

Célie Pierre

Sarah Lawrence

Reihaneh Maarefdoust

COS 457 Database Systems

Database Systems Project Part 3: Documentation

December 6, 2023

Introduction

Pawnee Commons is a database that both creators and fans can use as a way to quickly access information from the TV show Parks and Recreation. The TV show Parks and Recreation was picked due to the amount of data available for this particular show, specifically the availability of their transcripts. The Pawnee Commons database can be used to quickly find information about the cast and crew, characters, episodes, and transcripts.

Data File

The data files that were used to start the Pawnee Commons database were a combination of prior information and web scraping. The prior information we had was parks_and_rec_imdb.csv, parks_and_rec_episodes.csv, and a bunch of transcripts in a folder. The information was gathered from [kaggle](https://www.kaggle.com). Reihaneh created the Python code called Data Preparation for Database. This code organizes the data in the prior files to follow our ER diagram. It created the episode, transcripts, and writer&director csv files. The web scraping uses Reihaneh's organized episode (Episode - Sheet1.csv) and writer&director (Person-writer&director - Sheet1.csv) CSV files to help get the cast and character information. The big parts of the web scraping was the use of the selenium.webdriver to emulate opening the tabs. This was used to get hidden episode information. The comparison of episode titles from Reihaneh writer&director information to the web scraped information to get season number and episode number was also a big part. Finally the ID's for every person all correspond to their respective character. The web scraping uses these parts to create the person.csv, character.csv, and plays.csv.

Part 1

Part one was creating an entity relationship diagram. As a team, we were tasked with creating an ER diagram mapping out the idea of our database. We would then present the diagram to get live feedback.

For part one, we all split the workload among us. Célie did the introduction, person, and relationships slides. On the ER diagram, she did the Person and Works on. For the write-up she did the person entities and works on relationships. Reihaneh did the character and viewer slides. On the ER diagram, Reihaneh and Sarah did the Episode, Viewers, Characters, Views, and Plays entities and relationships. For the write-up, Reihaneh did the character and viewers entities and the plays and views relationships. Sarah did the episode and transcript slides. On the ER diagram Sarah and Reihaneh did the Episode, Viewers, Characters, Views, and Plays entities and relationships. For the write-up she did the episode, transcripts entities, and the Has, says relationships. Sarah then presented the presentation.

Part 2

Part two as a team were tasked with gathering scraped information and cleaning it. Next, we were tasked to write SQL Commands. The SQL Commands were based on searches we wanted users to be able to use in the database. Finally, a video explaining the SQL commands.

For part two, again we all split the workload among us. Célie wrote SQL commands. She later on made all the commands into procedures. She also wrote the `import_data_to_sql.py`. Finally, she wrote the Database Setup Instructions. Reihaneh cleaned up and organized .csv files based on ER diagram. Reihaneh also helped Célie a bit with the SQL commands and testing. She also did the explanation video for this section. Sarah wrote Python web scraping explained above. She also organized the scraped information based on the ER diagram.

Part 3

For Part Two creating a working user interface was the first task. Next we had to create three search functionalities. The first functionality being the trigger, the next being a join operation, and finally a transaction for the update operation. After that a presentation with a live demo was performed to get live feedback. As a team, once everything was done was tasked with documentation of the whole project. Next organize the codes on one platform, GitHub. Finally, the demo video shows the completed project.

Finally the workload was split one last time. Célie was the one to work with the GUI and created the front-end user interface. She also performed the live Demo. She also did the video demo. Reihaneh did two of the additional searches. The first additional search functionality she did was functionality B. This was the join functionality. The second additional search functionality she did was functionality C. This was the update functionality. Finally Reihaneh did the Sentence transformer. Sarah wrote the additional search functionality A. This was the Trigger functionality. She also did the formatting for the presentation slides.

Conclusion

Sarah's Conclusion

I learned that the processes of creating a database are far more complex than I first thought. There is a lot of thought that goes into creating a database. I learned that SQL is incredibly useful for a variety of tasks. A few examples are retrieving, inserting, updating, and so on. I also learned that scraping and cleaning information is difficult and time-consuming. I'll have to look into better methods in the future. Overall this project was helpful in expanding my knowledge and helped me gain the tools I need for making a database.

Sarah's contribution listing:

- Part 1: Slides: Episode, Transcript, and Why, ER Diagram: with Reihaneh we did Episode, Viewers, Character, Views, and Plays, Write up: Episode, Transcripts (entities) Has, says (relationship), Presented the presentation
- Part 2: Wrote python Web Scraping, Wrote code that organized scraped information with the previously collected .csv files and organized based on ER diagram
- Part 3: Additional search functionality A (Trigger), Some of the presentation slides,

Célie's Conclusion

For Part 1, I learned what makes a good database. During our brainstorming phase, we had quite a few options to consider. Some of those options worked, others did not. We landed on creating a database for a specific TV show. However, after going through the whole semester-long process, I think we may have chosen a larger data set. Our smaller data set ended

up working just fine and we were able to demonstrate most of the techniques we learned in class. During Part 1, I also learned how important a comprehensive ER diagram is to the database creation process. It helps to put things into perspective and makes implementation easier. Our ER diagram was referenced quite a bit during the creation of our tables and procedures in MySQL.

For Part 2, I was able to practice SQL commands. I was able to write commands that set up the initial schema, added all the necessary tables, created procedures to access those tables, created procedures that allowed for user input, and created more procedures with some of the advanced techniques such as joins and aggregate functions. I was also able to use Python to import our data into the MySQL database.

For Part 3, this part was quite challenging for me. I was unfamiliar with front-end design so I learned a lot with this section of the project. I tried several different options at first (Tkinter, Flask, PyQt5) but landed on PySimpleGUI which, in my opinion, has a misleading name! After much trial and error, I got the hang of it and ended up with a pretty basic user interface capable of displaying information pulled from our MySQL database.

Célie's contribution listing:

- Part 1: Slides: Introduction, Person, Relationships; ER Diagram: Person, Works on; Write up: Person, Works on Other: Organized team meetings, made various edits/updates/corrections
- Part 2: Wrote SQL commands, wrote import_data_to_sql.py, wrote Database Setup Instructions
- Part 3: Front-end user interface, additional SQL modifications, made edits to other team members code for GUI and MySQL compatibility

Reihaneh's Conclusion

In the first steps, I learned an overview of databases to design a system infrastructure more accurately and pay attention to all aspects of relationships. Sometimes a simple system can involve complicated relationships. In the second part, gathering data and cleaning up for inserting on a system based on ER makes new insights into critical roles in ER Design and this part is so time consuming for making suitable data in all aspects of the database.

In the second part, I start using AI methods. It is really interesting to work with real data and databases and concepts like sentence transformers, opensearch methods are so useful these days so I think these parts are more effective for me.

In the third part, I focus on search functionality for two topics. Evaluation is an advanced topic in MySQL. Knowing how to use the information in Mysql for analysis is valuable to me. And the transactions worked. It is really interesting to see what happened in the background and create high-quality code in all aspects. Another important part of this project is working with a team and improving skills in how to divide tasks in a project and manage time based on goals.

Reihaneh's contribution listing:

- Part 1: Slides : Character and Viewers, ER Diagram : with Sarah we did Episode, Viewers, Character, Views, and Plays, Write up: Character and Viewers (entities) Plays and Views (relationship)
- Part 2: Clean up and organized .csv files based on ER diagram, Video
- Part 3: Additional search functionality B (join), Additional search functionality C (update), AI-Sentence transformer