Proposal: Natural Language Interface (NLI) to an RDBMS

Title

ChatDB: A Natural Language Interface for Relational Databases

Team Details

Name: Yuchen Zhu
Role: Sole Developer (One-Person Group)

Team Member Background and Skills

I have a basic understanding of SQL and some experience with API integration. I am proficient in Python, which will help me implement and refine the system. While I am new to natural language interfaces (NLI), this project will be an opportunity for me to explore and develop an intuitive way for users to interact with relational databases.

Project Requirements

This project focuses on developing a natural language interface (NLI) for an RDBMS (MySQL). The system will allow users to interact with relational databases using natural language queries instead of SQL.

The core functionalities include:
1. Schema Exploration: Users can ask about database structure (e.g., available tables, attributes, sample records).
2. Query Execution: Users can input natural language queries, which will be converted to SQL statements and executed on the database.
3. Data Modification: Users can insert, update, and delete records using natural language commands.
4. Multi-Table Queries: The system will support SQL joins, aggregations, and filtering operations.
5. Integration with LLM APIs: A large language model (LLM) will be used to interpret user input and convert it into structured SQL queries.

Planned Implementation

The system will be built using the following tech stack:
-Python-based API with FastAPI or Flask
- Database: MySQL
- Natural Language Processing (NLP): OpenAI API (GPT-4)
- Query Conversion: Prompt engineering and fine-tuning for accurate SQL generation
- Security Measures: SQL injection prevention and user authentication

Workflow:
1. User Input: The user types a natural language query.

2. Processing: The query is passed to the LLM (GPT-4) to generate SQL code.
3. Execution: The SQL query runs on the RDBMS, fetching the required data.
4. Response: The results are displayed in a readable format.

Team Responsibilities

Since this is a one-person group, I will handle all aspects of the project.

Timeline (Initial Draft)

Milestone	Task	Deadline
Week 1 (Feb 1 - 7)	Finalize project scope & proposal submission	Feb 7
Week 2-3 (Feb 8 - 21)	Setup database & API integration	
Week 4-5 (Feb 22 - Mar 7)	Implement schema exploration & query translation	Midterm Report (Mar 7)
Week 6-8 (Mar 8 - Apr 4)	Add data modification & optimize accuracy	
Week 9-10 (Apr 5 - Apr 20)	Final testing & debugging	
Week 11 (Apr 21 - Apr 23)	In-Class Demo	
Week 12 (Apr 24 - May 9)	Prepare and submit final report	May 9

Dataset Requirement

I will use real-world datasets from publicly available sources (e.g., Kaggle, government open data) to demonstrate SQL query execution and multi-table joins.