

# 基于自然语言处理技术的数据库实验 教学实践与探索\*

郑吉平

南京航空航天大学计算机科学与技术学院  
南京 211106  
软件新技术与产业化协同创新中心, 南京 210093

董婕

南京航空航天大学计算机科学与技术学院  
南京 211106

**摘要** 自然语言处理和 SQL 查询语言融合引入数据库实验教学是一种全新的教学方法。通过两者融合, 学生可以使用自然语言描述查询需求并转化为 SQL 语句, 提高对数据库查询的直观理解和应用能力。数据库实验教学中交互式实验环境、实时反馈和实际案例的应用使实验教学更加直观和易于理解, 该教学方式可以激发学生兴趣, 培养跨应用领域的综合能力, 并推动数据库实验教学的创新发展。

**关键字** 数据库实验教学, 自然语言处理, SQL 查询语言

## Practice and Exploration of Database Experimental Teaching Based on Natural Language Processing

Zheng Jiping

College of Computer Science and Technology  
Nanjing University of Aeronautics and Astronautics  
Nanjing 211106, China  
Collaborative Innovation Center of  
Novel Software Technology and Industrialization,  
Nanjing 210093, China  
jzh@nuaa.edu.cn

Dong Jie

College of Computer Science and Technology  
Nanjing University of Aeronautics and Astronautics  
Nanjing 211106, China  
djnuaa@nuaa.edu.cn

**Abstract**—The integration of natural language processing and SQL query language into database experimental teaching is a new teaching method. By integrating the two techniques, students can use natural language to describe query requirements and convert them into SQL statements, improving their intuitive understanding and application ability of database queries. The application of interactive experimental environments, real-time feedback, and practical cases in database experimental teaching makes experimental teaching more intuitive and easier to understand. This teaching method can stimulate interest of students, cultivate comprehensive abilities across application fields, and promote the innovative development of database experimental teaching.

**Keywords**—Database Experimental Teaching, Natural Language Processing, SQL Query Language

## 1 引言

在当今信息技术飞速发展的时代, 数据库技术在各行各业扮演重要角色, “数据库原理”课程是计算机专业的核心课程之一。然而由于学生的基础和个体兴趣差异, 传统的数据库实验教学并不能兼顾所有学生。

同时, 数据库实验教学要求熟练掌握结构化查询语言 (Structured Query Language, SQL), 这对于一般学生存在门槛。当前, 自然语言处理技术取得了

长足的发展, 学生可以使用自然语言与数据库系统进行交流, 降低了对 SQL 语言的学习要求。

通过解析学生自然语言输入并转换为 SQL 查询语句, 可以实现自然语言查询数据库。该方法提高了数据库查询的易用性和效率, 为学生提供更直观、高效的交互方式, 提高了数据库实验教学效率和效果。

## 2 相关基础

### 2.1 研究背景

随着数据量、数据规模和数据复杂性的不断增长, 要求高效地访问数据库, 即对数据库进行查询。

\*基金资助: 南京航空航天大学研究生教育教学改革研究项目“新工科背景下基于 SHEL 模型的研究生培养模式探索”(2022YJXGG43)

而对数据库进行查询需要掌握特定的查询语言，如：关系数据库标准查询语言 SQL。这对普通学生来说存在一定的门槛和障碍<sup>[1]</sup>。同时，由于云计算的普及和发展，为满足市场的需求，国产数据库产品需要适配各大主流云厂商的数据库产品；从 2022 年开始，国内的信创领域开始逐步推进国产数据库的适配，如：达梦、人大金仓、OceanBase、GaussDB 等。然而，这些国产数据库的有效使用必须需要学生熟练掌握数据库查询语言 SQL。

当前，自然语言处理技术（Natural Language Processing, NLP）不断突破，大语言模型（Large Language Model, LLM）<sup>[2,3,4]</sup>应运而生，如：BERT、GPT-3/4/5 等。目前，学生完全可以使用自然语言与数据库交流，这使得使用自然语言查询数据库成为可能。自然语言查询允许学生以更直观、更高效的方式与数据库进行交互。显然，将人工智能领域中自然语言处理技术引入数据库实验教学适应了新时代人工智能领域蓬勃发展的需求。

