

Using APIs to Collect Data from Media/Social Media Sources

Greta Heng

jieheng2@illinois.edu

3/24/2021

The logo for The New York Times, featuring the words "The New York Times" in a black, serif font, centered within a white square. This square is set against a larger, light gray hexagonal background.

Goals

- Understand what is an API
- Why API is helpful for social sciences research
- How to collect data using NYT API with Python

Resources & Tools

- Python and Jupyter Notebook
- Github Repository:
<https://github.com/gretaheng/PythonAPIWorkshop>
- Library workshops:
<https://libcal.library.upenn.edu/calendar/wicshops>

What Is an API & Why It Is Used

- API: Application Programming Interface, a computing interface that defines interactions between multiple software applications or mixed hardware-software intermediaries
- Methods of accessing specific data / functionality
- Benefits:
 - Efficiency
 - Automation
 -

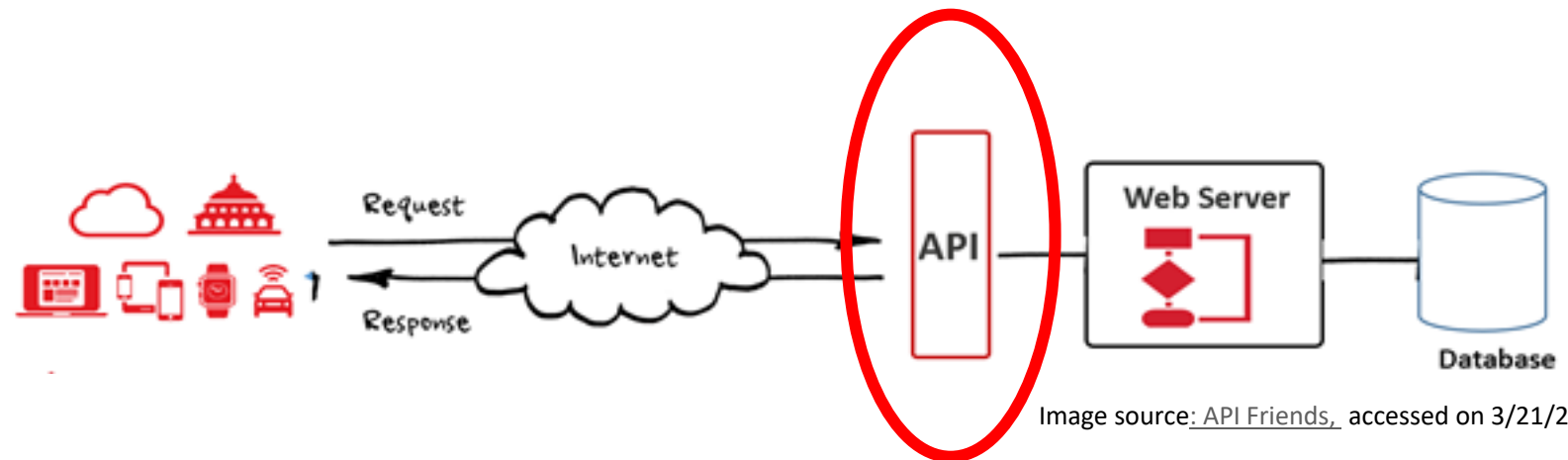


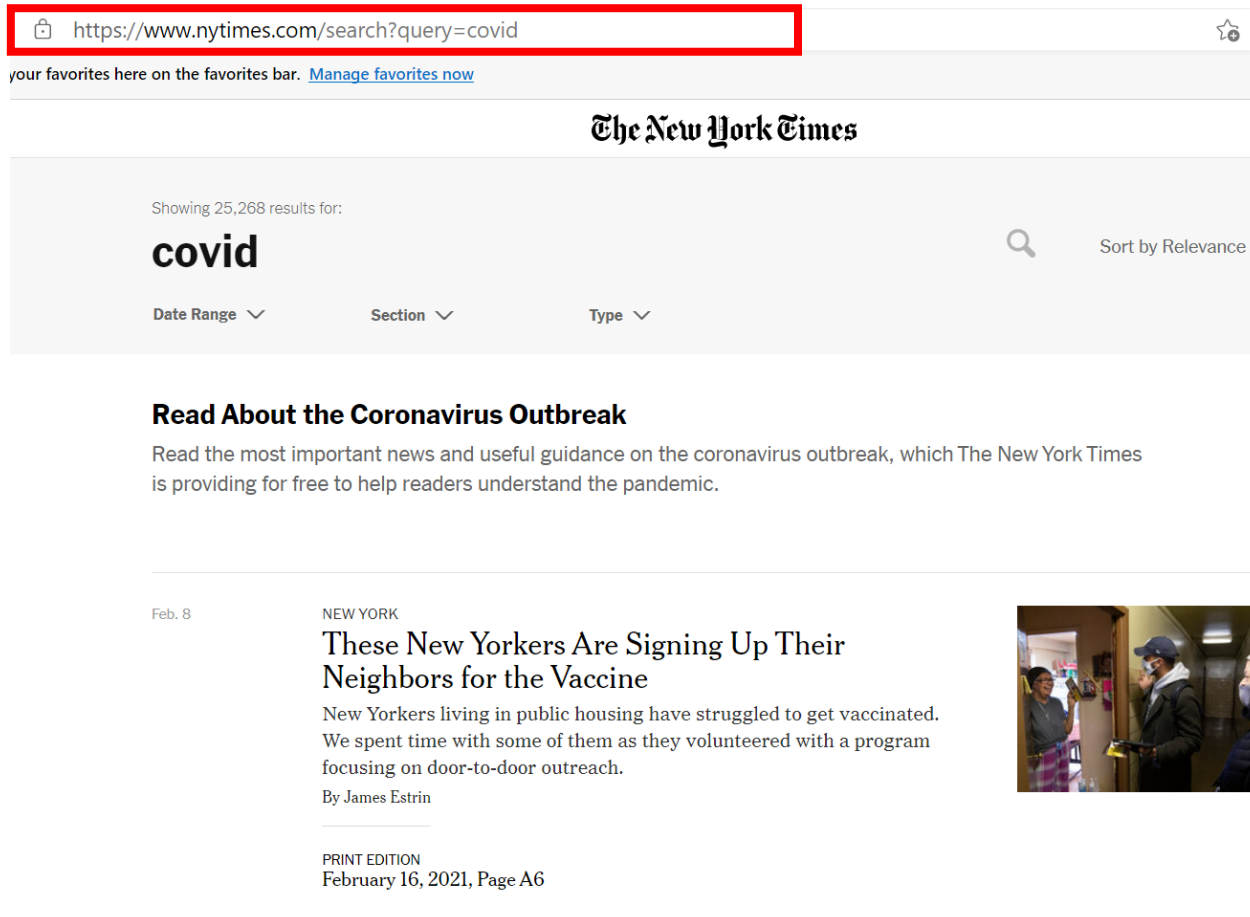
Image source: [API Friends](#), accessed on 3/21/2021

Media & Social Media APIs

- New York Times
- Bloomberg
- Guardian
- Financial Times
- YouTube
- Facebook
- Instagram
- Twitter



REST APIs



The screenshot shows a web browser with the address bar containing the URL `https://www.nytimes.com/search?query=covid`, which is highlighted with a red rectangle. Below the address bar, the text "your favorites here on the favorites bar. [Manage favorites now](#)" is visible. The main content area features the "The New York Times" logo at the top. Below the logo, it says "Showing 25,268 results for: covid". To the right of the search term is a magnifying glass icon and the text "Sort by Relevance". Below this, there are three dropdown menus labeled "Date Range", "Section", and "Type". The main article section is titled "Read About the Coronavirus Outbreak" and includes a sub-headline "Read the most important news and useful guidance on the coronavirus outbreak, which The New York Times is providing for free to help readers understand the pandemic." Below this, there is a date "Feb. 8" and a section header "NEW YORK These New Yorkers Are Signing Up Their Neighbors for the Vaccine". The article text reads: "New Yorkers living in public housing have struggled to get vaccinated. We spent time with some of them as they volunteered with a program focusing on door-to-door outreach." The author is listed as "By James Estrin". At the bottom, it says "PRINT EDITION February 16, 2021, Page A6". To the right of the text is a photograph showing two people in a hallway; one person is wearing a blue cap and a dark jacket, and the other is wearing a white shirt and a blue cap.

