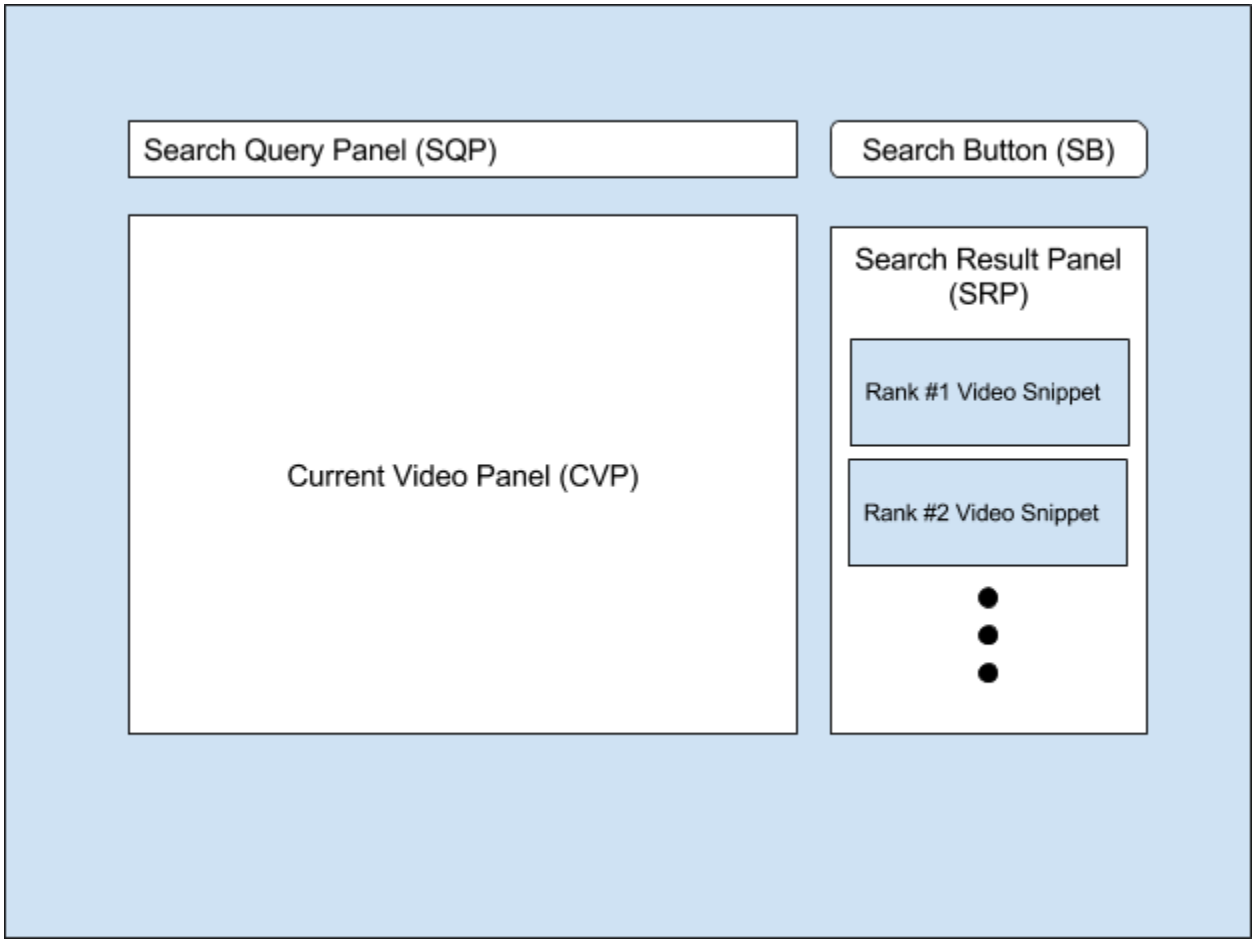
CSL4030 Data Engineering Course Project (Jul-Nov, 2023)

Problem Statement

This is a group project. The group formation details are given in the section ‘Group Formation’. Your objective is to build a **video search engine** application with a **GUI**.