作为计算机专业的核心课程之一，“数据库原理”课程的实验教学对培养学生的专业素养和实践能力具有重要意义，不仅需要学生深入数据库原理知识，同时需要通过实践教学使学生灵活运用，使学生在理论和实验两方面融会贯通<sup>[5,6]</sup>。但当前数据库教学的实验方案并不能完全兼顾每个学生的基础能力，欠缺对个体差异性的考虑，导致教学效果存在明显的差异。对于大部分学生来说，使用 SQL 进行有效的数据查询其难度较大，而将自然语言处理技术融入数据库实验教学充分利用了 SQL 与自然语言处理这两种技术的优势，高效帮助学生理解数据查询的相关知识，同时又提高了实验教学的便捷性与创新性。

## 2.2 相关技术简介

### （1）自然语言处理

自然语言处理是计算机科学和人工智能领域的一个重要分支，旨在使计算机能够理解和处理人类的自然语言，促进机器理解和处理人类的语言表达。自然语言查询允许学生更直观、更高效地与数据库进行交互。基于 LLM（如：Gemini、ChatGPT 等）的强大功能，可以创建应用程序，使学生能够使用自然语言查询数据库。

### （2）SQL 查询语言

SQL 是一种用于管理和操作关系型数据库的标准查询语言，它允许学生通过指定查询条件和操作来从数据库中检索、更新、删除和插入数据。学生可以使用 SQL 在关系数据库系统、大型数据库中进行存储检索等操作。由于大数据分析已成为业务决策

的核心要素，因此人与数据库之间更自然的通信需求变得更加明显，更加需要强大、灵活的工具来帮助人们查询和分析这些数据并加以理解。但是，对于某些学生来说，SQL 的使用可能具有挑战性，因为它需要精确的、语法正确的查询才能正常工作。对于不熟悉 SQL 的学生或编写 SQL 语句有困难的学生来说，这已经成为达到数据库实验课程要求的一个障碍。

## 2.3 研究现状

当前，NLP 技术逐渐成为近年来计算机应用和人工智能研究的热点，其目标是让计算机具有类似人类的语言智能，以便更好地理解、处理和生成自然语言。得益于深度学习和大规模数据集的发展，NLP 技术在过去几年取得许多重要的进展，例如机器翻译、问答系统以及辅助文本生成等<sup>[7]</sup>。机器翻译是 NLP 领域的热门研究方向，近年来，神经机器翻译（Neural Machine Translation, NMT）的发展迅速：NMT 使用深度学习模型，例如循环神经网络（Recurrent Neural Networks, RNN）和 Transformer，提高了翻译的质量和效率。问答系统旨在使计算机能够回答学生提出的自然语言问题。基于预训练的语言模型的方法在问答系统中取得了显著的进展。这些方法通过对大规模语料进行自监督学习，使得模型能够理解问题并生成准确的答案。在得益于 NLP 技术的重大突破及其强大的功能，其已被广泛应用于教育教学之中。例如，智能辅助教学系统利用 NLP 技术帮助学生自主学习和解决问题<sup>[8]</sup>，系统可以回答学生的问题，提供解释和示例，并根据学生的学习进度和需求，为其推荐适当的学习资源和练习。其次，NLP 技术还可以与其它模态数据（如图像、音频等）进行结合，实现多模态教学，比如结合语音识别可以开发智能语音助教系统，提供口语练习和反馈。

NLP 俨然已经成为智能教育未来发展过程中最重要的人工智能技术。同时，NLP 正在结合其它人工智能技术，给智能教育的未来发展带来新的机遇，一定程度上也给传统的教学方法带来了新的挑战。自然语言具有多义性、歧义性和上下文依赖性，因此准确理解学生的查询意图可能是一个挑战。除此之外，对于查询复杂性、错误处理以及学生综合体验上都将面临新的问题，因此在将 NLP 技术融入到 SQL 教学过程中需要考虑相关问题，以提升教学效果。

## 3 NLP 引入 SQL 查询的数据库教学融合

### 3.1 相关工作

NLP 是人工智能（Artificial intelligence, AI）的一个重要领域，它关注如何使计算机能够理解、处理和生成自然语言。研究者们通过开发算法和模型，致

