

Project Proposal

Kehao Guo, Kecheng Ye, Xiaojun Min

Eustrat Zhupa

CSC 261

September 10, 2020

1. Project Name: YouTube trending video search engine

Team Name: YourSQL (Group ID: 31)

Team Member: Kehao Guo (id: 29852529),

Kecheng Ye (id: 30642683)

Xiaojun Min (id: 31389082).

2. For this project, we want our database to give us a brief but accurate snapshot of public interest by analyzing the daily trending videos on world largest video website Youtube. We hope by looking into the different trending videos during different time period, we can get a rough idea about what interested general public or what do they care about during different time. Also, it is possible for us to dig into some specific content of the video (e.g: description, tags, etc) to understand why people get interested to such topic and give this meaningful information to our target users for further use.

The reason why we don't simply use a spreadsheet to accomplish this task is that:

- The raw data is relatively huge, and by loading the whole dataset into the memory and doing some query operation on it will be extremely computationally slow.
 - The spreadsheet does not allow distributed operation like several people simultaneously altering the same spreadsheet. One way they can accomplish this task is to download the spreadsheet on each PC and try to merge the changes after the change which is extremely difficult and inefficient. Another solution will be uploading the spreadsheet onto an open online platform (Google docs) such that people can edit the same dataset in the same time. However, this approach although solve the distributed operation part, it does not necessary guarantee on the information security part.
 - The spreadsheet cannot necessarily recognize different user with different authority so everyone will be able to do any operation on the database including some destructive operation like deleting all contents in the spreadsheet. This implies that it is very hard for dataset administrators to effectively manage and protect this database.
3. Target users for this database are expected to be employees at Youtube. Specifically, the data managers at Youtube would be responsible for authorizing access to the database, for monitoring its use, acquiring software and hardware resources, controlling its use and monitoring efficiency of operation. On the other hand, data analysts, data scientists, and those at the marketing and strategy departments would be the end users. They would be able to access, retrieve and modify entries in the database concurrently for different uses.

For instance, they would be able to conduct analysis on the latest trends on Youtube in order to create more effective promoting strategies to attract more Youtube viewers.

4. The relations for the database and their corresponding schemas are listed as followed:

- General information
 - Video_id (string), Title (string), publish_time (datetime), channel (string), category_id (int)
- Statistics
 - Video_id (string), views (int), likes (int), dislikes (int), comment_count (int), trending_date (string)
- Peripheral
 - Video_id (string), thumbnail_link (string), comments_disabled (boolean), rating_disabled (boolean), video_error_or_removed (boolean)
- Text
 - Video_id (string), title (string), tag (string), description (string)

5. The following is the designed UI for the database. Users can query or make updates to the dataset via this interface.

YouTube Search Engine

[Login](#)[Signup](#)

Sign Up

Please fill in this form to create an account!

☐ I accept the Terms of Use & Privacy Policy

YouTube Search Engine

Login

Signup

Login

Username

Password

Sign in

Lost your Password ?

Don't have an account? Signup here!

YouTube Search Engine

User123

Search by

Trending Date

Tags

Key words

likes ☐ Dislikes ☐

Channel

Advanced search

Result show all

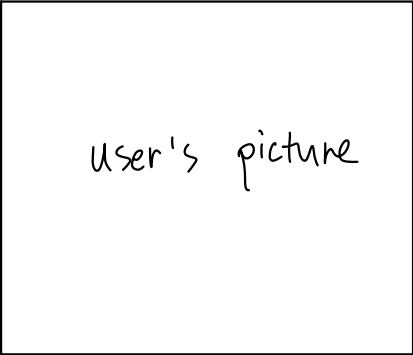
if click on "Advanced Search"

0 Comment Disabled

0 Rating Disabled

YouTube Search Engine

Me



Edit

My Favourite Video

upload / delete

≡ Sort By



x x x



x x x

YouTube Search Engine

User123

Title : x x x x x

Trending Date : / /

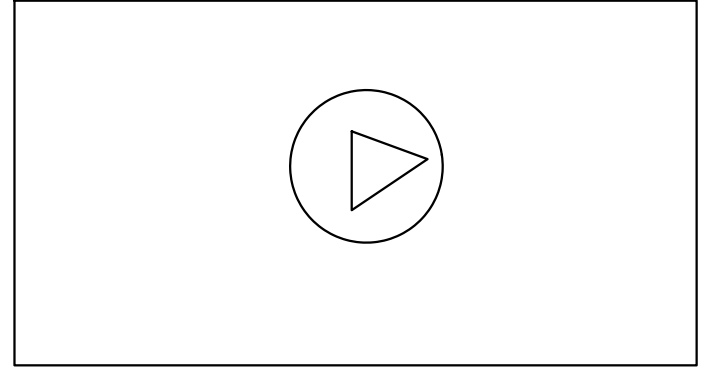
Tags :

key words :

Likes : 5361

Dislikes 119

Channel : x x x



Link: x x x x x

Add to my list

6. We planned to get data from a Kaggle project “Trending Youtube Video Statistics” (<https://www.kaggle.com/datasnaek/youtube-new?select=USvideos.csv>), where the statistics of top trending Youtube videos in US were recorded.