Project Mid Progress Report

DSCI 551 – Fall 2024

ChatDB: A SQL like chat database software based on console interaction

Team Member: Liqiang Deng (liqiangd@usc.edu).

Tech stack used for implementation

So far, I have been using Python to program, using mysql-connector-python to connect to the database, using re for matching patterns, using csv package to handle raw dataset.

The plan to implement query syntax

At present, my plan is to first check whether the SQL statement is valid through user input, and if so, run it directly to return the result. If not, use re to match the user's input pattern (such as regular expressions) to generate the corresponding query SQL, query the data, and return the display. If none of them match, it will prompt that the user's input has not been understood.

The database I choose to use

As I am a single member team, I have decided to use MySQL as the support.

Planned Implementation and Status of the project

Planed Implementation: My planned approach so far is the same as in the proposal. Complete user interaction on the console, convert user input into specific actions or SQL, and then interact with MySQL database through Python middleware code.

Status of the project: At present, the creation of the database, creation of data tables, processing of table data, and data import have been completed. The main process of program interaction has been completed. Completed displaying table data information and sample SQL to users. The next step will focus on natural language parsing of the user input, matching the user's input patterns to generate corresponding SQL and interact with the database.

Here are some simple screenshots of my current progress, not including all the achievements:

```
Enter your query (or type 'exit' to quit): example sql queries

Here are some example SQL queries with explanations:

Query: SELECT product_type, SUM(transaction_qty) FROM coffee_shop_sales GROUP BY product_type;
Explanation: This query summarizes the total quantity sold for each product type.

Query: SELECT * FROM coffee_shop_sales WHERE product_category = 'Beverages';
Explanation: This query retrieves all sales data where the product category is 'Beverages'.

Query: SELECT product_id, AVG(unit_price) FROM coffee_shop_sales GROUP BY product_id;
Explanation: This query calculates the average unit price for each product.
```

Challenges

At present, the biggest challenge I think is how to accurately judge the natural language input by users to understand their intentions. I currently plan to use a rather stupid method, which is to enumerate more models or patterns that users may input for matching, and mainly attempt to extract user action keywords to clarify their purpose, and generate SQL based on this.

Timeline and miletones

9/23 – 10/6	Complete the work of data collection, data cleaning, underlying database system, and writing the main process of the system.
10/7 – 10/18	Complete the task of outputting SQL examples, show basic information of tables and improve the integration with SQL database programs.
	Midterm progress report
10/19 – 11/3	Complete natural language pattern recognition and result output.
11/4 – 11/17	Complete the task of indirect execution of SQL statements through the program and output the results.
11/18 – 11/25	Complete exception handling and improve and check the entire program.
	Demo and Implementation
11/26 – 12/12	Final report