力于提高机器对自然语言的理解和处理能力,以实现自动化的语言处理任务<sup>[9]</sup>。这些研究成果为将 NLP 应用于教学提供了技术支持。

在教学领域, NLP 技术被广泛应用于语言学习、辅助教学和教育评估等方面。例如,语音识别和语音合成技术可以帮助学习者改善发音和听力;机器翻译技术能够为学生提供跨语言的学习支持;自动评估系统可以自动分析和评估学生的写作能力。数据库实验通常涉及 SQL 查询语句的编写和执行,而学生可能面临语法复杂性和抽象性的挑战。通过应用 NLP 技术,可以将自然语言查询转化为 SQL 查询,使学生能够以更直观和自然的方式与数据库进行交互<sup>[10]</sup>。与此同时,在教育过程中会降低学习门槛、提供实时学习反馈以及个性化学习支持,这都有助于学生更好地理解和应用 SQL 语言,同时有望拓宽相关数据库应用场景。采用 NLP 技术后不仅可以使 SQL 教学更加灵活和适应不同的应用场景,更有助于相应数据库系统的更新迭代,助力相关软件产业升级。

### 3.2 教学方法

将 NLP 技术引入数据库实验教学,即:将自然语言的查询需求转化为数据库的 SQL 查询语句,并通过设计交互式实验环境、提供实时反馈和纠错机制、结合案例和实际应用<sup>[11]</sup>,使学生能够更好地理解数据库查询的过程,促进数据库技术与自然语言处理技术的有机结合,具体如下。

(1) 设计交互式实验环境:开发一个基于 NLP 的交互式实验平台,让学生通过自然语言描述查询需求,并将其转换为相应的 SQL 查询语句。

(2) 提供实时反馈和纠错机制:在学生提交查询时,系统能够实时给予反馈和纠错建议,帮助他们理解和纠正语法或语义错误。

(3) 结合案例和实际应用:通过使用真实的数据库案例和实际应用场景,让学生在实验中应用 NLP 技术进行数据库查询,加深理解和实践能力。

(4) 引入自然语言生成技术:除了将自然语言转换为 SQL 查询语句,还可以考虑将 SQL 查询结果转换为自然语言描述,以双向反馈的形式帮助学生更好地理解 and 解释查询结果,提升学生对查询结果的理解和应用能力,达到教学目的。

### 3.3 教学设计

南京航空航天大学计算机科学与技术专业的数据库实验教学目标包括:理解 SQL 定义功能;熟练掌握 SQL 操纵功能;了解 SQL 数据控制功能。熟练掌握 Oracle、SQL Server、MySQL、或 PostgreSQL 等

对数据库的管理和操作,具体教学要求包括:实验一:SQL 定义功能、数据插入;实验二:数据查询;实验三:数据修改和删除;实验四:视图的操作;实验五:库函数与授权控制。在上述实验中,通过引入人工智能领域的 NLP 技术,将其与数据库操作和管理相结合,为学生提供开发智能化数据库应用的能力。

### 3.4 教学设计意义

学生可以用简单的自然语言表达查询需求,而无需深入学习特定数据库产品的语法。系统可以解析相关语言输入,并将其转换为相应的 SQL 查询语句,从而实现自然语言查询数据库的功能<sup>[12]</sup>。这种方法不仅提高了数据库查询的易用性和效率,还为学生提供了更直观、更高效的方式来与数据库进行查询和交互,从而节省了时间和资源<sup>[13]</sup>。

(1) 提升学生学习效率:传统的 SQL 语句编写可能需要学习复杂的语法和规则,而使用自然语言处理生成 SQL 查询语句可以简化学习过程。学生可以通过自然语言描述查询需求,然后自动转换为相应的 SQL 查询语句,减少了繁琐的语法细节,提高了学习效率。

(2) 增强学习兴趣:将自然语言处理引入数据库实验教学可以增加学生的学习兴趣。学生可以通过自然语言与数据库进行交互,将复杂的查询需求用自己熟悉的语言表达,从而更加主动地参与学习过程。

