Utilizing APIs for Humanities Research

Examples from The New York Times and The Guardian

Vincent J. Schober 13.12.2023
Overview

• API in General
• API Call/Response
• Data Structure and Conversion
• APIs in academia – why?
• New York Times API
• Guardian API
Prerequisites to Working with APIs

- Some understanding of data structures necessary
- Some programming interest/experience is helpful
What is an API?

• Application Programming Interface

• => basically a communication interface
  • With a specific set of rules
The „Call“

• Communicating with an API is done through an API call or request

• The basic building blocs:
  • URL
  • Authorization
  • Parameters

• Example:


• Tools: postman
The „Response“
The „Response“

• JSON – Java Script Object Notation
  • objects and arrays

• Needs to be transformed into a tabular format

• You need to understand the data!

• Data Science Tools:
  • Python/Jupyter/Pandas
  • R
The New York Times

- Need to make an account within the Developer Network
- Create an „app“ -> that gives you authorization
The New York Times

- Many different APIs
- Archive and Article Search API most interesting for research
• Need to register for a key as well
What do you get access to?

NYT
• Headlines
• Keywords
• Multimedia
• Abstract

The Guardian
• Headlines
• Keywords
• Full Text
What can you do with it?

• Descriptive statistics
• Interrelate with other data
The World According to Online Newspapers

**Background**
- Interest in spatial matters
- Background in geography
- Background at IfL and CRC 1199
- Knowledge about API's
- Interest in programming

**sources / data acquisition**
- The Guardian
- New York Times
- Die Zeit (?)

**data handling / coding**
- excel does not work
- handling data with python pandas
- coding data with gazetteers

**analysis**
- Descriptive statistics
- Network analysis with Gephi

**results**
- Online newspapers construct spatial entities!
- Which become thematically charged!
Figure 4 - Guardian Sections in different Editions

Figure 5 - New York Times Editions and Sections

Figure 6 - General Characteristics of the Dataset of The Guardian and the New York Times

The Guardian Database
- 1,333,611 articles
- 7,685,643 keywords
- 578,620 spatial keywords
- 1,524,630 wikidata classes

The New York Times Database
- 1,079,366 articles
- 5,485,588 keywords
- 514,270 spatial keywords
- 1,269,359 wikidata classes
The World According to Online Newspapers

Background
• Interest in spatial matters
• Background in geography
• Background at IfL and CRC
• Knowledge about API's
• Interest in programming sources/data acquisition
• The Guardian
• New York Times
• Die Zeit (?)

Data handling/coding
• excel does not work
• handling data with python pandas
• coding data with gazeteers

Analysis
• Descriptive statistics
• Network analysis with Gephi

Online newspapers construct spatial entities!
Which become thematically charged!

Figure 31 - New York Times - Heatmap of spatial Keyword Locations
Figure 32 - New York Times - Wikidata Classes
Other Interesting APIs

• Google Books API
• UNESCO API
• JSTOR Data for Research
• Wikipedia API
• OpenStreetMap API
• Archive.org API
Thank You!

Vincent J. Schober
vincent_j.schober@uni-leipzig.de