The following features are mandatory:

- The **application** must be combinedly managed by all three database programs.
    - MongoDB should be used for indexing the video files. Use the video dataset shared with this file (‘dataset.tar’). Please feel free to add more video documents, e.g., you can create JSON files of your favorite videos and add them to the database.
    - Neo4j will be used for managing the relationships between the videos and entities in the videos.
    - MySQL will be used to store relational information, such as click-through information.
- The **GUI** should at least have the following components:
    - Search Query Panel (SQP) - for submitting video search queries
    - Search Button (SB)
    - Search Result Panel (SRP)
    - Current Video Panel (CVP)



- When the application is launched, all three panels should remain empty.
- When the user inputs a search query in SQP and clicks on SB, the GUI should display a list of **relevant videos** in SRP. The search query can be a single word query or a multi-word query, e.g., 'Kohli' or 'Virat Kohli'. You may choose the number of videos you would like to show in SRP (at least two are expected). Each result should be displayed in the result list with appropriate snippets, such as – the title of the video, thumbnail image, and short description.
- If a user clicks on a video snippet, the video details (description, like count, etc.) should be shown in CVP. At the same time, SRP should be updated to display the new list of **related videos** i.e. the videos related to the current video in the Neo4j database. You can choose the type(s) of relationship between the videos to improve the search results.
- The **click-through information** should be stored in the MySQL database. In a click-through log, traditionally relevant information about clicks against a search query is stored. In general, every transaction in the relation may capture information such as search query, video ID, rank of the video in the search result against the query, timestamp of the click, user ID <if available>, etc. You may additionally consider any other relevant information that you feel is important for improving the search results. Improvement includes, but not limited to, the following:
    - Inclusion or exclusion of a video in the search result
    - Change in the rank of a video in the search result

Group Formation

- Each group can have min. 3 and max. 4 members.
- A Google Form will be made available for submitting the members' details of a group. Only one member of the group needs to fill up the form.
- Group information submission deadline: **11:59 PM, November 5, 2023, Sunday**.
- If you are unable to form a group, you shall be randomly allocated to a group.
- A **Group ID** will be generated for each group and posted on Google Classroom on **November 8, 2023, Wednesday**.

Evaluation

- Code submission: An assignment will be created for code submission on Google Classroom. Please keep all the required files in one folder and zip it. Rename the zip file to '<your group ID>.zip'. Then submit the zip file through the given assignment link on or before **November 22, 2023, Wednesday**. Only one of the group members needs to submit the code.
- The evaluation session will be conducted on **Nov 22, 2023, at CC Lab 1 during 3-6 PM**. Each group will be provided a 5-minute slot to give a demo of their application. It will be followed by a 5-minute Q&A session by the teachers and other students.
- All the members of a group must be present during the evaluation. But the demo can be given by a subset of members.
- The search engine should be accessible from anywhere on the intranet through a browser. The demo needs to be given on a computer specified by the teachers, not on your laptop. During the demo, the teachers must be able to access the search engine.
- **The total marks are 20.** Each member of the same group will be awarded the same marks.
    - Database design: 5 marks
    - GUI design: 5 marks
    - Additional features: 10 marks.
- Code plagiarism will be penalized heavily.
- Please note that higher weightage is given to the **additional features** section in the marking scheme. Additional features can be any features that you would like to offer to your application users. For example, recommended videos based on the click history of a given user. Be as creative as you want to be.

Important Dates

Group information submission deadline	**11:59 PM, Nov 5, 2023, Sunday**
Generation of Group IDs	**Nov 8, 2023, Wednesday**
Code submission	**11:59 PM, Nov 22, 2023, Wednesday**
Evaluation session	**3-6 PM, Nov 22, 2023,** at CC Lab 1