arXiv:2303.07351v1 [cs.DB] 11 Mar 2023

面向数据库的上下文本体建模：赋能 ChatGPT 进行语义数据库管理

Wenjun Lin[1,2],PaulBabyn[1],YanYan[2] 和 Wenjun Zhang[1*]1*萨斯喀彻温大学工程学院，加拿大萨斯喀彻温省萨斯卡通市 Camous Dr 57 号，邮编 S7N5A9。[2]汤普森河大学理学院，加拿大不列颠哥伦比亚省坎卢普斯市 University Dr 835 号，邮编 V2C 0C8。

通讯作者。电子邮箱：chris.zhang@usask.ca；贡献作者：randy.lin@usask.ca；paul.babyn@gmail.com；yyan@tru.ca；

摘要

本文探讨了 ChatGPT 在数据库管理中的应用。ChatGPT 是一种人工智能驱动的聊天机器人，由于缺乏用于表示数据库语义的标准化词汇和语法，因此在执行数据库管理相关任务时存在局限性。为解决这一问题，本文提出了一种方案：开发一套语法，以自然语言表示数据库语义。利用这些语法，可将数据库模式转换为自然语言格式，从而为 ChatGPT 在数据库管理中的应用开辟新途径。本文通过一个案例研究展示了该方案：使用 ChatGPT 执行语义集成和表连接两项任务。结果表明，与未采用语义表示的情况相比，使用语义数据库表示能够获得更精确的结果，并避免常见错误。该方法有望加快数据库管理流程，降低对数据库领域知识的理解要求，并使系统无需访问实际数据即可自动执行数据库操作，从而缓解使用人工智能时对隐私保护的担忧。本文为基于人工智能的数据库管理领域研究提供了一个富有前景的新方向。

关键词： ChatGPT、数据库管理、语义数据库表示、语义集成、表连接

1 引言

ChatGPT 是一种对话式聊天机器人，结合自然语言处理（NLP）方法，运用人工智能（AI）和机器学习（ML）技术生成类人文本。它于 2022年11月推出，并迅速走红，仅用五天就吸引了超过一百万用户 [1]。ChatGPT 生成类人文本并执行各种任务的能力，使其成为许多用户青睐的工具，可用于回答问题、撰写短篇故事、创作音乐、解决数学问题、进行语言翻译，甚至编写计算机程序。

数据库操作是指使用特定语法或命令来处理数据和信息，这与计算机编程类似。在数据库操作中，这些命令称为数据库查询，用于检索、更新和处理存储在数据库中的数据。同样，在计算机编程中，人们使用 C语言 或 Python 等编程语言编写指令，以指定计算机程序的预期行为。

人们一直期待 ChatGPT 能够像协助编写计算机程序一样，协助生成数据库查询。然而，生成数据库查询需要理解数据库本身，而目前还没有表示数据库语义的常规方法。这一问题限制了 ChatGPT 执行数据库管理相关任务的能力。

本文通过开发一套语法来解决这一问题。这套语法能够以自然语言表示数据库语义，例如表结构和关系。由此生成的数据库语义表示可供 ChatGPT 理解，并使其能够执行数据库管理任务。我们通过一项案例研究展示了这项工作，其中使用 ChatGPT 执行了两项任务：语义集成和表连接。结果表明，与不使用语义表示的情况相比，使用语义数据库表示能够获得更精确的结果，并避免常见错误。

所提出的方法将数据库模式转换为自然语言格式，为 ChatGPT 在数据库管理中的应用开辟了新途径。本研究有望加快数据库管理流程，降低对数据库领域知识的理解要求，并在不访问实际数据的情况下实现数据库操作自动化，从而为使用人工智能时的隐私保护问题提供新的思路。

