DS4UX: How to get data with Python

[HCDE598] RESTful APIs and data crawling

Johnathan Morgan & Sungsoo (Ray) Hong

We are making many questions while we conduct our work, and research.
For example,
How many of people are suffered from starving in food deserts?
How much of carbon dioxide has been emitted from vehicles in Washington state last year?
How the consumption of red meat related to global warming?
Who will win in the presidential election in 2017?

How many, How much, How frequent, Which party will, How A and B is related things : these types of questions are important for understanding the current status of the world.
Past - Intuition based. Build the hypothesis and find out if the hypothesis is correct, from interview. – make new data by themselves. This is valid thing, and I'm also doing this approach.
"Fermi problem. Fermi estimation".
Sample are not enough, and sometimes result is different from the reality.
Not shared. Small scale. Not precise.

Now, you can capture the current status of the world more accurately by relying on

Big Picture

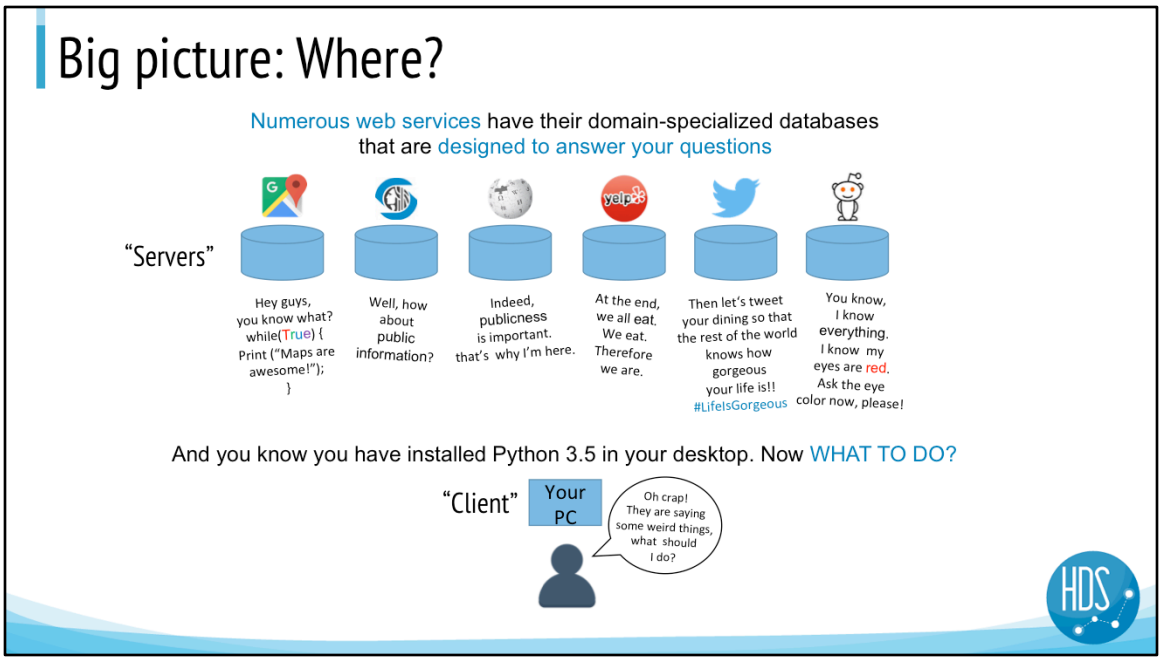
In the big picture, we see the high-level concept of how you get the data.

And will find out what type of things are needed to learn.

Big picture

Data,
from where to find
and how to get?

?



-. How many of you are familiar with server-client model?

Here's the big picture of the world and you.

-. Servers, or several web services ?

-. Explain the services

Then

-. Explain you
-. Variables (e.g., list, and dictionary), for and while loop, if and elif, else.

Big picture: Where?

These "Numerous web services" include:

- Shared knowledge and public data (e.g., Wikipedia, Data.seattle.gov, Reddit)
- Map services (e.g., Google Maps, Yelp, OpenStreetMap)
- Multimedia (e.g., YouTube, Spotify)
- Social networks (e.g., Flicker, Twitter, Facebook, Instagram)
- e-Commerce (e.g., Amazon, eBay)
- Reviews (e.g., foursquare, Rotten Tomatoes)