(3) 增强数据库查询实践能力:通过应用该技术学生可以增强数据库查询实践的能力,从问题定义到查询语句编写,增强他们的实际操作能力。

(4) 强化理解和应用:使用该技术可以帮助学生更好地理解查询语句的语义和结构,同时也让学生初步理解认识到人工智能领域的实用性与便捷性。学生可以通过自然语言描述查询需求,然后观察生成的 SQL 查询语句,进一步理解 SQL 语句的构成和执行过程。

(5) 培养综合能力:自然语言处理生成 SQL 查询语句需要学生综合运用数据库知识、自然语言理解和转换技术,有助于学生培养解决问题的能力,使他们能够更好地应对实际的数据查询和分析任务。

## 4 实例教学

实验教学采用 LangChain 大模型框架,将学生的自然语言转化为数据库 SQL 语句。

### 4.1 LLM 模型 LangChain

相比其它的 LLM 模型,LangChain 是一个开源的项目,可以让学生用自然语言和数据库聊天。它支

持多种数据库，如 MySQL, PostgreSQL, MongoDB 等，同时也支持多种语言。LangChain 运用了当前最先进的 NLP 技术，例如 BERT, GPT-3 等，来理解问题，并生成适用于 SQL 语句或数据库命令的合适策略。它还可以根据数据生成可视化图表，让学生更直观地显示数据分析结果。该模型目前提供了 SQL Chain (SqlDatabaseChain) 和 SQL Agent (SqlAgent) 两种方式来实现与存储在数据库中的数据进行交互。如图 1 所示，LangChain 框架由七个模块组成：

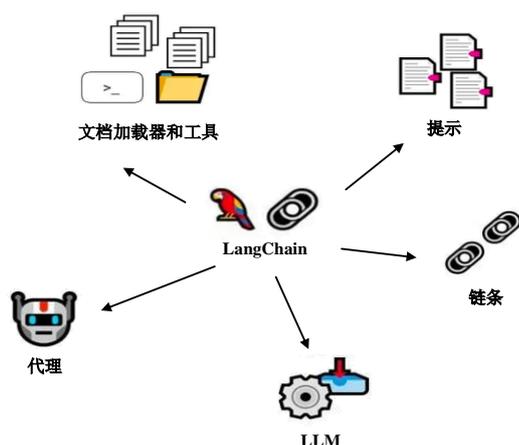


图 1 LangChain 框架

(1) LLM: LLM 是 LangChain 的基础组件。它是一个对 LLM 的封装，可以让学生利用模型的相关功能。

(2) 链条: 解决任务有时单独调用一个 LLM 的 API 是不够的。例如，在特定的 URL 获取数据，并对返回的文本进行摘要处理，借助生成的摘要回答相关问题。此模块可以把多个工具连接起来，以解决复杂的任务。

(3) 提示: 提示是任何自然语言处理应用的核心。它是学生和模型交互的方式，试图从模型那里得到一个输出。LangChain 提供了提示模板，可以让学生格式化输入和其它工具。

(4) 文档加载器和工具: LangChain 的文档加载器和工具模块可以帮助用户连接到数据源和计算，工具模块提供了 Bash 和 Python 解释器会话等。适合于那些需要直接和底层系统交互或者需要用代码片段来计算一个特定的数学量或者解决一个问题，而不是一次性地计算答案的应用。

(5) 代理: LangChain 库提供了代理的能力，其可以根据输入采取行动，并观察其结果来调整下一步行动，以此循环直到完成既定任务。与硬编码的确定性序列的代理不同，LangChain 库的代理更具灵活性。

(6) 索引: 最好的模型通常是那些与文本数据结合使用的模型，可以为模型提供上下文信息或解释相关内容。

(7) 内存: 此模块可以让学生在模型的调用之间创建一个持久化的状态，能够记住过去对话内容，提供更好的学生体验和应用功能。

## 4.2 教学实例

通过 LangChain 模型目前有两种自然语言查询数据库的办法：一是使用 prompt，令 ChatGPT 生成 SQL 语句，然后执行。这种方法操作简单，但是需要学生熟悉表结构；二是使用 LLM 将 ChatGPT 和数据库连接起来，直接用自然语言进行查询，对学生友好高效，但存在数据安全和成本高的问题。

