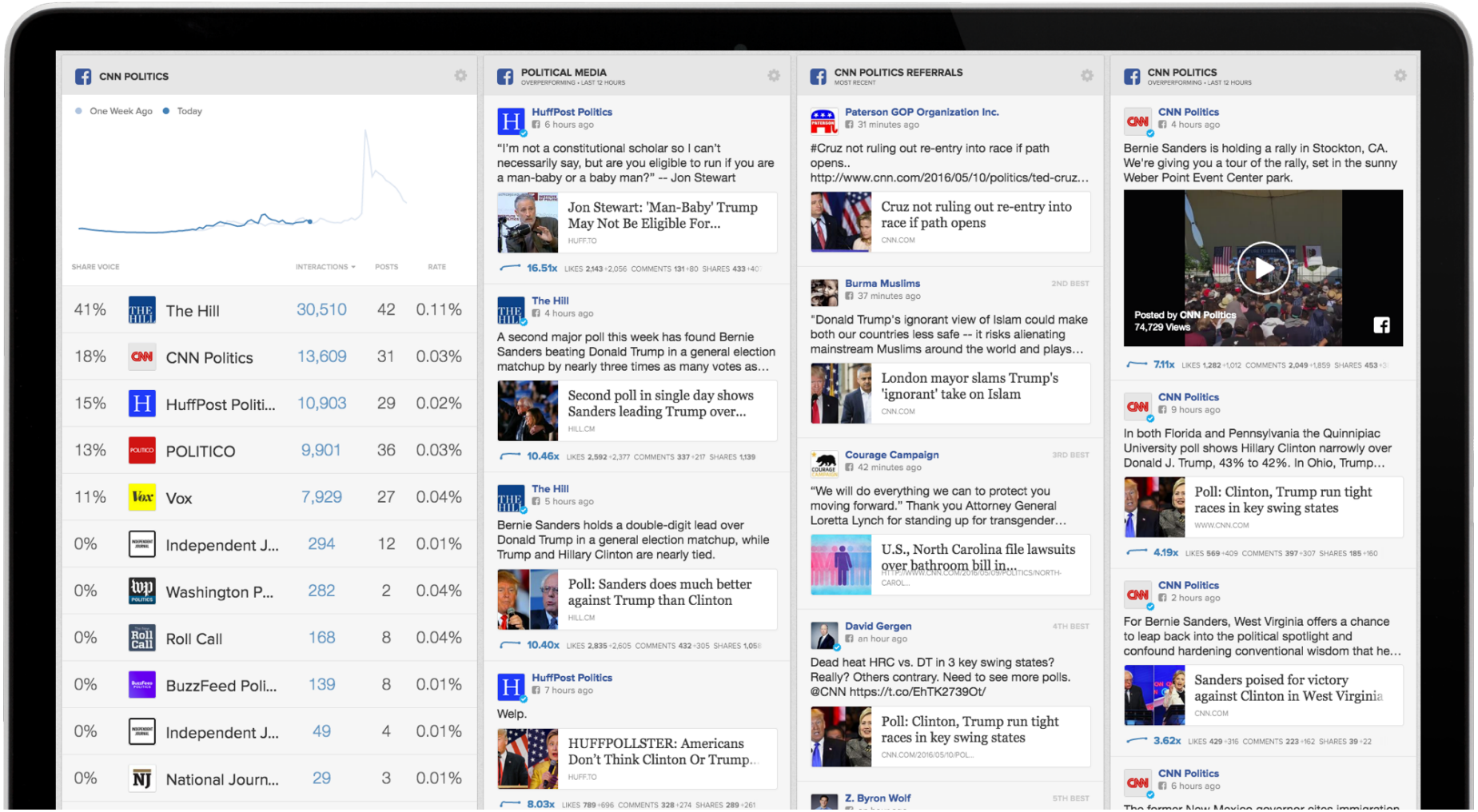
CrowdTangle

- 7M+ pages, groups, and verified profiles
- 2M+ public Instagram accounts
- 20K+ of the most active sub-reddits[1]

Source: https://help.crowdtangle.com/en/articles/1140930-what-data-is-crowdtangle-tracking

[1] More details: https://www.crowdtangle.com/reddit

create a list of accounts (related to a topic)
to browse or get the content posted by the accounts


