

Data Science

Lab. Sheet 6 – Accessing Data through API's

In this lab you will learn how to access online data through unauthenticated and authenticated API's. Please refer to the 'Getting Data' notes online for detailed instructions in the case of the unauthenticated API and the authenticated Twitter API.

Exercise 1 – Unauthenticated API - Github

Following the Getting Data notes implement the necessary code to gather details of the software repositories on Github for Joel Grus as returned through the following endpoint

<https://api.github.com/users/joelgrus/repo>

You will need to gather the data and then convert from JSON structure to a python dictionary for analysis.

Implement the basic code provided to gather the data and then extend the code to

- a) Print the id's the repositories and the language
- b) Print the size, the number of watchers and the avatar_url of the owner of the repository.
- c) Analyse the created dates of the repositories to find the number of repositories created in each calendar month and also the number created on each different day to see if there is a trend.

Exercise 2 – Other Unauthenticated API's

You are required to search online for two other examples of unauthenticated API's and to develop python implementations which gather data through the API and restructure in the form of a dictionary for analysis.

Exercise 3 – Authenticated API - Ebay

Step 1:

To begin with you should first register with the ebay developer program at

<https://developer.ebay.com/signin>

You will then receive a keyset which includes an AppID to use in your code.

Step 2: You then need to install the ebaysdk package which provides python calls to the ebay API. You install by running from the command line

pip install ebaysdk

Information on this package is available at

<https://github.com/timotheus/ebaysdk-python>

Step 3: Using the demo code provided at the link above find the results of a search using the search term “Legos” and store this in a dictionary. Note that the returned data is not in the form of a json string in this case so you will simply need to use “response.dict()” to convert to a dictionary and use the ‘searchResult’ element of the dictionary to show the results.

Step 4: Explore the structure of the data dictionary and then implement the code to print out the ID of the item and the price. Use a loop to count the items printed out.

Step 5: Explore the structure of the data dictionary and implement the code to print out the total number of items available to be returned by the search. By exploring the ‘finditemsadvanced’ ebaysdk call online implement code to extract into an array the prices of the first 1000 items returned through the “Legos” keyword search. Find the mean, median and standard deviation of this data.

Step 6: Implement code to find the mean, median and standard deviation of the available items to purchase on ebay with exactly the title “Lego Star Wars Han Solo Mudtrooper 40300 Polybag”

Exercise 4 – Authenticated API - Twitter

Following the final section of the notes, register with Twitter and implement the code to gather and parse data.