基于 3.3 的教学设计，实验采用方法一进行实例化，部分查询实例如表 1 所示。

## 4.3 教学效果

一方面，从表 1 可以看出，将 NLP 与数据库技术结合，可以有效地表达查询需求，切实可行地将其转换为可执行的 SQL 语句，从而实现使用自然语言进行数据库查询的功能，进而使学生能够更直观地与数据库进行交互；另一方面，在实际操作过程中可能会面对复杂语义理解、存在歧义和灵活性限制的挑战，如表 1 中加粗字体所示，具体如下：

(1) 实验一中的数据插入 SQL 语句所需要描述的自然语言可能需要更多的上下文信息，以便正确解析插入哪些值；

(2) 实验二中的查询语句“查询不选 1 号课程的学生学号与姓名”，可能存在歧义：将“1 号课程”解释为课程名而不是课程号，从而导致生成的 SQL 语句错误；

(3) 实验二中的查询语句“查询不选 1 号课程的学生学号与姓名。”，使用 LEFT JOIN 左连接不仅仅是连接列所匹配的行，如果左表的某行在右表没有匹配行，则在相关联的结果集行中右表的所有选择列都为空值。查询结果虽然正确，但是不易于学生理解，在数据库课程实验中通常使用 EXISTS/NOT EXISTS 子句进行查询；

(4) 实验二中的查询语句“查询选修数据库原理成绩第 2 名的学生姓名。”，所生成的 SQL 语句中“LIMIT 1 OFFSET 1”的使用频率并不高，在数据库实验中通常使用两次 MAX 函数获取结果，因此该 SQL 语句的易用性较差；

表 1 自然语言查询数据库实例

实验	自然语言查询	生成的 SQL 查询
一	建立选课表SC(Sno, Cno, Grade) 选课 (学号, 课程号, 成绩)  在 S 表中增加 1 列 Saddress 在 C 表中增加一个元组(101, Math, NULL, 3)	CREATE TABLE SC( Sno VARCHAR(9), Cno VARCHAR(4), Grade INT, PRIMARY KEY (Sno,Cno), FOREIGN KEY (Sno) REFERENCES S(Sno), FOREIGN KEY (Cno) REFERENCES C(Cno)); ALTER TABLE S ADD COLUMN Saddress VARCHAR(100); INSERT INTO C (Cno, Cname, Cpno, Ccredit) VALUES (101, 'Math', NULL, 3);
二	查询不选 1 号课程的学生学号与姓名  查询选修数据库原理成绩第 2 名的学生姓名。  查询选课门数唯一的学生的学号	SELECT S.Sno, S.Sname FROM S LEFT JOIN SC ON S.Sno = SC.Sno AND SC.Cno = 1 WHERE SC.Sno IS NULL;  SELECT S.Sname FROM S JOIN SC ON S.Sno = SC.Sno JOIN C ON SC.Cno = C.Cno WHERE C.Cname = '数据库原理' 'ORDER BY SC.Grade DESC LIMIT 1 OFFSET 1; SELECT S.Sno FROM S JOIN SC ON S.Sno = SC.Sno GROUP BY S.Sno HAVING COUNT(DISTINCT SC.Cno) = COUNT(SC.Cno);
三	把 1 号课程的非空成绩提高 10% 在 SC 表中删除课程名为数据结构的学号 在 S 和 SC 表中删除学号为 201215122 的所有数据	UPDATE SC SET Grade = Grade * 1.1 WHERE Cno = 1 AND Grade IS NOT NULL; DELETE FROM SC WHERE Cno IN (SELECT Cno FROM C WHERE Cname = '数据结构'); DELETE FROM S WHERE Sno = '95002'; DELETE FROM SC WHERE Sno = '95002';
四	建立男学生的视图, 属性包括学号、姓名、选修课程名和成绩  男学生视图中查询平均成绩大于 80 分的学生学号与姓名	CREATE VIEW Male_Students AS SELECT S.Sno, S.Sname, C.Cname, SC.Grade FROM S JOIN SC ON S.Sno = SC.Sno JOIN C ON SC.Cno = C.Cno WHERE S.Ssex = '男'; SELECT Sno, Sname FROM Male_Students GROUP BY Sno, Sname HAVING AVG(Grade) > 80;
五	计算每个学生有成绩的课程门数、平均成绩  把对基本表 S、SC、C 的使用权限授给其它用户	SELECT SC.Sno, COUNT(SC.Cno) AS CourseCount, AVG(SC.Grade) AS AverageGrade FROM SC WHERE SC.Grade IS NOT NULL GROUP BY SC.Sno; GRANT SELECT, INSERT, UPDATE, DELETE ON S TO <其他用户>;

