Project Proposal

DSCI 551 – Fall 2024

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

Team Member: Liqiang Deng (liqiangd@usc.edu), an MSCS graduate student.

Proficient in programming with Python, JavaScript, and Golang, with a basic understanding of database knowledge and proficiency in web development. As this is a single person team, MySQL database has been chosen as the main support. And as a student majoring in CS, I will do my best to complete and do the project well.

Project requirements

Explore databases and show table

After the user enters the program in the console, a welcome message is displayed. Afterwards, users can choose data, currently only MySQL is the only option available. Users can use **'show table'** to view all tables in the database, and use **'show [specific] tables'** to obtain attribute field information in the tables, as well as obtain some data examples, such as the first 5 items.

Obtain sample queries

Users can enter the example keyword to obtain some SQL query example statements, which are randomly generated according to specific patterns. And these statements can be directly input and executed in this system to obtain corresponding results. The pattern may be **group by, having, order by, where,** etc. I will support at least 5-10 basic SQL statement keywords.

1. Users who simply input **'example'** related information no patterns will receive several SQL query statements with random patterns.
2. When the user inputs **'example with [pattern]'**, several SQL query statements with specific patterns will be randomly generated.

Execute the queries

Users can directly input SQL statements for database operations in the console, and only **valid SQL** statements are supported. Illegal statements will prompt corresponding errors.

Ask questions in natural language

Design several natural language pattern statements in advance, so that when the user inputs the corresponding pattern statements, they can understand the content they want to query, and the query results are returned and displayed. For example, the pattern can be **Total <A> by <B>，Count <X> by <Y>，Average <A> by <B>，Max/Min <A> by <B>，Top N <A> by <B>，Filter <A> where <B>**. I will implement at least 5 natural language patterns.

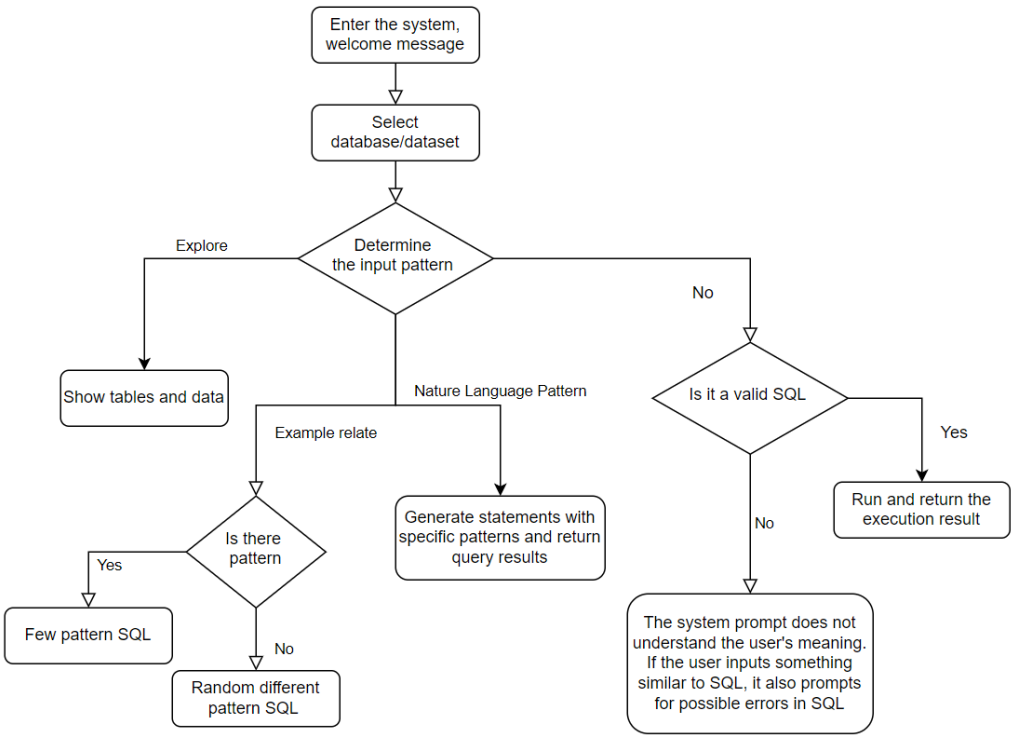
Planned Implementation

Prepare the dataset and design the database system

Choose Python as the programming language, and users can interact through the console. Programming intermediate layer code mobilizes database resources, returns and displays.
As I am a single member team, I have decided to use MySQL as the support. And I will use PyMySQL or mysql-connector-python libraries in the middle layer of the code to interact with the database. And I found 3 datasets on Kaggle as alternative references. Here are the links to the datasets:
https://www.kaggle.com/datasets/ahmedabbas757/coffee-sales
https://www.kaggle.com/datasets/meruvulikith/90000-cars-data-from-1970-to-2024
https://www.kaggle.com/datasets/ankitkr60/advertisement-and-sales-data-for-analysis
When I deal with the dataset, I may need to perform data cleaning and table partitioning operations, which will be implemented during specific operations. When specific query results need to be returned, generate corresponding SQL statements for querying after matching user input in the middle layer.

Process Control

After the user starts running the system's main.py program, the system basically follows the following process to determine the user's input and return the result. User can exit the program through exit command.



Team responsibilities

As this is a single person team, I will be responsible for all the work of the project. Including system design, data collection, data cleaning, code writing, database system implementation, language pattern matching, exception handling, project document writing, etc.

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