

选题

2.13基于 text2sql 的大模型的智能问答系统实现（4 人） 难易程度：高

- 需求简介实现一个基于数据库的智能问答系统，明确智能问答系统的目标应用领域（如特定知识领域的问答等），用户能以自然语言提问，系统借助基于 text2sql 的大模型将其转化为 SQL 查询，从特定数据库中获取答案，提高数据利用效率与便捷性，满足用户快速获取数据洞察的需求。
- 功能点
 - 自然语言理解：准确解析用户输入的自然语言问题，理解其语义、意图与涉及的关键信息，如查询目标、筛选条件、数据范围等，无论问题表述简单或复杂、正式或口语化。
 - text2sql 转换：基于大模型将理解后的自然语言问题精准转化为正确的 SQL 语句，涵盖常见的数据库操作，包括数据查询 (select)、添加 (insert)、删除 (delete)、更新 (update) 等，且能处理多表关联查询、嵌套查询等复杂情况。
 - 数据库交互：与各类主流数据库（如 MySQL、PostgreSQL、SQL Server 等）建立稳定连接，将生成的 SQL 语句安全、高效地发送至数据库执行，并获取执行结果。
 - 结果解析与呈现：对数据库返回的查询结果进行处理，以直观、易懂的格式呈现给用户，例如表格、图表、简洁文本等，方便用户快速解读与应用数据。
 - 上下文关联：支持多轮对话，在会话过程中结合上下文信息理解用户新问题，避免用户重复输入背景信息，提供连贯流畅的问答体验。
 - 错误处理与反馈：通过测试集进行测试，得到系统的问答准确性指标。

成员名单

- 组名： 软件系统我会永远实践你 🍷
- 组长： 郑涵丹(2024302111379)
- 组员(2): 熊益婕(2024302111232) 代芷涵(2024302111249)

分工方案

- 一、代芷涵：自然语言处理与 text2sql 转换模块
- 主要职责：
- 自然语言理解：负责准确解析用户输入的自然语言问题，运用 NLTK、spaCy 等自然语言处理库，实现对问题语义、意图及关键信息（查询目标、筛选条件、数据范围等）的深度理解，处理简单或复杂、正式或口语化等多种表述形式。
- text2sql 转换：基于选定的 text2sql 大模型（如 Seq2SQL、SyntaxSQLNet 等），将理解后的自然语言问题精准转化为 SQL 语句，支持数据查询、添加、删除、更新等常见数据库操作，以及多表关联查询、嵌套查询等复杂情况。同时，研究大模型的微调方法，根据项目特定数据集和任务优化模型，提高 SQL 转换准确性。
- 单元测试与优化：编写单元测试用例，对自然语言理解和 text2sql 转换模块进行全面测试，依据测试结果优化模块性能，确保功能的正确性和稳定性。
- 二、熊益婕：数据库交互与结果解析呈现模块
- 主要职责：
- 数据库交互：负责与 MySQL、PostgreSQL、SQL Server 等各类主流数据库建立稳定连接，使用 pymysql、psycopg2 等数据库连接库，将生成的 SQL 语句安全、高效地发送至数据

库执行，并获取执行结果。同时，学习数据库事务处理方法，确保数据库操作的原子性、一致性、隔离性和持久性。

结果解析与呈现：对数据库返回的查询结果进行处理，运用 Matplotlib、Seaborn、Pandas - Plotting 等数据可视化库，将结果以直观、易懂的格式（表格、图表、简洁文本等）呈现给用户。此外，设计友好的错误提示，处理数据库操作过程中出现的连接失败、SQL 语法错误等问题。

单元测试与优化：编写单元测试用例，对数据库交互和结果解析呈现模块进行测试，根据测试反馈优化模块，保证功能正常运行。

三、郑涵丹：系统集成、上下文关联与错误处理反馈模块

主要职责：

上下文关联：实现多轮对话功能，使用会话机制记录用户历史对话信息，在会话过程中结合上下文信息理解用户新问题，避免用户重复输入背景信息，提供连贯流畅的问答体验。

错误处理与反馈：通过构建测试集对系统进行全面测试，计算问答准确性指标（精确率、召回率、F1 值等），并将结果反馈给用户。同时，使用日志记录和监控工具，记录系统运行过程中的错误信息和性能指标，以便及时发现和解决问题。

系统集成与测试：负责将各个功能模块进行集成，进行系统集成测试，模拟用户实际使用场景，对系统各项功能进行全面测试，发现并修复模块间协同工作存在的问题。根据测试结果对系统进行性能优化、功能优化和用户体验优化，编写系统需求文档、设计文档、使用手册等。

项目管理与协调：承担部分项目管理工作，协调成员 A 和成员 B 的工作进度，组织团队会议，确保项目按计划推进；在项目总结阶段，向团队成员和相关人员展示项目成果，分享开发经验和教训。