

# Lecture 10

## Web APIs

Pierre Biscaye  
Université Clermont Auvergne

Data Science for Economics

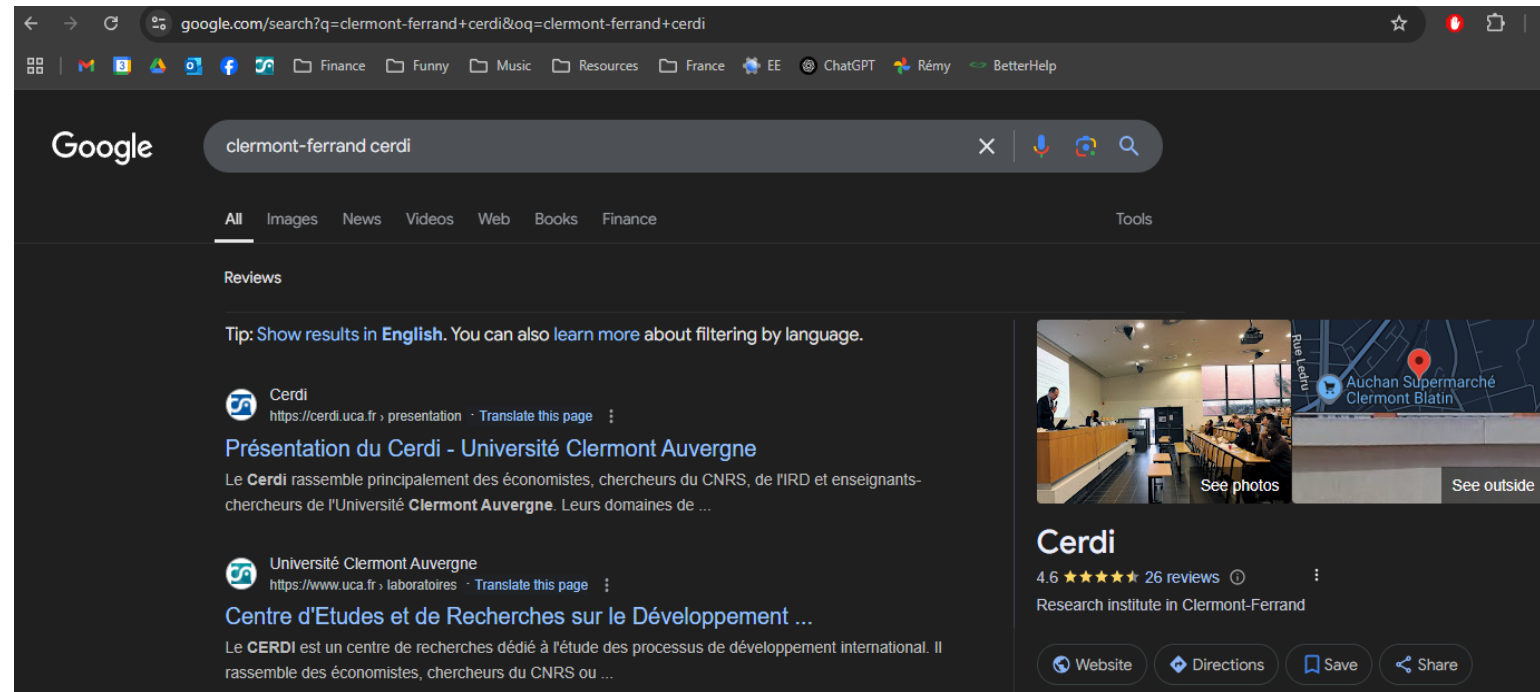
Note: Materials for this lecture are drawn from the UC Berkeley D-Lab's Python Web APIs course.

# What is an API?

- "API" stands for Application Programming Interface.
- Broadly defined, an API is a set of rules and procedures that facilitate interactions between computers and their applications.
- A very common type of API is the **Web API**, which, among other things, allows users to query a remote database over the internet.
  - For example, a web service such as Reddit has many databases that may be of use to us: Posts, Users, Subreddits, etc.
  - If we want to access some portion of these databases, it'd be helpful to have a set of rules and protocols in place to outline how we access this information. This is the motivation for an API.

# RESTful Web APIs

- Web APIs take on a variety of formats, but the vast majority adhere to a particular style known as Representational State Transfer or REST.
- What makes these "RESTful" APIs so convenient is that we can use them to query databases using URLs.
- Using RESTful APIs is essentially formatting URLs so that you can get the response you want.



# Terminology

- **Uniform Resource Locator (URL):** a string of characters that, when interpreted via the Hypertext Transfer Protocol (HTTP), points to a data resource, notably files written in Hypertext Markup Language (HTML) or a subset of a database. This is often referred to as a "**call**".
- HTTP Methods/Verbs:
  - GET: requests a representation of a data resource corresponding to a particular URL. The process of executing the GET method is often referred to as a "**GET request**" and is the main method used for querying RESTful databases.
  - HEAD, POST, PUT, DELETE: other common methods, though mostly never used for database querying.
- Surfing the web is basically equivalent to sending a bunch of GET requests to different servers and asking for different files written in HTML.

