TranscriptQuery: Our CS410 Project

By Adnan Noorullah, Anshul Goswami, Matt Straczek, Heet Parikh, and Vrush Patel

Roles

- Vrush - Helped with frontend implementation, helped connect web scraper to database and database to search engine, helped add display functionality to search engine frontend.
- Adnan - Frontend applications for query page and search engine page, making sure links work for each page, made sure that input fields passed data to web scraper, helped with search engine backend implementation and display, and assisted Matt with layout of web scraper.
- Heet - BackEnd applications, setting up SQL databases to perform POST requests to queries, and then parsing data from SQL to rank queries based on frequency.
- Matt - Setup the web scraper and linking the frontend and backend. Integrated the web scraper with the search engine to display the correct transcript results to the front end.
- Anshul - BackEnd applications, setting up SQL databases to perform POST requests to queries, and then parsing data from SQL to rank queries based on frequency.

Techstack

- FrontEnd:
  - React
  - Tailwind CSS
  - Daisy UI
- Backend:
  - Node.JS
  - Express
  - Python (Selenium and WebDriver for web scraper)
  - Javascript
  - SQLite

What is our project?

- YouTube Transcript Query Search Engine
- Channel Name
- Date Range
- Search Query
- Display Results

Challenges

- Getting BackEnd to connect with local SQLite Database
- Working together without getting merge conflicts
- Connecting the database properly to the search engine
- Lack of experience creating web scrapers

Future improvements (extensions)

- Creativity on the search engine page
  - YouTube thumbnails
  - More formatted results
- More functionality to scrap specific videos
- Ranking algorithm to provide more relevant results
- Date range could be exact date