Using APIs to Collect Data from Media/Social Media Sources

Greta Heng

jieheng2@illinois.edu

3/24/2021

Overview

- Introduction to APIs
  - What Is an API & Why It Is Used
  - Media & Social Media APIs
  - REST APIs
  - Request
    - HTTP Methods
    - Parameters
    - API key & Access Token
  - Response
    - JSON Format
- Collect data using APIs

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:
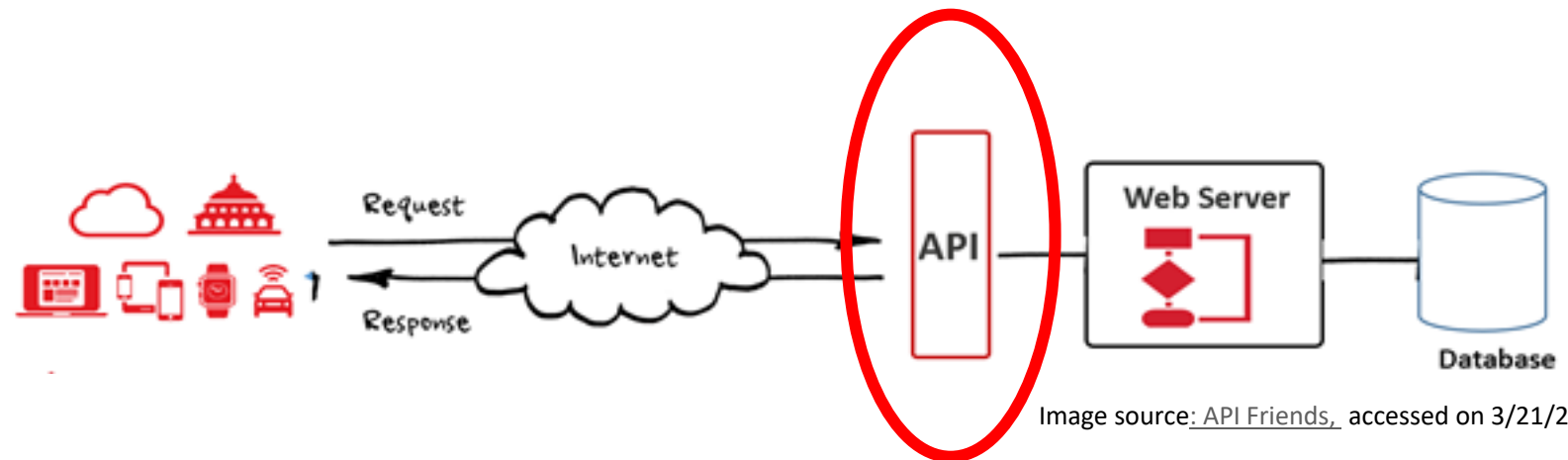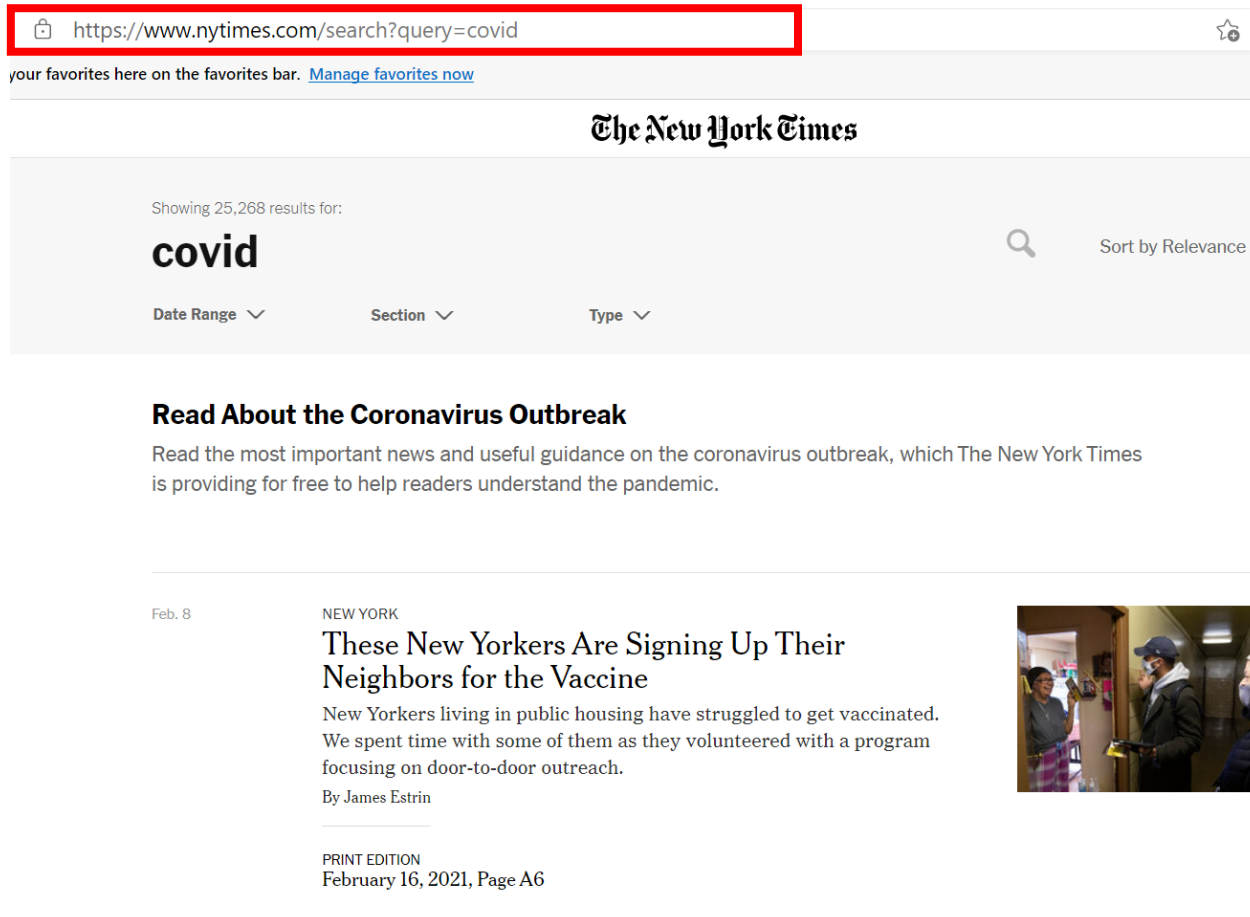 - Efficiency
  - Automation
  - ……

Image source: API Friends, accessed on 3/21/2021

Media & Social Media APIs

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

Media & Social Media APIs

Topic Modeling

Semantic Network Analysis

REST APIs

- 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": [                                                   [975 lines]
                "headline": {                                                       [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

- GET: retrieve resource representation/information only

Request: Parameters

Showing 25,269 results for:

COVID

Date Range ∨   Section ∨   Type ∨

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

- 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
  - ……

Request: API key & Access Token

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

Response: JSON

```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": [                                    [975 lines]
                "headline": {                                       [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