



APIs: Application Programming Interface

D-Lab Python Web Data Workshop



What is an API

- API stands for Application Programming Interface
- A set of rules and procedures that facilitate interactions between computers and their applications

Web APIs

- Allows users to query a remote database over the internet
- Take on a variety of formats
- Majority adhere to a particular style known as Representational State Transfer or REST
- "RESTful" APIs are convenient because we can use them to query databases using URLs



RESTful Web APIs are All Around You...

Consider a simple Google search.

Go ahead and search something.

Ever wonder what all that extra stuff in the address bar was all about?

RESTful Web APIs are All Around You...

It looks like Google makes its query by taking the search terms, separating each of them with a "+", and appending them to the link:

<https://www.google.com/#q=>

So that we have

<https://www.google.com/#q=search1+search2>

So can change our Google search by adding some terms to the URL.

Some Basic Terminology: URL

- Uniform Resource Location
- 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.

Some Basic Terminology: 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.

How Do GET Requests Work? A Web Browsing Example

- Surfing the Web = Making a bunch of GET Requests
- For instance, I open my web browser and type in `http://www.wikipedia.org`. Once I hit return, I'd see a webpage.
- Several different processes occurred, however, between me hitting "return" and the page finally being rendered.

RESTful Database Querying: The Response

- Unlike web browsing, the content of the server's response that is extracted by Curl is unlikely to be HTML code.
- Will likely be raw text response that can be parsed into one of a few file formats commonly used for data storage.
-
- Usual suspects include .csv, .xml, and .json files.



RESTful Database Querying: The Formatting

- Web browser parsed the HTML code
- But we need R, Python, or other programming languages to parse the server response and convert it into a format for local storage (e.g. matrices, dataframes, databases, lists, etc.)
- Most APIs have a Python/R package that makes this process easier.

API Examples

- [Twitter](#)
 - Used for pulling/streaming twitter data, posting status updates, and more. Check out their [academic API](#).
- [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.

The New York Times API



- For our hands-on workshop we'll be using the New York Times' API to pull and analyze data from their articles.
- But first we all need to sign up for an API key which can be easily done at this link on [their page](#).