Ch 2. Interacting with APIs to import data from the web

In this chapter, you will gain a deeper understanding of how to import data from the web. You will learn the basics of extracting data from APIs, gain insight on the importance of APIs, and practice extracting data by diving into the OMDB and Library of Congress APIs.

1. Introduction to APIs and JSONs

1.1 Pop quiz: What exactly is a JSON?

Which of the following is NOT true of the JSON file format?

Possible Answers

1. JSONs consist of key-value pairs.
2. JSONs are human-readable.
3. The JSON file format arose out of a growing need for real-time server-to-browser communication.
4. The function json.load() will load the JSON into Python as a list.
5. The function json.load() will load the JSON into Python as a dictionary.

Answer: 4

1.2 Loading and exploring a JSON

Now that you know what a JSON is, you'll load one into your Python environment and explore it yourself. Here, you'll load the JSON `'a_movie.json'` into the variable `json_data`, which will be a dictionary. You'll then explore the JSON contents by printing the key-value pairs of `json_data` to the shell.

Instructions

- Load the JSON `'a_movie.json'` into the variable `json_data` within the context provided by the `with` statement. To do so, use the function `json.load()` within the context manager.
- Use a for loop to print all key-value pairs in the dictionary `json_data`. Recall that you can access a value in a dictionary using the syntax: dictionary [`key`] .

```python
import json
# Load JSON: json_data
with open("a_movie.json") as json_file:
    json_data = json.load(json_file)

# Print each key-value pair in json_data
for k in json_data.keys():
    print(k + ': ', json_data[k])
```

```
Title:  The Social Network
Year:  2010
Rated:  PG-13
Released:  01 Oct 2010
Runtime:  120 min
Genre:  Biography, Drama
Director:  David Fincher
Writer:  Aaron Sorkin (screenplay), Ben Mezrich (book)
Actors:  Jesse Eisenberg, Rooney Mara, Bryan Barter, Dustin Fitzsimons
Plot:  As Harvard student Mark Zuckerberg creates the social networking site
that would become known as Facebook, he is sued by the twins who claimed he s
tole their idea, and by the co-founder who was later squeezed out of the busi
ness.
Language:  English, French
Country:  USA
Awards:  Won 3 Oscars. Another 165 wins & 168 nominations.
Poster:  https://m.media-amazon.com/images/M/MV5BOGUyZDUxZjEtMmIzMC00MzlmLTg4
MGItZWJmMzBhZjE0Mjc1XkEyXkFqcGdeQXVyMTMxODk2OTU@._V1_SX300.jpg
Ratings:  [{'Source': 'Internet Movie Database', 'Value': '7.7/10'}, {'Sourc
e': 'Rotten Tomatoes', 'Value': '96%'}, {'Source': 'Metacritic', 'Value': '9
5/100'}]
Metascore:  95
imdbRating:  7.7
imdbVotes:  576,495
imdbID:  tt1285016
Type:  movie
DVD:  11 Jan 2011
BoxOffice:  $96,400,000
Production:  Columbia Pictures
Website:  N/A
Response:  True
```

1.3 Pop quiz: Exploring your JSON

Load the JSON 'a_movie.json' into a variable, which will be a dictionary. Do so by copying, pasting and executing the following code in the IPython Shell:

```
import json with open("a_movie.json") as json_file: json_data = json.load(json_file)
```

Print the values corresponding to the keys 'Title' and 'Year' and answer the following question about the movie that the JSON describes:

Which of the following statements is true of the movie in question?

Instructions

1. The title is 'Kung Fu Panda' and the year is 2010.
2. The title is 'Kung Fu Panda' and the year is 2008.
3. The title is 'The Social Network' and the year is 2010.
4. The title is 'The Social Network' and the year is 2008.

Answer: 3

APIs and interacting with the world wide web

2.1 Pop quiz: What's an API?

Which of the following statements about APIs is NOT true?

Answer the question

1. An API is a set of protocols and routines for building and interacting with software applications.
2. API is an acronym and is short for Application Program interface.
3. It is common to pull data from APIs in the JSON file format.
4. All APIs transmit data only in the JSON file format.
5. An API is a bunch of code that allows two software programs to communicate with each other.

