Course1 : Foundation of information

Assignment 1: Data extraction and analysis from social media platform Youtube by generating Youtube API key (30 Marks)

Problem statement

Videos are a fast growing medium where people communicate, share knowledge, showcase skills etc. YouTube is one of the biggest platforms which hosts videos. The YouTube platform hosts content from many different professions/arts/ cultures across the world.

People can express their opinion about the video in the form of likes, dislikes, comments which are features provided by the YouTube platform which provides the information on the sentiment about the video.

The assignment involves the steps on programmatic data extraction from YouTube on which analysis can be conducted to understand various attributes related to a video.

Steps to be performed

1) Connect to the Youtube API using a Python client (5 Marks)

a) Create a YouTube API key (3 marks)
b) Install the Google API python client (2 marks)

Reference links :
https://developers.google.com/youtube/v3/getting-started

https://developers.google.com/youtube/v3/quickstart/python

2) Search and extract the data (10 Marks)

a) Search videos related to the query string “avatar movie” .

(For this part, choose/search **one video of your choice** and perform data collection steps on that specific video) (3 marks)

Output expected : ID, Snippet, Channel ID, Video Description, Channel Title, Video Title Reference links : https://developers.google.com/youtube/v3/docs/search/list

b) Provide the following statistics **for query string "avatar movie"** of **top 50 videos** sorted by **relevance in the US region (regionCode)** (7 marks)

Output expected: video ID, title, no of views, no of likes,no of comments exported to CSV file

Reference links: https://developers.google.com/youtube/v3/docs/videos/list

3) Analyze the exported data obtained in 2.b and carry out the following tasks (15 marks)

a) Sort the data **2.b** by top 10 comments in descending order and consider the video IDs and Titles of top 10 videos which have highest comments. (3mark)

b) Use a suitable method to retrieve comments of those top 10 videos from **3.a.** For doing this, write a program to loop through each video id from 3.a and pass in the part parameter set to "snippet", to retrieve basic details about the comments. Execute this request and print the response using the pprint() method. (6 marks)

Note: pprint() will print out the response from the API in a more human-readable format.

Reference link: https://developers.google.com/youtube/v3/docs

Output expected : Use the python library " pprint " to print the output of the program with the following properties etag, items, id , kind snippet and snippet to have the text display field which represents the comment of videos

c) Write a program to export the output of question 3.b in JSON file format and submit the file as part of the assignment (3 marks)

d) Write a function to get the likes vs views ratio of the top 10 videos obtained in 3.a with the highest comments (3 marks)

Supporting documentation:

- Creating youtube api and credentials

  https://developers.google.com/youtube/v3/getting-started

- For utilizing the functions such as channels, searching the videos , videos - to identify the video etc in python code to fetch appropriate information

  https://developers.google.com/youtube/v3/docs