(5) 实验三的数据修改和删除的 SQL 语句中, 自然语言描述可能无法准确表达特定的修改和删除操作, 从而导致生成的 SQL 语句不符合预期;

(6) 实验四中的查询语句“男学生视图中查询平均成绩大于 80 分的学生学号与姓名。”，所生成的 SQL 语句结果有可能不正确, SELECT 中缺少关键词 DISTINCT 于返回唯一不同的值。

因此, NLP 技术所生成的部分 SQL 语句存在可解释性差的问题, 需要综合考虑其优势和局限性。尽管在面临新的教学实践技术要不断的进行探索和试错, 但是学生总体上从中得到了知识和技能的提升, 因此可以看出基于 NLP 技术融入教学实验的方案在强化计算机专业学生高阶能力的培养的方面存在可行性和必要性, 我们在教学实践中也希望强化学生问题分析和建模能力、系统性思维、实时反馈和自主学习能力和主动进行案例研究和实际应用, 最终启发学生独立研究和创新意识达到终生学习的目的。

## 5 结束语

本文探索了基于自然语言处理的数据库 SQL 查询语句的教学实践, 我们发现使用人工智能领域的自然语言处理技术可以大大简化学生理解和掌握 SQL 查询语句的过程。引入此方法进行数据库实验教学, 可以帮助学生更好地理解 SQL 语法和查询逻辑, 并提高实际应用能力。

## 参考文献

- [1] 刘喜平, 舒晴, 何佳壕等. 基于自然语言的数据库查询生成研究综述[J]. 软件学报, 2022, 33(11): 4107-4136.
- [2] A. Bazaga, N. Gunwant and G. Micklem. Translating Synthetic Natural Language to Database Queries with a Polyglot Deep Learning Framework. Scientific Reports, 2021, 11 (1), 18462.
- [3] S. Karimi, A. A. Rasel and M. S. Abdullah. Non-English Natural Language Interface to Databases: A Systematic Review, IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2022: 391-397.

- [4] A. Kolarkar and S. Kumar. A Peer Review on Natural Language Interface: Various Challenges and Scope, International Conference on Disruptive Technologies (ICDT), 2023: 252-258.
- [5] 邓芳, 叶文, 卢向群等. 新工科背景下融合 OBE 的《数据库系统原理》实验环节教学改革与实践[J]. 计算机技术与教育学报, 2021, 9(2): 54-58.
- [6] 白琳. 信息化技术赋能的“3+6”混合式教学创新与实践——以数据库课程为例[J]. 计算机技术与教育学报, 2023, 11(3): 88-91.
- [7] 李良炎. 基于词链接的自然语言处理技术及其应用研究[D]. 重庆大学, 2005.
- [8] 张博, 董瑞海. 自然语言处理技术赋能教育智能发展——人工智能科学家的视角[J]. 华东师范大学学报(教育科学版), 2022, 40(09): 19-31.
- [9] 陈程. 基于自然语言接口的数据库查询系统的研究[D]. 华北电力大学, 2015.
- [10] Z Rehman and K. Stefania. Teaching Natural Language Processing (NLP) Using Ontology Based Education Design. Balkan Region Conference on Engineering and Business Education, 2015: 206 - 214.
- [11] 滑贝贝, 何鸣皋. 基于 NLP 和 NLU 的智能教学系统设计[J]. 无线互联科技, 2022, 19(1): 54-55.
- [12] 赵飞, 刘文婷. 基于自然语言的数据库查询接口探究[J]. 赤峰学院学报(自然科学版), 2018, 34(6): 54-55.
- [13] 钱雅萍. 数据库自然语言接口技术研究在教学中的应用[J]. 咸宁学院学报, 2012, 32(8): 143-144+157.