And more. Find more at http://www.programmableweb.com/apis/directory

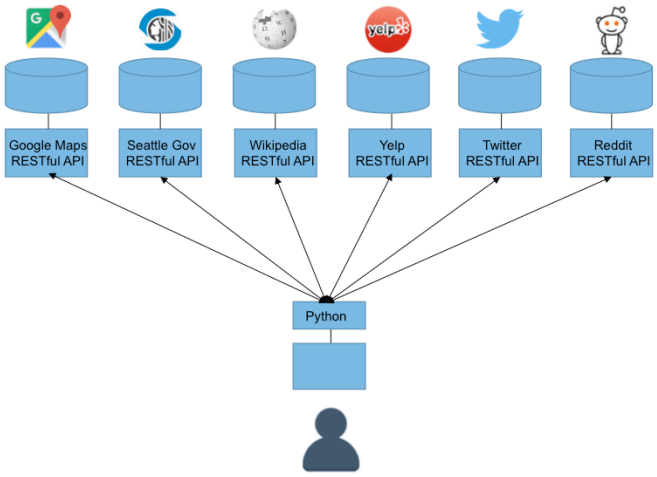
NBA

Big picture: Then how?

Several web services provide what is called RESTful APIs* as a resource that you can access and conduct some data crawling **.

* REST – Representational State Transfer: https://en.wikipedia.org/wiki/Representational_state_transfer
**Web crawler: https://en.wikipedia.org/wiki/Web_crawler

Big picture: Then how?



Big picture: Step by Step

1) You construct a query which specifies what you want to know.
2) You cast the query to a target RESTful API.
3) Then the API will answer your query by
sending a file to you which is encoded in
some standardized format (e.g., JSON).
4) You access to the information by
decode a file you received from the API, and save it.
+
5) You can even periodically cast queries to REST API and let the Python save them, so that you can analyze the information later.
(i.e., data crawling).

Big picture: Things to learn today

1) What type of queries can we can make and
: will cover this tonight.
2) How to actually cast to REST APIs ?
: will cover this tonight.
3) How to decode the standardized data format you got from RESTful APIs and how to save it, so that you can access to the information?
: will cover this tonight

Big picture: Things to learn today

4) How can you periodically cast queries to RESTful APIs and systematically store information?
: not today :p

IMAGE CREDIT: WWW.QUICKMEME.COM & HBO

Constructing a query for RESTful API with examples

Query Construction

RESTful APIs will answer your questions properly
only if
1) you construct your query based on the APIs' specification, and
2) you have permission to ask questions to the APIs.

Let's see one by one.

Query Construction

Q: What do you mean by constructing a query based on APIs' specifications?

A: This means that you construct the query by specifying necessary parameters that are specified in the API that you are planning to use. The parameters should be documented in APIs.

Let's see some examples.

Query Construction

But before we go, wait. Parameters?

Let's quickly take a look at the link below:

https://courses.cs.washington.edu/courses/cse154/13sp/lectures/slides/lecture09-forms.shtml#slide2

Generally, the web service receives a set of variables made from
1) name of parameter and 2) the actual value of parameter.
e.g., http://www.google.com/maps?q=MIT means
the google map API receives the parameter name q (i.e., query) and the paired parameter value "MIT".
If you don't specify the q parameter name with value, the API will not know what to find.

Query Construction

#1. The case of Twitter

Query Construction

https://dev.twitter.com/
https://dev.twitter.com/overview/documentation
https://dev.twitter.com/rest/public
https://dev.twitter.com/rest/reference/get/search/tweets
https://dev.twitter.com/rest/tools/console

Query Construction

#2. The case of Yelp

Query Construction

https://www.yelp.com/developers/documentation/v2/overview
https://www.yelp.com/developers/documentation/v2/authentication
https://www.yelp.com/developers/manage_api_keys
https://www.yelp.com/developers/documentation/v2/search_api
https://www.yelp.com/developers/documentation/v2/business
https://www.yelp.com/developers/api_console

Query Construction

Q: How the APIs can check I have permission to ask questions?

A: It varies. But in general, you have limited number of queries within a certain amount of time window.
For example, Twitter allows you to ask 180 search queries within 15 minutes (see here). Yelp allows you to throw 25,000 queries per day.

So you have to be wise when you decide which parameters (e.g., query term) to throw and when.

Plus, the APIs check this via an authentication protocol called Oauth.

HOW to Cast A Query to APIs with Python?