- REST: Representational State Transfer
- HTTP protocol and JSON data format
- Request: a URL
- Response: a result of your request in JSON format

REST APIs

url = `https://api.nytimes.com/svc/search/v2/articlesearch.json?q=COVID&api-key=[YOUR_API_KEY]`

```
{
  "status": "OK",
  "copyright": "Copyright (c) 2021 The New York Times Company. All Rights Reserved.",
  "response": {
    "docs": [
      {
        "abstract": "New Yorkers living in public housing have struggled to get vaccinated. We spent time with some of them as they volunteered with a program focusing on door-to-door outreach.",
        "web_url": "https://www.nytimes.com/interactive/2021/02/08/nyregion/covid-vaccine-public-housing.html",
        "snippet": "New Yorkers living in public housing have struggled to get vaccinated. We spent time with some of them as they volunteered with a program focusing on door-to-door outreach.",
        "lead_paragraph": "New Yorkers living in public housing have struggled to get vaccinated. We spent time with some of them as they volunteered with a program focusing on door-to-door outreach.",
        "print_section": "A",
        "print_page": "6",
        "source": "The New York Times",
        "multimedia": [
        "headline": {
```

[975 lines]

[8 lines]

- REST: Representational State Transfer
- HTTP protocol and JSON data format
- Request: a URL
- Response: a result of your request in JSON format

Request: HTTP Methods

{T} Developers

Article Search

ARTICLE SEARCH

Overview

PATHS

/articlesearch.json **GET**

GET /articlesearch.json

Search for NYT articles by keywords, filters and facets.

- GET: retrieve resource representation/information only

Request: Parameters



- Optional key-value pairs that could narrow or sort the search results
- After the question mark in the URL
- NYT Article Search API parameters:
 - begin_date
 - end_date
 - q
 -

`https://api.nytimes.com/svc/search/v2/articlesearch.json?begin_date=20200101&end_date=20200331&q=COVID&api-key=[YOUR_API_KEY]`

Request: API key & Access Token

WORKSHOP

Twitter API Workshop1

Settings **Keys and tokens**

Consumer Keys ⓘ

API key & secret

Regenerate

Authentication Tokens ⓘ

Bearer token
Generated March 21, 2021

Regenerate Revoke

Access token & secret
For @gretasweet118

Generate

- API key
 - Identification and authorization
- Access Token
 - User-specific
 - Part of authentication

Response: JSON

```
{
  "status": "OK",
  "copyright": "Copyright (c) 2021 The New York Times Company. All Rights Reserved.",
  "response": {
    "docs": [
      {
        "abstract": "New Yorkers living in public housing have struggled to get vaccinated. We spent time with some of them as they volunteered with a program focusing on door-to-door outreach.",
        "web_url": "https://www.nytimes.com/interactive/2021/02/08/nyregion/covid-vaccine-public-housing.html",
        "snippet": "New Yorkers living in public housing have struggled to get vaccinated. We spent time with some of them as they volunteered with a program focusing on door-to-door outreach.",
        "lead_paragraph": "New Yorkers living in public housing have struggled to get vaccinated. We spent time with some of them as they volunteered with a program focusing on door-to-door outreach.",
        "print_section": "A",
        "print_page": "6",
        "source": "The New York Times",
        "multimedia": [
        "headline": {
```

[975 lines]
[8 lines]

- JSON: JavaScript Object Notation
- A data interchange format
- Name-value pairs
 - "status": "OK"
 - "response": {}
 - "multimedia": []

Let's Have Some Fun

- Download and open the Jupyter Notebook
(<https://github.com/gretaheng/PythonAPIWorkshop>)

Thank you!

Greta Heng
jieheng2@illinois.edu
3/24/2021