本文其余部分安排如下：第2节回顾了人工智能在数据库管理领域的相关工作。随后，我们在第3节介绍所提出的解决方案。第4 节展示研究结果

并讨论我们的案例研究。最后，我们讨论本研究的潜在优势与局限性，并总结未来的研究方向。

2 文献综述

2.1 基于人工智能的数据库查询生成

使用人工智能模型通过自然语言生成数据库查询，是多项研究关注的重点。Bais 等人提出的一种模型 [2] 利用 NLP 技术，通过执行形态分析、句法分析和语义分析来分析和解释用户查询，从而生成有效的 SQL 数据库查询。类似地，Sawant 等人 [3] 实施了一个系统，该系统运用 NLP 和长短期记忆（LSTM）等深度学习技术，根据文本和语音输入生成 SQL 查询。

其他研究，例如 Ghosh 等人 [4], Nagare 等人 [5], 以及 Kombade 等人 [6] , 也采用了词法分析、语法分析和语义分析等技术，从自然语言输入中提取 SQL 查询。Kombade 等人 [6] 甚至考虑在 NLP 中使用缩写来生成 SQL 查询。这些研究的实现使用 Python 和图形用户界面（GUI）进行输入和输出，用户可以通过语音或文本提供输入。

尽管该领域已取得进展，但由于自然语言复杂且含义模糊，同时缺乏用于表示数据库结构的标准化词汇和语法，人工智能模型在根据自然语言准确生成数据库查询方面仍存在局限。例如，Nagare 等人 [5] 提到，系统会检查用户查询的有效性，但并未明确说明如何判定查询是否有效。此外，这些研究仅考虑了基本的数据库操作，如选择、删除和更新。多表连接和语义集成等复杂操作尚未得到研究。

2.2 语义集成

语义集成对于解决相关数据库之间的数据表示不匹配至关重要。在当今的数字世界中，机构会在各种数据库中生成并存储大量数据，这往往会导致数据表示不一致。例如，一个数据库可能使用“社会安全号码”这一属性来识别个人，而另一个数据库可能使用“SSN”这一属性。在这种情况下，确定这些属性之间的关系至关重要，这样才能准确比较和合并来自这些数据库的数据。

近年来，人们开发了多种语义集成方法。Silk [7], LIMES [8], 和 PARIS [9]等工具利用字符串相似度指标、函数属性和手动配置来检测匹配的属性。WebPie [10]和 LINDA [11] 是全自动系统，采用邻域检查和块放置等技术。MateTee [12]和 RDF2VEC [13] 是较新的方法，它们利用嵌入和

机器学习来寻找相似性。然而，当实体以缩写形式出现时，实体匹配任务可能会变得复杂 [14]。

现有方法高度依赖领域专业知识，且需要进行复杂的准备工作。例如，基于数据集的方法 [12, 13] 需要确定属性之间的关系。基于关键词的方法 [7, 8] 取决于元数据的准确性，而基于 URI 的方法 [10]则严重依赖 HTTP URI 的解引用。

3 方法论

3.1 ChatGPT

ChatGPT 是 OpenAI [15]开发的一种语言模型。它是一种人工智能算法，根据前文语境预测给定词序列出现的可能性。该技术基于自注意力机制 [16]，并通过海量文本数据集进行训练，因此能够生成复杂且看似智能的文字。ChatGPT 旨在用英语及其他语言与用户就广泛主题进行对话，因此非常适合用于聊天机器人、客户服务、内容创作和语言翻译等任务。

ChatGPT 的一项应用是辅助编程，这可以通过两种方式实现。首先，ChatGPT 可以充当编程助手或工具。例如，开发人员可以向 ChatGPT 提出与编程相关的问题，并获取有关一般工作流程和步骤的建议。其次，ChatGPT 可以直接生成代码片段，从而提高开发人员的工作效率并节省时间。

尽管 ChatGPT 具备先进的自然语言处理能力，并在辅助编程方面取得了成功，但它仍无法为数据库生成查询，因为数据库模式包含有关数据库结构的重要信息，而这些模式往往以图而非自然语言的形式编写。