Casting a Query with Python

#1. Placekitten APIs

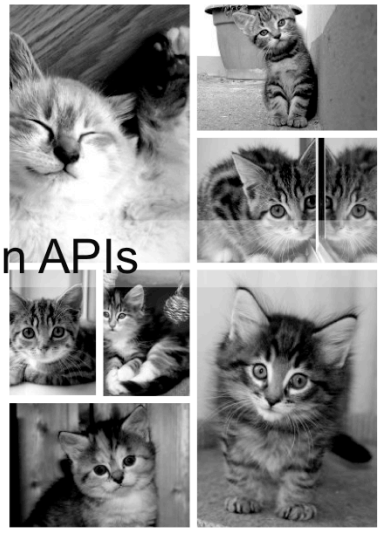
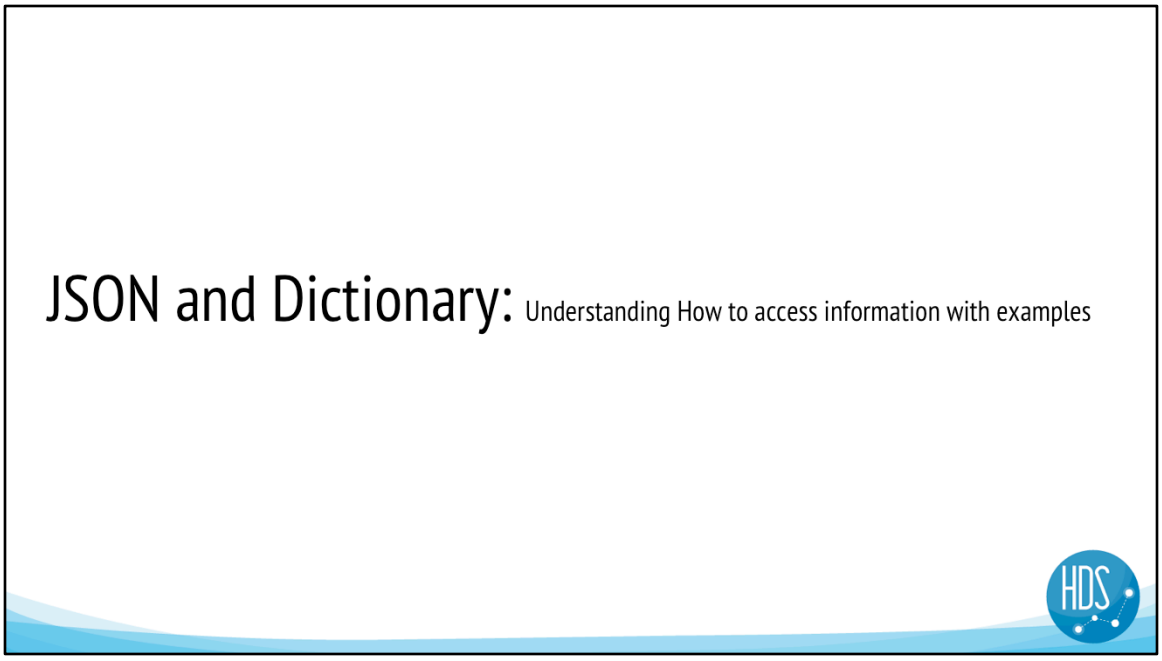


IMAGE CREDIT: Placekitten APIs

JSON and Dictionary: Understanding How to access information with examples



JSON and Dictionary

Every information can be structured.
Let's see how we can structure e-mail, for example.

Subject: Tomorrow's event~!
From: Alice Smith (alice@example.com)
To: Robert Jones (roberto@example.com) ; Charles Dodd (cdodd@example.com)

Hey guys, don't forget to call me this weekend.

Sent : 22nd of April, 2016
Private message: Yes

JSON and Dictionary

Introduce JSON object:
the way in which you can use to structuralize every type of information.

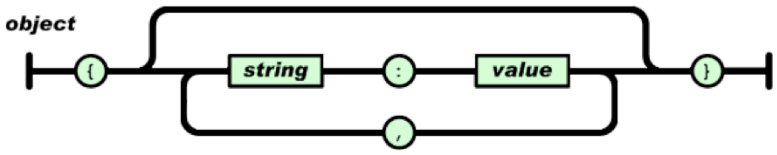


IMAGE CREDIT: http://www.json.org/

Review and
http://ruliweb.daum.net/game/search/game_map.daum?key=0

JSON and Dictionary

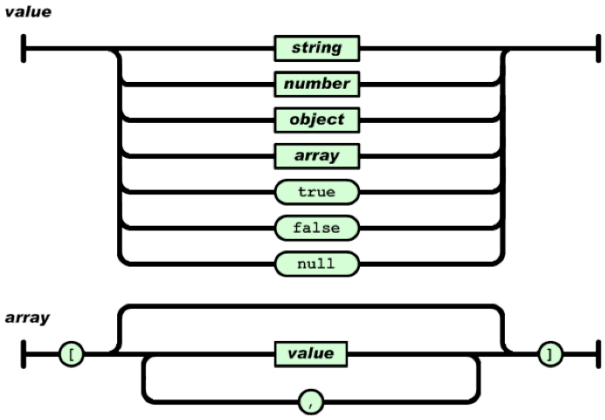


IMAGE CREDIT: http://www.json.org/

Review and
http://ruliweb.daum.net/game/search/game_map.daum?key=0

JSON and Dictionary

Subject: Tomorrow's event~!
From: Alice Smith (alice@example.com)
To: Robert Jones (roberto@example.com) ; Charles Dodd (cdodd@example.com)

Hey guys, don't forget to call me this weekend.