CNN POLITICS
One Week Ago
Today
SHARE VOICE
INTERACTIONS
POSTS
RATE
41% The Hill 30,510 42 0.11%
18% CNN Politics 13,609 31 0.03%
15% HuffPost Politi... 10,903 29 0.02%
13% POLITICO 9,901 36 0.03%
11% Vox 7,929 27 0.04%
0% Independent J... 294 12 0.01%
0% Washington P... 282 2 0.04%
0% Roll Call 168 8 0.04%
0% BuzzFeed Poli... 139 8 0.01%
0% Independent J... 49 4 0.01%
0% National Journ... 29 3 0.01%
POLITICAL MEDIA
OVERPERFORMING • LAST 12 HOURS
HuffPost Politics
6 hours ago
"I'm not a constitutional scholar so I can't necessarily say, but are you eligible to run if you are a man-baby or a baby man?" -- Jon Stewart
Jon Stewart: 'Man-Baby' Trump May Not Be Eligible For...
HUFF.TO
16.51x LIKES 2,143 +2,056 COMMENTS 131 +80 SHARES 433
The Hill
4 hours ago
A second major poll this week has found Bernie Sanders beating Donald Trump in a general election matchup by nearly three times as many votes as...
Second poll in single day shows Sanders leading Trump over...
HILL.CM
10.46x LIKES 2,592 +2,377 COMMENTS 337 +217 SHARES 1,139
The Hill
5 hours ago
Bernie Sanders holds a double-digit lead over Donald Trump in a general election matchup, while Trump and Hillary Clinton are nearly tied.
Poll: Sanders does much better against Trump than Clinton
HILL.CM
10.40x LIKES 2,835 +2,605 COMMENTS 432 +305 SHARES 1,05
HuffPost Politics
7 hours ago
Welp.
HUFFPOLLSTER: Americans Don't Think Clinton Or Trump...
HUFF.TO
8.03x LIKES 789 +696 COMMENTS 328 +274 SHARES 289 +261
CNN POLITICS REFERRALS
MOST RECENT
Paterson GOP Organization Inc.
31 minutes ago
#Cruz not ruling out re-entry into race if path opens..
http://www.cnn.com/2016/05/10/politics/ted-cruz...
Cruz not ruling out re-entry into race if path opens
CNN.COM
Burma Muslims
2ND BEST
37 minutes ago
"Donald Trump's ignorant view of Islam could make both our countries less safe -- it risks alienating mainstream Muslims around the world and plays...
London mayor slams Trump's 'ignorant' take on Islam
CNN.COM
Courage Campaign
3RD BEST
42 minutes ago
"We will do everything we can to protect you moving forward." Thank you Attorney General Loretta Lynch for standing up for transgender...
U.S., North Carolina file lawsuits over bathroom bill in...
David Gergen
4TH BEST
an hour ago
Dead heat HRC vs. DT in 3 key swing states? Really? Others contrary. Need to see more polls. @CNN https://t.co/EhTK2739Ot/
Poll: Clinton, Trump run tight races in key swing states
CNN.COM/2016/05/10/POL...
Z. Byron Wolf
5TH BEST
CNN POLITICS
OVERPERFORMING • LAST 12 HOURS
CNN Politics
4 hours ago
Bernie Sanders is holding a rally in Stockton, CA. We're giving you a tour of the rally, set in the sunny Weber Point Event Center park.
Posted by CNN Politics
74,729 Views
7.11x LIKES 1,282 +1,012 COMMENTS 2,049 +1,859 SHARES 453
CNN Politics
9 hours ago
In both Florida and Pennsylvania the Quinnipiac University poll shows Hillary Clinton narrowly over Donald J. Trump, 43% to 42%. In Ohio, Trump...
Poll: Clinton, Trump run tight races in key swing states
WWW.CNN.COM
4.19x LIKES 569 +409 COMMENTS 397 +307 SHARES 185 +160
CNN Politics
2 hours ago
For Bernie Sanders, West Virginia offers a chance to leap back into the political spotlight and confound hardening conventional wisdom that he...
Sanders poised for victory against Clinton in West Virginia
CNN.COM
3.62x LIKES 429 +316 COMMENTS 223 +162 SHARES 39 +22
CNN Politics
6 hours ago

search

exports and imports

interactive plots

also, an API…

✓

content
from public pages, public groups, and verified profiles.

