MSc Project Proposal

Title: Text-to-SQL System for Project Management Applications

Student: Md Tanzeem

Supervisor: Professor Kuo-Ming Chao

Industry Partner: EZPS

Program: MSc Computing

1. Introduction

In many companies, project managers and team members need to look at project data—like progress, deadlines, or who is working on what. But this data is stored in databases and usually requires knowledge of SQL (Structured Query Language) to access. Many people in project management do not know SQL, so they must wait for IT or data analysts to help them.

This project will create a smart system that lets people type questions in everyday language (like English) and automatically translates them into SQL queries. For example, if someone types: "What tasks are due this week?", the system will convert that into a SQL query and fetch the correct data.

The goal is to make it easier for non-technical staff to get the information they need quickly, without relying on IT staff. This saves time and helps people make better decisions in their projects.

The motivation behind this project is to reduce the communication gap between business users and technical systems. With the help of modern technologies like Natural Language Processing (NLP) and AI, this can be made possible and practical.

2. Problem Statement

Many people involved in managing projects—like managers, team leaders, and business analysts—need access to project data every day. But that data is stored in databases, and only those who understand SQL can easily access it.

Let's say a project manager wants to know: "How many tasks are still not finished for Project A?" They will either:

- Have to ask someone from IT or data team to write the SQL query.

- Or they need to learn SQL themselves, which takes time.

This creates delays, miscommunication, and dependency on others.

Who is affected?

- **Project Managers**: Can't get quick answers to daily project questions.

- **Executives**: Can't get summary reports easily.

- **IT teams**: Are overloaded with report requests.

- **Clients**: Get delayed updates.

Why is this a problem?

- It slows down decision-making.

- It increases pressure on technical staff.

- It causes delays in projects.

- It adds training costs for SQL.

- It limits who can access valuable data.

With so many tools already using AI (like chatbots or voice assistants), it is time that project management systems also allow users to simply "ask a question" in plain language and get answers directly.

This project will solve the problem by creating a simple web-based interface where users can type questions about their projects. The system will understand the question using AI and convert it to SQL, run it on the database, and return the result.

3. Aims and Objectives

Main Aim:
To build a smart system that allows users to ask project-related questions in natural language and get answers from a project database using automatically generated SQL queries.

Key Objectives:

1. **Understand Project Language**: Train a model to recognize words like "task," "deadline," or "milestone" and match them to the database.

2. **Translate Text to SQL**: Use NLP models to convert plain English into SQL queries.

3. **Build a Web Interface**: Make a website where users can type their questions and view answers.

4. **Test with Users**: Try the system with real or test users and improve it based on feedback.

5. **Measure Accuracy**: Check how many times the system gives the correct answer.

Research Questions:

- How well can AI understand project management language?

- What are the best models to use for text-to-SQL conversion?

- How accurate is the system in giving correct answers?

- How easy is it for users to use the system?

Technologies to Use:

- NLP tools: Hugging Face Transformers, spaCy

- Programming: Python, SQL

- Backend: Fast API

- Frontend: React

- Database: MySQL or PostgreSQL

- Hosting: Docker

Approach:

- Collect real or fake project queries and SQL pairs.

- Train an NLP model.

- Build and connect the frontend and backend.

- Test performance and accuracy.

4. Legal, Social, Ethical and Professional Considerations

Data Privacy:
If using real project data, it may include sensitive company or personal information. The project must follow GDPR rules and make sure all data is anonymized or fake.

Ethical Use of AI:
The system should not give wrong answers. If the AI is unsure, it should notify the user instead of guessing. Users must be told about the system's limits.

Fair Access:
The system must work well for all users, regardless of how they phrase their question. It should not only work with very specific sentence formats.

Professional Responsibility:
Wrong answers from the system can lead to bad business decisions. The system must be well-tested and include warnings or error messages when needed.

Intellectual Property:
If using data or tools from a company, permission must be taken. Also, agreements must be made about who owns the final product (student, university, or company).

5. Background

The idea of converting text to SQL is not new. Researchers have been working on it since the 1970s. But older systems needed a lot of hand-coding and were limited to small domains. Now, with modern AI and NLP models, like BERT, GPT, and T5, systems can understand complex questions and generate good SQL code.

However, most existing tools are trained on general data. They don't work well in specific areas like project management where special words are used (e.g., "tasks due," "sprints," "resources assigned"). Also, real business data often spans multiple tables and involves complex relationships.

In business, tools like Microsoft Power BI and Tableau now allow basic natural language queries, but they are often limited and expensive. This project aims to make a lightweight, open, and specific solution for project management.

Innovation of this project:

- Focused on project management, not general data
- Uses latest NLP models
- Simple web interface
- Designed for non-technical users

6. References

Core Text-to-SQL and Semantic Parsing References

1. **Zhong, V., Xiong, C., & Socher, R. (2017)**
   Seq2SQL: Generating structured queries from natural language using reinforcement learning.
   arXiv preprint arXiv:1709.00103.
   ➤ Introduced a SQL generation model using reinforcement learning on WikiSQL dataset.

2. **Yu, T. et al. (2018)**
   Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-SQL task.
   Proceedings of EMNLP, 3911–3921.
   ➤ A widely used benchmark dataset for complex Text-to-SQL tasks.

Transformer Models and NLP Techniques

3. **Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019)**
   BERT: Pre-training of deep bidirectional transformers for language understanding.
   NAACL-HLT.
   ➤ Groundbreaking transformer-based language model widely used in NLP.

Schema Linking and Data Representation

4. **Guo, J., et al. (2019)**
   Towards complex text-to-SQL in cross-domain database with intermediate representation.
   ACL, 4524–4535.
   ➤ Addresses schema mapping and generation for complex query structures.

Applications in Business and Project Management

5. **Vassilakis, C., & Lepouras, G. (2003)**
   Natural language interaction in e-commerce applications.
   Decision Support Systems, 34(2), 211–226.
   ➤ Discusses early use of natural language interfaces in enterprise settings.

6. **Chaudhuri, S., & Narasayya, V. (2011)**
   Providing olap and data mining capabilities on relational databases.
   Proceedings of VLDB.
   ➤ Bridges gap between business needs and technical data access.

<table>
<tr><th colspan="4"><u>Student and First Supervisor Project Sign-Off</u></th></tr>
<tr><td></td><td>Name</td><td>Signature</td><td>Date</td></tr>
<tr><td>**STUDENT:**
I agree to complete this project:</td><td>Md Tanzeem</td><td></td><td></td></tr>
<tr><td>**SUPERVISOR:**
I approve this project proposal:</td><td>Kuo-Ming Chao</td><td></td><td></td></tr>
<tr><td>Supervisor Comments/Feedback</td><td colspan="3"></td></tr>
</table>