Sent : 22nd of April, 2016
Private message: Yes

JSON and Dictionary

Then how can we structuralize the previous e-mail with JSON object?

```
{
    "subject": "Tomorrow's event~!“,
    "from": "Alice Smith (alice@example.com)",
    "to": [
        "Robert Jones (roberto@example.com)",
        "Charles Dodd (cdodd@example.com)"
    ],
    "message": {
        "text": "Hey guys, don't forget to call me this
weekend",
        "language": "english"
    },
    "private": "true"
}
```

JSON and Dictionary

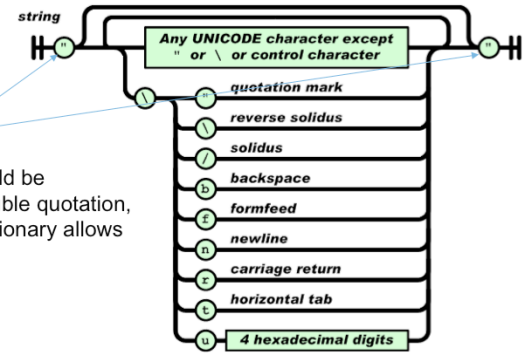
This seems very^100 familiar. Isn't it?

"JSON is more useful as a Python dictionary."

- Prof. David McDonald, UW HCDE

JSON and Dictionary

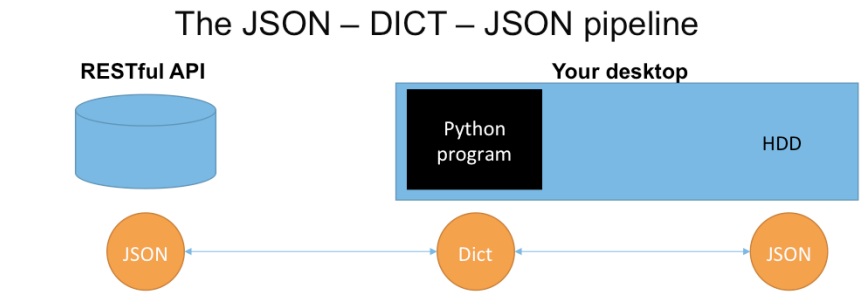
JSON and Python dictionary seems very similar,
but not exactly the same thing. (visit http://jsonlint.com/)



For example,
The JSON string should be always starts from double quotation, whereas a Python dictionary allows single quotation.

IMAGE CREDIT: http://www.json.org/

JSON and Dictionary



Stage 1. RESTful API provides you a JSON

Stage 2. *Request package* automatically coverts a JSON file to a dictionary to load that JSON to your python program. (You can do some process after a JSON is converted to a dictionary)

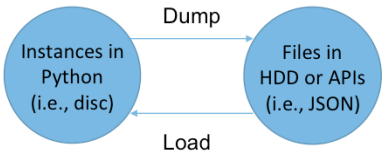
Stage 3. After you finish process a dictionary, you save a dictionary as a JSON again in your HDD. This will allow you to load that again later.

JSON and Dictionary

Thus, you need to know
How to convert a JSON file to a dictionary, and
convert a dictionary to a JSON.

JSON and Dictionary

Fortunately, this is easy.
Just remember dump and load



JSON and Dictionary

When you load

```
import json
Jsdict = {}
Fin = open("some_input.json", "r")
Jsdict = json.load(Fin)
```

And when you dump

```
import json
Jsdict = {}
…
Fout = open("some_output.json", "w")
json.dump(Jsdict, Fout, indent = "4")
Fout.close()
```

JSON and Dictionary

Demonstration:
Open jsonexample.py

Next things to learn

A Couple of Items will be discussed in next week.
(e.g., how to find a locations with Yelp)

Knowing some new RESTful APIs takes some time. So try to find your own things and learn:

http://www.programmableweb.com/apis/directory
http://www.computersciencezone.org/50-most-useful-apis-for-developers/
https://www.quora.com/What-are-some-cool-fun-APIs
https://www.reddit.com/r/webdev/comments/3wrswc/what_are_some_fun_apis_to_play_with/

Thanks!