Application Programming Interface (API) for Web-Scraping & Text Analysis

Individual Class Info

Research Methods
Alex Alden
Spring 2020
Number of Students: 20

Summary of Module

This module introduces students to Web-Scraping & APIs for Data Collection, outlines and explores the New York Times API, presents how to use python and jupyter notebooks to make use of the New York Times Article API and collect articles relating to the "Opioid Epidemic." From here, students learn how to go from unstructured data to structured data and begin asking and answering questions with this data. Students then have a hands on opportunity to conduct text analysis on Opioid Epidemic articles collected from the New York Times API using the browser-based computational text analysis tool Lexos. The module includes a brief discussion on how computational text analysis is useful, the browser tools that can be used to do this method of analysis, and how to interpret results.

Learning Goals

- Understand the definition and purpose of an API and web-scraping.
- Understand the importance of API documentation.
- Understand the affordances and limitations of using APIs to build a corpus
- Start to understand how to use digital tools to pull out novel insights and findings from text data

Learning Objectives

- Learn about Web-Scraping & APIs for Data Collection
- Introduce and Explore the New York Times API
- Collect "Opioid Epidemic" News Coverage
- Conduct a Text Analysis of Opioid Epidemic News Coverage in Lexos

Materials for Module

Slides (Power Point)

Slides (PDF with Presentation Notes)

Jupyter Notebook (Global Warming)
Jupyter Notebook (Donald Trump)

Data

- Global Warming Data
    - New York Times Articles relating to the search term “Global Warming.” Data collected from the New York Times Article API, comprising the 10 most relevant articles per year relating to global warming from 2010 to 2019.
- Donald Trump Data
    - New York Times Articles relating to the search term “Donald Trmup.” Data collected from the New York Times Article API, comprising the 10 most relevant articles per year relating to global warming from 2010 to 2019.

Stopwords File