DIGITAL DEMOGRAPHY: ANALYZING WEB AND SOCIAL MEDIA DATA

WORKING WITH WEB-APIS

EDSD NOVEMBER 2022 BARCELONA

TOM THEILE

LAB OF DIGITAL AND
COMPUTATIONAL DEMOGRAPHY

DISADVANTAGES OF SCRAPING

- Websites are finicky and prone to change
- CSS-selectors that work today might not work tomorrow
- Webpages are made to be read by humans, not R-scripts

You are not alone if you experienced problems while webscraping.

APIS

APIs!

→ Application Programming Interfaces

→ Interfaces that are meant to be used by machines

APIS

APIs!

→ Application Programming Interfaces

→ Interfaces that are meant to be used by machines

There are different classes of APIs:

- Operating System APIs (e.g.: write a file to a folder)
- R-package APIs (e.g.: read a dataframe from a csv-file with read.csv())
- Web-API
- == webpage for machines

WEB-APIS

Web APIs are "websites for machines". They provide machine-readable access to to data and services.

While websites are designed to be read by humans, APIs are designed to be read by machines.

That means that the answers of APIs don't need to be formatted in a fancy style (like a flashy website), but have to follow predictable rules.

FROM SCRAPING TO WEB-APIS

Webpages are for Humans, APIs are for computers

Web APIs use (almost) the same protocols

APIs are structured and machine readable

RESTful APIs - most common nowadays

GraphQL - can make complex queries, newer, not as common

APIs can response with any data type, but most often: JSON

RESTFUL WEB-APIS

- Uniform interface
- Client-server decoupling
- Statelessness
- Cacheability
- Layered system architecture

FROM SCRAPING TO WEB-APIS

APIs can be accessed through the browser:

https://v2.jokeapi.dev/joke/Any

You can pass variables to the server (just like r-functions):

https://v2.jokeapi.dev/joke/Any?lang=de&type=single&amount=5

https://v2.jokeapi.dev/joke/Any?lang=es&type=twopart&amount=2

FROM SCRAPING TO WEB-APIS

https://open-meteo.com/en

https://archive-api.open-meteo.com/v1/era5?latitude=41.39&longitude=2.16&start_date=1985-10-24&end_date=1985-10-26&daily=temperature_2m_max,temperature_2m_min&timezone=auto

https://docs.openalex.org/about-the-data/work

https://api.openalex.org/authors?filter=display_name.search:tom+theilehttps://api.openalex.org/works?filter=author.id:A773951722

WEB-APIS

Please open
03_02_using_simple_web_APIs.R

WEB-APIS

Simple APIs, test them in your browser:

- https://dog-facts-api.herokuapp.com/api/v1/resources/dogs?number=3
- https://docs.openalex.org/about-the-data/work
  https://api.openalex.org/authors?filter=display_name.search:tom+theile
- https://api.openalex.org/works?filter=author.id:A773951722 as webpage

THANK YOU FOR YOUR ATTENTION!

Tom Theile

Research Software Engineer

theile@demogr.mpg.de