3.2 面向数据库的上下文本体建模

本研究提出了一种新方法——面向数据库的上下文本体建模（COM-DB），旨在将数据库模式转换为自然语言。COM-DB 建立在我们此前关于本体建模的研究之上 [17], ，该研究采用上下文关系、单向关系和双向关系等构造来描述数据库中概念之间的关系。

在本研究中，我们着重探讨如何使用“context-of” 构造来描述数据库模式，尤其是在概念数据模型层面。概念模式或概念数据模型是用于数据库的一张概念及其关系的图谱。COM-DB 的关键特性是能够将这些关系转换为自然语言，从而让 ChatGPT 更容易理解。为了展示 COM-DB 的有效性，我们提供

两个示例，说明如何使用“context-of” 构造来描述同一张表中表头之间的关系，以及同一数据库中各表之间的关系。

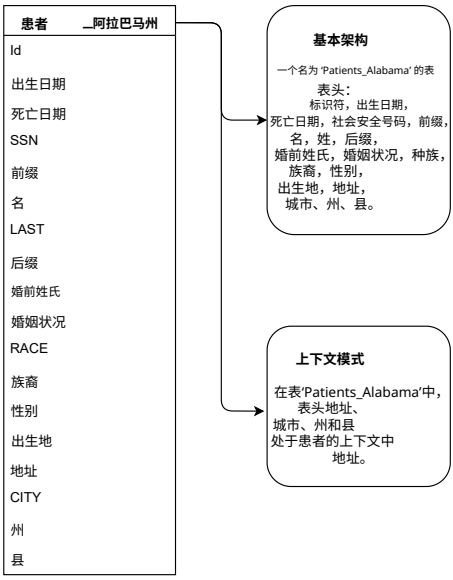


图 1 使用 COM-DB 描述同一表中表头之间的关系

图 1 中的示例展示了如何使用 COM-DB 描述数据库表中表头之间的关系。图的左侧显示了典型的数据库表模式，其中包含表名和表头名称。

COM-DB 将此模式转换为两部分：基础模式，用于描述表中表头的名称；以及上下文模式，它使用“context-of” 构造来描述表头之间的关系。在此示例中，“ADDRESS”、“CITY”、“STATE”和“COUNTY”彼此相关，用于存储患者地址信息。

上下文模式使用“context-of” 构造来描述这些表头与患者地址之间的关系，从而将这些信息浓缩为一个句子。具体而言，它指出“ADDRESS、CITY、STATE、COUNTY 处于患者地址的上下文中”。这用一个句子表达了四种关系，使信息更易于理解，也更加简洁。