interactions
# of reactions, comments, shares - historical data at the post level

page likes
# of likes of the page since added to CrowdTangle

Facebook video views
owned, crossposted, shared

✗

comment text

demographic data

page reach

traffic & clicks

private posts & profiles

pages that are geo-gated or age-gated

paid or boosted posts
CrowdTangle can't tell if a post was boosted or differentiate between organic and paid engagement

gesis Leibniz-Institut für Sozialwissenschaften

Meet the Experts

Academic and Research Access

https://help.crowdtangle.com/en/articles/4302208-crowdtangle-for-academics-and-researchers
support@crowdtangle.com

REQUESTER INFORMATION

First Name *

Last Name *

Please use your university email address and ensure it is valid in order to receive the Terms of Service agreement.
We cannot process applications using a personal email address.

Email Address *

AFFILIATED ACADEMIC INSTITUTION

Under List Institution Name dropdown, type in a few characters from your Institution name to see if it shows up and then click on it.
If it doesn't, select "OTHER" and a new "Other Institution Name" box will display where you can enter it.

List Institution Name *

GESIS - Leibniz-Institute for Social.::.

Were you referred by a Facebook team? *

-- Select one --

What is your research about? *

In one paragraph, please describe what your research is about.

How do you plan to use CrowdTangle to support your research? *

In one paragraph, please describe your plan for using CrowdTangle data to support your research.

Submit

gesis Leibniz-Institut für Sozialwissenschaften

Meet the Experts

Social Media (APIs) comparison

gesis Leibniz-Institut für Sozialwissenschaften

Meet the Experts

What are they best for?

Messages from "influencers"

- ✓ most prolific 10% create 80% of tweets[1]
- ✓ media follow and propagate Tweets of personalities
- ✓ for instance, politicians:
  - ✓ topic saliences (Stier et al., 2018)
  - ✓ communication networks (Lietz et al., 2014)

- Twitter use in Germany: 5.2%, +16yo (GLES Cross-Section Survey 2017[2])

Engagement and content popularity

- ✓ 2nd most visited website and social network (US and Germany)[3,4]
- ✓ does not strongly encourage being logged in to watch its content
- ✓ ~19% US adults considers it an important way to get news[5]
- ✓ presence of "traditional" media
- ✓ capture some TV consumer market[6]

- users across many demographic groups

Emergence of moderation rules

- ✓ >100K active "subreddits" (topical communities)
- ✓ individual norms and cultures, and moderation practices
- ✓ moderators are volunteers (no moderator, no community)
- ✓ home to (organized) events, e.g. controversial: Boston-city bombing terrorists, GameStop stock

- subreddit members are often very engaged

[1] Pew Research (2019)
[2] GLES (2017). *https://doi.org/10.4232/1.13213*
[3] https://www.alexa.com/topsites/countries/DE
[4] https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/
[5] https://www.pewresearch.org/journalism/2020/09/28/many-americans-get-news-on-youtube-where-news-organizations-and-independent-producers-thrive-side-by-side/
[6] https://www.cnbc.com/2021/07/28/youtube-is-a-proven-juggernaut-that-rivals-netflix-in-the-streaming-wars.html

What has data been collected for?

- election prediction (Singh et al., 2020; Tumasjan et al., 2010; Wang & Gan, 2017)
- migration movements (Zagheni et al., 2014)
- depression (De Choudhury et al., 2013; Nadeem, 2016; Tsugawa et al., 2015)
- hate speech (Burnap & Williams, 2015; Watanabe et al., 2018; Zhang & Luo, 2019)

- sentiment of captions/transcripts (Soldner et al., 2019)
- misinformation (Donzelli et al., 2018; Hussein et al., 2020; Khatri et al., 2020; Li et al., 2020)
- news spreading (al Nashmi et al., 2017; Al-Rawi, 2019)
- recommendation system: filter bubbles & radicalization (Heuer et al., 2021; Tomlein et al., 2021)
- toxicity of comments (Obadimu et al., 2019)

- Baumgartner et al. (2020)[1]:
  community governance
  extremism
  disinformation
  health science
- Proferes et al. (2021)[2]:
  mental health
  moderation
  depression
  anonymity
  gender
  eating disorders

[1] Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The Pushshift Reddit Dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*(1), 830-839. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/7347

[2] Proferes, N et al. (2021). Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics. *Social Media + Society*, *7*(2)

gesis Leibniz-Institut für Sozialwissenschaften

Meet the Experts

How does the API access compare?