Answer: 4

2.2 API requests

Now it's your turn to pull some movie data down from the Open Movie Database (OMDB) using their API. The movie you'll query the API about is The Social Network. Recall that, in the video, to query the API about the movie Hackers, Hugo's query string was `'http://www.omdbapi.com/?t=hackers'` and had a single argument `t=hackers`.

Note: recently, OMDB has changed their API: you now also have to specify an API key. This means you'll have to add another argument to the URL: `apikey=72bc447a`.

Instructions

- Import the requests package.
- Assign to the variable url the URL of interest in order to query `'http://www.omdbapi.com'` for the data corresponding to the movie The Social Network. The query string should have two arguments: `apikey=72bc447a` and `t=the+social+network`. You can combine them as follows: `apikey=72bc447a&t=the+social+network`.
- Print the text of the reponse object `r` by using its `text` attribute and passing the result to the `print()` function

```
In [2]:  # Import requests package
         import requests

         # Assign URL to variable: url
         url = 'http://www.omdbapi.com/?apikey=9cd6fcc8&t=social+network'

         # Package the request, send the request and catch the response: r
         r = requests.get(url)

         # Print the text of the response
         print(r.text)
```

{"Title":"The Social Network","Year":"2010","Rated":"PG-13","Released":"01 Oct 2010","Runtime":"120 min","Genre":"Biography, Drama","Director":"David Fincher","Writer":"Aaron Sorkin (screenplay), Ben Mezrich (book)","Actors":"Jesse Eisenberg, Rooney Mara, Bryan Barter, Dustin Fitzsimons","Plot":"As Harvard student Mark Zuckerberg creates the social networking site that would become known as Facebook, he is sued by the twins who claimed he stole their idea, and by the co-founder who was later squeezed out of the business.","Language":"English, French","Country":"USA","Awards":"Won 3 Oscars. Another 165 wins & 168 nominations.","Poster":"https://m.media-amazon.com/images/M/MV5BOGUyZDUxZjEtMmIzMC00MzlmLTg4MGItZWJmMzBhZjE0Mjc1XkEyXkFqcGdeQXVyMTMxODk2OTU@._V1_SX300.jpg","Ratings":[{"Source":"Internet Movie Database","Value":"7.7/10"},{"Source":"Rotten Tomatoes","Value":"96%"},{"Source":"Metacritic","Value":"95/100"}],"Metascore":"95","imdbRating":"7.7","imdbVotes":"576,495","imdbID":"tt1285016","Type":"movie","DVD":"11 Jan 2011","BoxOffice":"$96,400,000","Production":"Columbia Pictures","Website":"N/A","Response":"True"}

2.3 JSON–from the web to Python

Wow, congrats! You've just queried your first API programmatically in Python and printed the text of the response to the shell. However, as you know, your response is actually a JSON, so you can do one step better and decode the JSON. You can then print the key-value pairs of the resulting dictionary. That's what you're going to do now!

Instructions

- Pass the variable url to the `requests.get()` function in order to send the relevant request and catch the response, assigning the resultant response message to the variable `r` .
- Apply the `json()` method to the response object r and store the resulting dictionary in the variable `json_data` .
- Hit `Submit Answer` to print the key-value pairs of the dictionary `json_data` to the shell.

```
In [3]: # Import package
        import requests

        # Assign URL to variable: url
        url = 'http://www.omdbapi.com/?apikey=72bc447a&t=social+network'

        # Package the request, send the request and catch the response: r
        r = requests.get(url)

        # Decode the JSON data into a dictionary: json_data
        json_data = r.json()

        # Print each key-value pair in json_data
        for k in json_data.keys():
            print(k + ': ', json_data[k])
```

```
Title:  The Social Network
Year:  2010
Rated:  PG-13
Released:  01 Oct 2010
Runtime:  120 min
Genre:  Biography, Drama
Director:  David Fincher
Writer:  Aaron Sorkin (screenplay), Ben Mezrich (book)
Actors:  Jesse Eisenberg, Rooney Mara, Bryan Barter, Dustin Fitzsimons
Plot:  As Harvard student Mark Zuckerberg creates the social networking site
that would become known as Facebook, he is sued by the twins who claimed he s
tole their idea, and by the co-founder who was later squeezed out of the busi
ness.
Language:  English, French
Country:  USA
Awards:  Won 3 Oscars. Another 165 wins & 168 nominations.
Poster:  https://m.media-amazon.com/images/M/MV5BOGUyZDUxZjEtMmIzMC00MzlmLTg4
MGItZWJmMzBhZjE0Mjc1XkEyXkFqcGdeQXVyMTMxODk2OTU@._V1_SX300.jpg
Ratings:  [{'Source': 'Internet Movie Database', 'Value': '7.7/10'}, {'Sourc
e': 'Rotten Tomatoes', 'Value': '96%'}, {'Source': 'Metacritic', 'Value': '9
5/100'}]
Metascore:  95
imdbRating:  7.7
imdbVotes:  576,495
imdbID:  tt1285016
Type:  movie
DVD:  11 Jan 2011
BoxOffice:  $96,400,000
Production:  Columbia Pictures
Website:  N/A
Response:  True
```