图 2是医院数据库的实体-关系模式。与第一个示例类似，COM-DB 将数据库分为两部分进行表示。第一部分是基本架构，用于说明每个表中的表头。以下是一些基本架构：

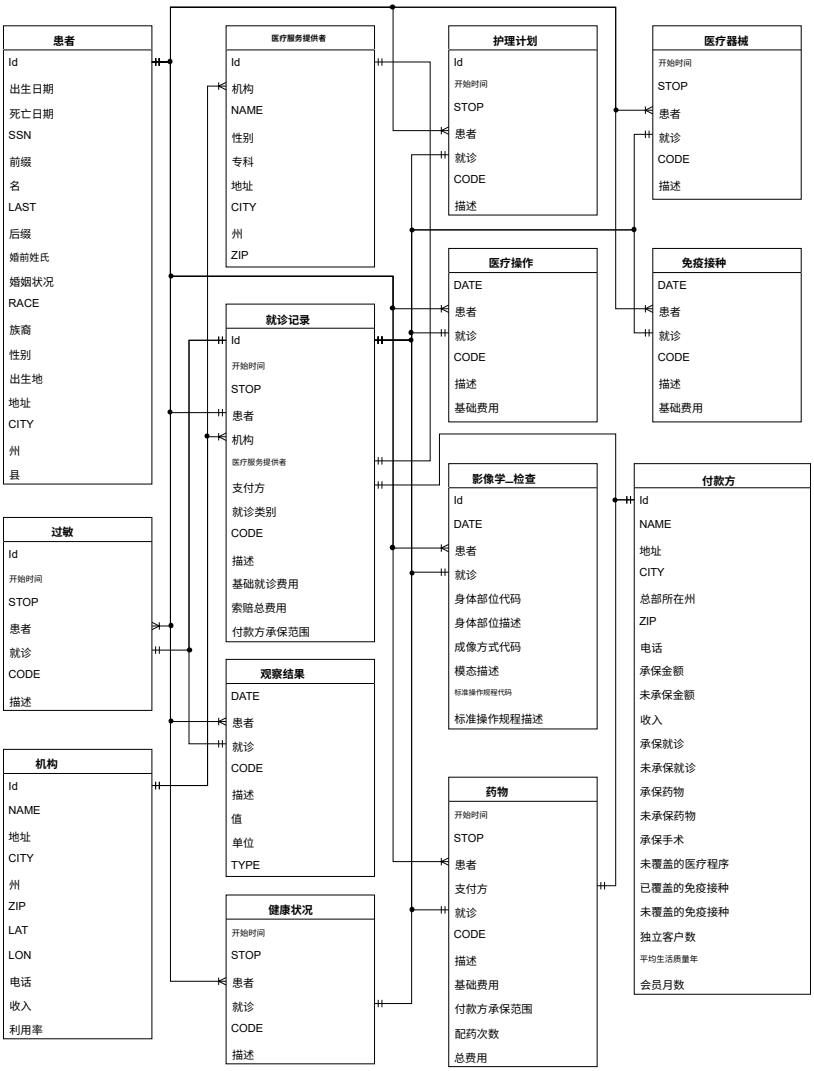


图 2 医院数据库的实体-关系模式 [18]

给定一个名为 'allergies' 的表，其表头为：标识符、开始时间、结束时间、患者、就诊、代码、描述。另有一个名为 'careplans' 的表，其表头为：标识符、开始时间、结束时间、患者、就诊、代码、描述。还有一个名为 'conditions' 的表，其表头为：开始时间、结束时间、患者、就诊、代码、描述。以及一个名为 'devices' 的表，其表头为：开始时间、结束时间、患者、就诊、代码、描述、医疗器械唯一标识。

还有一个名为 'encounters' 的表，表头为：Id、START、STOP、PATIENT、ORGANIZATION、PROVIDER、PAYER、ENCOUNTERCLASS、CODE、DESCRIPTION、BASE_ENCOUNTER_COST、TOTAL_CLAIM_COST、PAYER_COVERAGE。还有一个名为 'imaging_studies' 的表，表头为：Id、DATE、PATIENT、ENCOUNTER、BODYSITE_CODE、BODYSITE_DESCRIPTION、MODALITY_CODE、MODALITY_DESCRIPTION、SOP_CODE、SOP_DESCRIPTION。还有一个名为 'immunizations' 的表，表头为：DATE、PATIENT、ENCOUNTER、CODE、DESCRIPTION、BASE_COST。还有一个名为 'medications' 的表，表头为：START、STOP、PATIENT、PAYER、ENCOUNTER、CODE、DESCRIPTION、BASE_COST、PAYER_COVERAGE、DISPENSES、TOTALCOST。还有一个名为 'observations' 的表，表头为：DATE、PATIENT、ENCOUNTER、CODE、DESCRIPTION、VALUE、UNITS、TYPE。还有一个名为 'organizations' 的表，表头为：Id、NAME、ADDRESS、CITY、STATE、ZIP、LAT、LON、PHONE、REVENUE、UTILIZATION。还有一个名为 'patients' 的表，表头为：Id、BIRTHDATE、DEATHDATE、SSN、PREFIX、FIRST、LAST、SUFFIX、MAIDEN、MARITAL、RACE、ETHNICITY、GENDER、BIRTHPLACE、ADDRESS、CITY、STATE、COUNTY。还有一个名为 'payers' 的表，表头为：Id、NAME、ADDRESS、CITY、STATE_HEADQUARTERED、ZIP、PHONE、AMOUNT_COVERED、AMOUNT_UNCOVERED、REVENUE、COVERED_ENCOUNTERS、UNCOVERED_ENCOUNTERS、COVERED_MEDICATIONS、UNCOVERED_MEDICATIONS、COVERED_PROCEDURES、UNCOVERED_PROCEDURES、COVERED_IMMUNIZATIONS、UNCOVERED_IMMUNIZATIONS、UNIQUE_CUSTOMERS、QOLS_AVG、MEMBER_MONTHS。还有一个名为 'procedures' 的表，表头为：DATE、PATIENT、ENCOUNTER、CODE、DESCRIPTION、BASE_COST。还有一个名为 'providers' 的表，表头为：Id、ORGANIZATION、NAME、GENDER、SPECIALITY、ADDRESS、CITY、STATE、ZIP。