- ✓ academic access: submit form
- ✓ limits: 10M Tweets/month
- ✓ well documented, lots of resources and examples
- ✓ fairly complete access to the data
  - ✓ likes, followers, retweets
  - ✓ full-text search

- ✓ easy to get access: Google account
- ✓ limits: 10K units per day ~ 10K videos (direct access) or 5000 videos (search)
- ✓ well documented, lots of resources and examples
- ✓ fairly complete access to the data:
  - ✓ comments, (dis)likes, favorites, views, subscribers
  - ✓ searches by terms, channels, categories

- ✓ easy to get access: Reddit account
- ✓ limits: 60 requests / min -> 2.6M / month
- ✓ space for improvement: **pushshift.io**[1]
- ✓ everything is accessible but not so easy to navigate
- ✓ but Reddit is open regarding the use of the content, e.g., compared to Twitter and deletion policies

pushshift.io

- ✓ collecting live data (when is posted) since 2015
- ✓ and a retroactive dataset that goes back to 2005
- ✓ even easier to get access! No need of an account
- ✓ full-text search
- ✓ easy to use, well organized
- ✓ possible to find removed and moderated posts and comments
- ✓ raw data files accessible to download
- ✓ free (donation based)

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The Pushshift Reddit Dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*(1), 830-839. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/7347

Other cases

Wikipedia

- Not just the most widely used encyclopedia, but a social network of collaboration:
  - >120K active editors per month (English Wikipedia)
  - Wikipedia page talks (where discussions about article revisions happen)
  - record of all revisions
  - sources (references) that support content[1]

- A myriad of APIs and tools associated to it, e.g.:
  - statistics: https://xtools.wmflabs.org
  - knowledge base: https://www.wikidata.org
  - classification systems: https://ores.wikimedia.org
  - tracking changes: https://www.wikiwho.net[2]

- Relevance of Wikipedia as a corpus for machine learning (NLP) systems

- A free API (or direct download of dumps)

[1] Zagovora, O., Ulloa, R., Weller, K., & Flöck, F. (2020) http://arxiv.org/abs/2010.03083
[2] Flöck, F., & Acosta, M. (2014).. https://doi.org/10.1145/2566486.2568026

Common Crawl

- Monthly Internet snapshot since 2013
  - Petabytes of historical web pages
- Free to use (donation based)
- Due to its size, not the most user-friendly system
- Use cases:
  - mining old content related specific topics (e.g. climate change)
  - evolution of Internet (e.g. URLs)
  - find dead links (content that is “lost”)
  - as a corpus for NLP

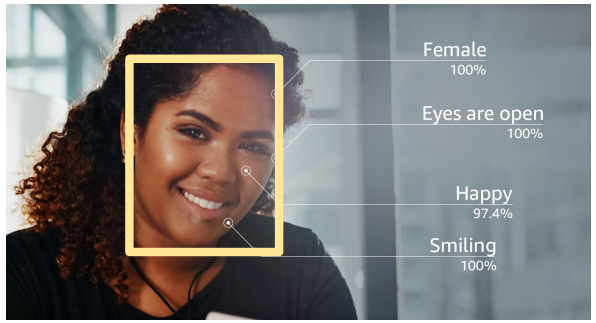
Annotation APIs

Perspective

- Scores to classify comments:
  - toxicity, insult, profanity, identity attack, threat, sexually explicit
  - different languages: English, Spanish, French, German, Portuguese, Italian and Russian
- Free to use: 1 query per second, this can be increased if requested
- You should always evaluate the performance
- By Jigsaw LLC, under Google management

Amazon Rekognition

- Tag images and videos:
  - general labels (objects)
  - text detection
  - face analysis (e.g. happy, eyes are open)
  - celebrity recognition



Source: https://aws.amazon.com/rekognition

- Offers coordinates, so one can calculate area, e.g. face-to-body ratio
- 5000 requests per month for one year (AWS Free Tier):
  - After that, 1$ per 1000 images
- It could be biased, evaluate the labels
- Easy to visually evaluate the results, as opposed to manually tagging pictures