# API or web scraping?

- When deciding between using an API or web scraping, you should consider both the method's legality and efficiency.
- APIs provide structured, authorized access to data, often with clear documentation and rate limits to manage server load.
- Web scraping, on the other hand, involves extracting data directly from web pages.
  - This may violate a site's terms of service or lead to challenges in navigating complex page structures.
- While scraping can be useful when no API is available, APIs are generally the preferred method for accessing web data due to their reliability and compliance with legal standards.

# API Examples

- [Reddit](#): Used for pulling Reddit data, posting status updates, and more.
- [Spotify](#): Access to rich song data data such as valence, energy, and danceability metrics.
- [Watson IBM Natural Language Inference API](#): Use state of the art NLP models to analyze text sentiment, extract named entities, and classify text.
- [New York Times](#): Access to a treasure trove of data about their articles that is easily accessible and available for free.

# The New York Times API

- This course will use the NYT API as a case study for learning how to use web APIs.
- For most APIs, a key or other user credentials are required for any database querying. Generally, this requires that you register with the organization.
- Most APIs are set up for developers, so you'll likely be asked to register an "application".
  - All this really entails is coming up with a name for your project, and providing your real name, organization, and email.
  - Note that some more popular APIs (e.g. Twitter, Facebook) will require additional information, such as a web address or mobile number.
- We'll now get set up with API keys so that we can make some API calls to the NYT servers.

# Getting NYT API Access

1. Go to the [NYT Developer Page](#) and create an account
  1. Follow the required steps to register
2. Login with your new username and password

## Get Started

Get started using the APIs in four steps:

### Create an Account

If you don't already have an account create one:

1. Click [Sign In](#).
2. Click [Create account](#) and follow prompts to register.

### Sign In

To sign in to the portal:

1. Click [Sign In](#).
2. Enter your email address and password.
3. Click [Sign In](#).

### Register apps

To register an app:

1. Select **My Apps** from the user drop-down.
2. Click **+ New App** to create a new app.
3. Enter a name and description for the app in the New App dialog.
4. Click **Create**.
5. Click the **APIs** tab.
6. Click the access toggle to enable or disable access to an API product from the app.

### Access the API keys

1. Select **My Apps** from the user drop-down.
2. Click the app in the list.
3. View the API key on the **App Details** tab.
4. Confirm that the status of the API key is Approved.

### APIs

The [APIs](#) page has information on the different APIs. The documentation for each API includes an interactive reference for trying out the API.



# Getting NYT API Access

3. Click on your email in the top right corner and you'll see a dropdown menu that says Apps. Click on Apps and then click on the + New App button.
4. You'll see the page where you'll be prompted to add a name for your App. You can call it anything.
5. Then click enable on the APIs that are enabled in the screenshot on the next slide. You can enable them all but make sure you at least enable the ones on the screenshot.

# NYT APIs to select

## Web Scraping Course

### APIs \*

| Name               | Description   | Status    | Actions                  |
|--------------------|---|-----------|--------------------------|
| Archive API        | Get all NYT article metadata for a given month.                               | ✓ Enabled | <button>Disable</button> |
| Article Search API | Search for New York Times articles.   | ✓ Enabled | <button>Disable</button> |
| Books API          | Get NYT Best Sellers Lists and lookup book reviews.                           | —         | <button>Enable</button>  |
| Most Popular API   | Popular articles on NYTimes.com.  | ✓ Enabled | <button>Disable</button> |
| Movie Reviews API  | Search for movie reviews. DEPRECATED  | ✓ Enabled | <button>Disable</button> |
| RSS Feeds          | NYT RSS section feeds.  | —         | <button>Enable</button>  |
| Semantic API       | Get semantic terms (people, places, organizations, and locations). DEPRECATED | ✓ Enabled | <button>Disable</button> |
| Times Tags API     | NYT controlled vocabulary. DEPRECATED   | ✓ Enabled | <button>Disable</button> |
| Times Wire API     | Real-time feed of NYT article publishes.                                      | —         | <button>Enable</button>  |
| Top Stories API    | Get articles currently on a section front or the home page.                   | ✓ Enabled | <button>Disable</button> |

# Your API key

6. You'll see an API key next to your App ID. Have that key ready to copy into the python notebook.

## Overview

App Name \*

Web Scraping Workshop

Description

App ID

df2ec1c4-33e2-46ab-ab9a-5927a6b9f043

## API Keys

Key

Secret

Status

Created

Expires

Actions

**YOUR KEY IS HERE**

Show secret



✓ Active

Jan 26, 2022, 8:59 PM

never

Revoke

ADD KEY

# API limits

- When registering for any web API, you will be assigned one or more keys, tokens, or other credentials that must be supplied to the server as part of any API call you make.
- To make sure that users aren't abusing their data access privileges (e.g. by making many rapid queries), each set of keys will be given several **rate limits** governing the total number of calls that can be made over certain intervals of time.
- For the NYT Article API, we have relatively generous rate limits: 10 calls per minute and 4,000 calls per day.