2.4 Checking out the Wikipedia API

You're doing so well and having so much fun that we're going to throw one more API at you: the Wikipedia API (documented here). You'll figure out how to find and extract information from the Wikipedia page for Pizza. What gets a bit wild here is that your query will return nested JSONs, that is, JSONs with JSONs, but Python can handle that because it will translate them into dictionaries within dictionaries.

The URL that requests the relevant query from the Wikipedia API is

> https://en.wikipedia.org/w/api.php?
> action=query&prop=extracts&format=json&exintro=&titles=pizza
> (https://en.wikipedia.org/w/api.php?
> action=query&prop=extracts&format=json&exintro=&titles=pizza)

Instructions

- Assign the relevant URL to the variable `url`.
- Apply the `json()` method to the response object `r` and store the resulting dictionary in the variable `json_data`.
- The variable `pizza_extract` holds the HTML of an extract from Wikipedia's Pizza page as a string; use the function `print()` to print this string to the shell.

```
In [4]:  # Import package
         import requests

         # Assign URL to variable: url

         url = 'https://en.wikipedia.org/w/api.php?action=query&prop=extracts&format=js
         on&exintro=&titles=pizza'

         # Package the request, send the request and catch the response: r
         r = requests.get(url)

         # Decode the JSON data into a dictionary:
         json_data = r.json()

         # Print the Wikipedia page extract
         pizza_extract = json_data['query']['pages']['24768']['extract']
         print(pizza_extract)
```

```
<p class="mw-empty-elt">
</p>

<p><b>Pizza</b> (<small>Italian: </small><span title="Representation in the I
nternational Phonetic Alphabet (IPA)">[ˈpittsa]</span>, <small>Neapolitan: </
small><span title="Representation in the International Phonetic Alphabet (IP
A)">[ˈpittsə]</span>) is a savory dish of Italian origin, consisting of a usu
ally round, flattened base of leavened wheat-based dough topped with tomatoe
s, cheese, and various other ingredients (anchovies, olives, meat, etc.) bake
d at a high temperature, traditionally in a wood-fired oven. A small pizza is
sometimes called a pizzetta.
</p><p>In Italy, pizza served in formal settings, such as at a restaurant, is
presented unsliced and eaten with the use of a knife and fork. In casual sett
ings it is cut into wedges to be eaten while held in the hand.
</p><p>The term <i>pizza</i> was first recorded in the 10th century in a Lati
n manuscript from the Southern Italian town of Gaeta in Lazio, on the border
with Campania. Modern pizza was invented in Naples, and the dish and its vari
ants have since become popular in many countries. It has become one of the mo
st popular foods in the world and a common fast food item in Europe and North
America, available at pizzerias (restaurants specializing in pizza),  restaur
ants offering Mediterranean cuisine, and via pizza delivery. Many companies s
ell ready-baked frozen pizzas to be reheated in an ordinary home oven.
</p><p>The <i>Associazione Verace Pizza Napoletana</i> (lit. True Neapolitan
Pizza Association) is a non-profit organization founded in 1984 with headquar
ters in Naples that aims to promote traditional Neapolitan pizza. In 2009, up
on Italy's request, Neapolitan pizza was registered with the European Union a
s a Traditional Speciality Guaranteed dish, and in 2017 the art of its making
was included on UNESCO's list of intangible cultural heritage.</p>
```