- AWS provides other types of annotations, e.g.:
  - text analysis (Comprehend)
  - OCR incl. tabulated data (Textract)
  - voice recognition (Transcribe)

- Other services that can help annotating text, voice and images:
  - Google Cloud Services (https://cloud.google.com/products/ai)
  - Azure Cognitive Services (https://azure.microsoft.com/en-us/services/cognitive-services)
  - IBM Watson (https://www.ibm.com/cloud)
  - Clarifai (https://www.clarifai.com)

- Domain (URL) classification:
  - Amazon Alexa (https://www.alexa.com)
  - Webshrinker (https://www.webshrinker.com)
  - Klazify (https://www.klazify.com)

Many more APIs to access data…

- Search Engine APIs (Google, Bing)
- Governmental data (abgeordnetenwatch.de, data.gov, data.gov.uk, open-data.europa.eu)
- International agencies: UN, WHO, the World Bank
- News organizations: BBC, The New York Times, The Guardian, NPR, USA Today and ZEIT Online
- Scholarly archives and journals: arXiv, PLoS, Mendeley
- Metadata of data: Dryad (https://datadryad.org/api/v2/docs/), Figshare (https://docs.figshare.com/)
- Music: Spotify, Soundcloud
- …

Takeaways

- Lots of possibilities to explore different research questions
- The platform that you choose for your research matters
- APIs were not meant for researchers to access data
- APIs offer access to machine learning models:
  - If you have a boring annotation task, look for an API. Chances are you will find an API for it.
- Evaluate the annotations

gesis Leibniz-Institut für Sozialwissenschaften

Meet the Experts

Collaborators

Colleagues at the CSS department

Dr. Mattia Samory

PhD Candidate Indira Sen

PhD Candidate Olga Zagovora

Maria Zens

Dr. Johannes Breuer

Thank you !

gesis Leibniz-Institut für Sozialwissenschaften

Leibniz Gemeinschaft

gesis Leibniz-Institut für Sozialwissenschaften

Meet the Experts

Expert Contact & GESIS Consulting

Contact: you can reach the speaker/s via e-mail:

roberto.ulloa@gesis.org

GESIS Consulting: GESIS offers individual consulting in a number of areas – including survey design & methodology, data archiving, digital behavioral data & computational social science – and across the research data cycle.

Please visit our website www.gesis.org for more detailed information on available services and terms.

More Services from GESIS

- Get materials for capacity building in computational social science and take advantage of our expanding expertise and resources in digital behavioral data.
- Use GESIS data services for finding data for secondary analysis and sharing your own data.
- Check out the GESIS blog "Growing Knowledge in the Social Sciences" for topics, methods and discussions from the GESIS cosmos – and beyond.
- Keep up with GESIS activities and subscribe to the monthly newsletter.
- 

for publications, tools & services.

gesis Leibniz-Institut für Sozialwissenschaften

Meet the Experts

More from CSS Experts in the Series

June 24 Katrin Weller: **A Short Introduction to Computational Social Science and Digital Behavioral Data**

July 01 Fabian Flöck, Indira Sen: **Digital Traces of Human Behavior from Online Platforms – Research Designs and Error Sources**

July 08 Sebastian Stier, Johannes Breuer: **Combining Survey Data and Digital Behavioral Data**

Sept 16 Oliver Watteler, Katrin Weller: **Research Ethics and Data Protection in Social Media Research**

Sept 30 Roberto Ulloa: **Introduction to Online Data Acquisition**

Oct 07 Roberto Ulloa: **Auditing Algorithms: How Platform Technologies Shape our Digital Environment**

Oct 14 Marius Sältzer, Sebastian Stier: **The German Federal Election: Social Media Data for Scientific (Re-)Use**

Nov 04 Arnim Bleier: **Introduction to Text Mining**

Nov 11 Haiko Lietz: **Social Network Analysis with Digital Behavioral Data**

Dec 2 Olga Zagovora, Katrin Weller: **Altmetrics: Analyzing Academic Communications from Social Media Data**

Dec 16 Andreas Schmitz: **Online Dating: Data Types and Analytical Approaches**

Jan 13 Gizem Bacaksizlar: **Political Behavior and Influence in Online Networks**

Jan 27 David Brodesser: **SocioHub – A Collaboration Platform for the Social Sciences**