Harry Vu

Professor John Mattox

CSE 385

20 November 2023

CSE 385: FINAL PROJECT DOCUMENTATION

Project Name: Movie Database Application

Overview: This project focuses on developing a graphical user interface (GUI) that enables users to filter movies from multiple streaming platforms and manage their personal watchlists. The data is maintained in a database, with SQL used for data retrieval.

1. **Source Code:**

Source code files are included in the "source_code" folder which has these 2 files:

- 'gui.py': Python source code for the GUI
- 'database.sql': SQL file to create the schema and tables for GUI functionality.

Extra:

- "data_downloaded" folder: Contains 4 CSV files with original data before SQL conversion.

2. **Documentation about the project**

a. **Reference to datasets:**

Dataset Descriptions: This dataset comprises movie information from four distinct streaming services: Amazon Prime, Disney+, Hulu, and Netflix.

- Amazon Prime Movies and TV Shows. https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows. Accessed 20 Nov. 2023.
- Disney+ Movies and TV Shows. https://www.kaggle.com/datasets/shivamb/disney-movies-and-tv-shows. Accessed 20 Nov. 2023.
- Hulu Movies and TV Shows. https://www.kaggle.com/datasets/shivamb/hulu-movies-and-tv-shows. Accessed 20 Nov. 2023.
- Netflix Movies and TV Shows. https://www.kaggle.com/datasets/shivamb/netflix-shows. Accessed 20 Nov. 2023.

b. References to external libraries:

- 'sys': Used for interacting with the Python interpreter, accessing command-line arguments, and system-specific parameters.
- 'PyQt5': Creates graphical user interfaces (GUIs) for Python applications using the Qt framework.
- 'mysql.connector': Connects Python applications to MySQL databases for executing SQL queries, data retrieval, and database operations.
- 'pycountry': Provides access to ISO standards data like country, language, and currency codes for internationalization purposes.

c. Installation instructions:

- Make sure to have MySQL server, MySQL Workbench, and python3 installed on your local machine.

- Run ‘database.sql’ in MySQL Workbench to create the schema and tables.
- Install libraries via Terminal:
    - python -m pip install --upgrade pip
    - pip install PyQt5
    - pip install mysql-connector-python
    - pip install pycountry
- Open ‘gui.py’ and modify the password for the localhost database, root user in the ‘create_db_connection()’ function to your own password that you set (Where you created schema and tables).
- Open the file 'gui.py' using your chosen Python Integrated Development Environment (IDE). Ensure that you select the appropriate Python interpreter before running the code.

d. Usage instructions:

- Feature 1 (Filter movies):
    - Function: Enables filtering movies by Title, Type, Country, Year, Rating, Director, Actor/Actress.
    - Usage: Select a streaming service or to-watch list, choose your filters, and click “Search.” The table displays matching movies. If there is no match, a message will be shown. Not all filter fields are mandatory, but you must choose a streaming service or to-watch list.
- Feature 2 (Show Cell Content):
    - Function: Displays full content of table cells with lengthy data in a pop-up window.
    - Usage: Click on a table cell to view its complete content.

- Feature 3 (Add Movie to Watchlist):
    - Function: Adds movies to a personal watchlist.
    - Usage: After filtering movies, click “Add to Watchlist” in the Action column to add your selected movie to your To-watch list. Access your watchlist by choosing “My To-watch list” in the “Select Streaming Service/Your to-watch list” dropdown menu.
- Feature 4 (Remove Movies from Watchlist):
    - Function: Removes movies from your watchlist.
    - Usage: Select “My To-watch list” in the dropdown menu of the “Select Streaming Service/Your to-watch list,” then click “Remove” in the Action column to delete a movie from your watchlist when you have finished watching the movie or do not want to watch it anymore.
- Feature 5 (Dynamic Dropdown Options):
    - By utilizing the 'SELECT DISTINCT' and 'UNION' in SQL, all the dropdown options will be populated by the data stored in the database (not prefixed), ensuring each dropdown option is relevant to the database. For example, if there are no movies in the database from the year 2000, then the dropdown options for 'Year' will not include 2000. However, if later on, one or more movies from the year 2000 are added, then 2000 will be included as an option for the 'Year' filter. The same logic applies to the 'Type' and 'Rating' filter options.
- Feature 6 (Notifications):
    - Whenever users attempt to add a movie to their watchlist or remove it, a pop-up window will notify them whether the action is successful or not. This is helpful in

terms of data consistency and makes the user feel confident that their desired movie has been correctly added.

- Feature 7 (Block User from Modifying the Table):
    - By default, users can modify cell content (just the display part, not the real database in the backend). However, this function has been disabled to maintain a higher level of data integrity.
- Feature 8 (Changing Style of Button):
    - By utilizing stylesheets (within Python), when users hover over a button, it will change the background color and more. This greatly aids in distinguishing one button from another and makes the GUI more interactive and user-friendly.

e. Additional comments:

- After obtaining the data from the source, I performed data preprocessing to enhance the quality and integrity of the data. This included moving data that was entered into incorrect columns to their correct positions and changing ratings for consistency (e.g., changing NULL to NOTRATED). These modifications were done using SQL, meaning the data in the 4 CSV files within the 'data_downloaded' folder might slightly differ from the data generated by running 'database.sql.' To ensure the GUI in 'gui.py' operates correctly, it is essential to execute 'database.sql.' The 4 CSV files can be disregarded if your goal is simply to run and test the program.
- Occasionally, the program may temporarily freeze for 2-5 seconds when processing large sets of returned data, before updating the table with this new data. This delay is a limitation related to the processing power of the machine being used.