

Crawl iTunes and Google Play

Part 1 (Due February 17, 2020)

Introduction:

The aim of this project is to build a dataset from (Google Play & Apple Store). You need to extract the below information and store it in CSV file. Then, filter those App to find the best match of provided list of APIs. The API list can be found in *APIsName.CSV*

<https://apps.apple.com/>

<https://play.google.com/>

Data requirements:

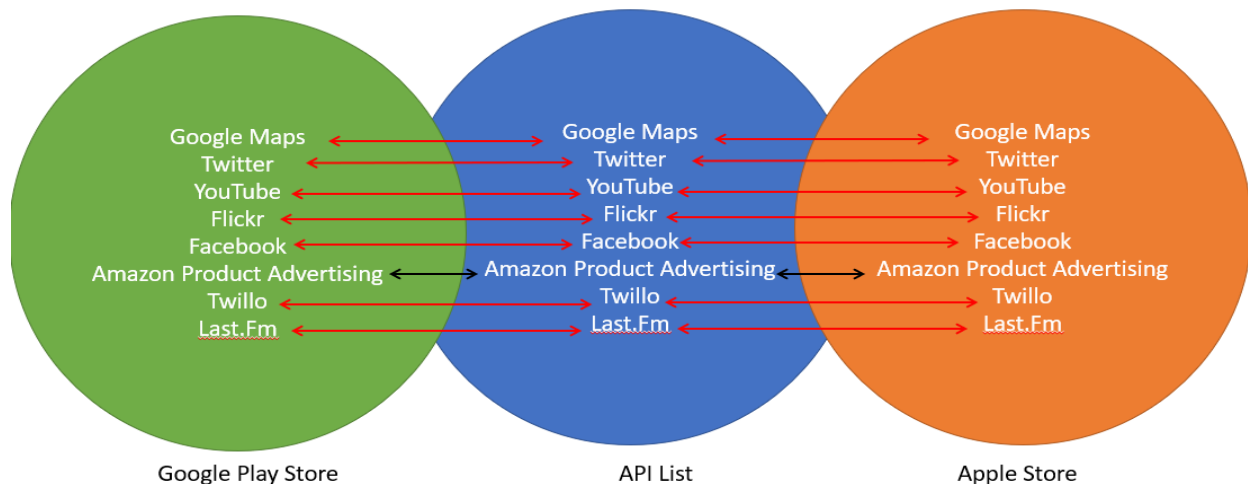
- 1- App ID: App's ID.
- 2- App Name: App name.
- 3- Review number: Number of user reviews.
- 4- Description: App description.
- 5- Rating value: User rating value.
- 6- Price: Price of app.
- 7- Version: Version number of app.
- 8- Downloads: Download numbers for app.
- 9- Author: Name of publisher

The first generated files would include the above information. And you should save this file in CSV format e.g. google-play.csv and app-apple.csv

Part 2 (Due March 19, 2020)

App and API matching:

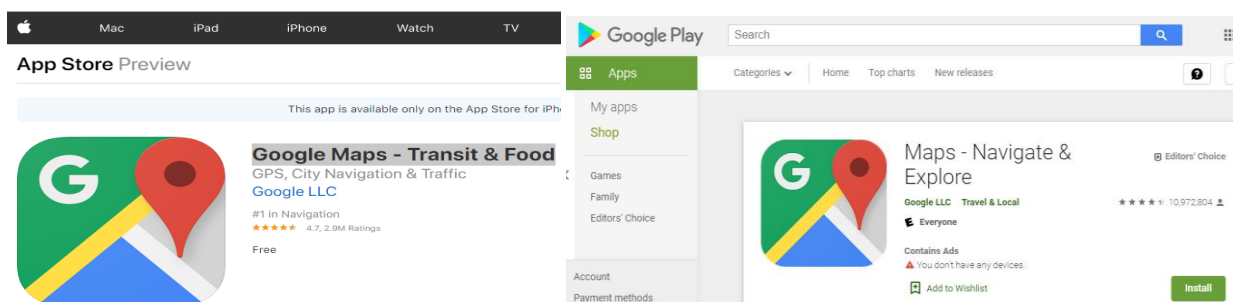
In this task, after collecting Apps information you need to match those Apps with their API's names by using *Fuzzy Matching technique* with best similarity threshold. However, some of those App and API may need manual matching. You may use the publisher name to find best match.



For example: Google Maps API, in App Store the name of the API is *Google Maps – Transit & Food* while in Google Play its name is *Map -Navigate & Explore*.

As a result:

Both of those names refer to one API which is *Google Maps API*



Useful links:

<https://nycdatasience.com/blog/student-works/web-scraping-apples-app-store/>

<https://github.com/pr3sh/AppStore>

<https://github.com/manojps/google-play-apps-crawler-scrapy>

Please feel free to contact me for any clarification:

Phone: 786 556 4167

Email: Faisalb@umich.edu