此外，第二部分——上下文模式——如下：

过敏、护理计划、健康状况、医疗器械、免疫接种、观察结果、医疗操作、影像学检查均处于患者和就诊记录的上下文中。就诊记录处于患者、机构、医疗服务提供者和付款方的上下文中。药物处于患者、就诊记录和付款方的上下文中。医疗服务提供者处于机构的上下文中。

请注意，上下文模式以精简形式呈现。例如，“过敏、护理计划、健康状况、医疗器械、免疫接种、观察结果、医疗操作、影像学检查处于患者和就诊记录的上下文中。”表示$^{-}$ $8 \times 2 = 16$ 个关系。这些关系存在于 8 个表“过敏、护理计划、健康状况、医疗器械、免疫接种、观察结果、医疗操作、影像学检查”与 2 个表“患者、就诊记录”之间。

我们用这些示例展示 COM-DB 的实际应用，以及如何利用它生成复杂数据库模式的自然语言描述。总体而言，本研究的方法论包括设计

并实现 COM-DB 方法，其中包括运用“context-of”构造及其他本体建模构造，将数据库模式转换为自然语言。通过两个示例展示了该方法的有效性，这些示例说明了如何利用该方法完成两项复杂的数据库管理任务。案例研究进一步证明了该方法的有效性。

4 案例研究

本案例研究旨在展示所提出的 COM-DB 系统的有效性。该系统的主要特点是“context-of”构造，它利用自然语言捕捉数据库语义，例如表结构和关系。该系统的主要目标是创建易于 ChatGPT 理解的数据库语义表示，使其能够执行各种数据库管理任务。

本案例研究提供了实证证据，以支持 COM-DB 的有效性。研究从文献中收集了两个样本数据库：Synthea Alabama [18] 和 BDA EHR [19]。基于这些数据库，研究开展了两项实验，分别代表数据库集成中的典型任务：语义集成和表连接。在两项实验中，均使用 ChatGPT 在有无基于 COM-DB 的模式的情况下执行任务。为确保结果可靠并消除 ChatGPT 性能可能存在的不一致，研究将每项实验重复进行 10 次。结果展示了重复实验的平均结果。

4.1 实验1：语义集成

语义集成是指合并两个包含同一类信息的表。不同表中的表头名称各异，往往会导致兼容性问题。’患者 A’和’患者 B’分别是来自 BDA EHR 和 Synthea Alabama 的患者信息表。’患者 A’包含以下表头：患者标识符、名称、姓氏、出生日期、出生地、地址、性别、血型、职业。’患者 B’包含以下表头：标识符、出生日期、死亡日期、社会安全号码、前缀、名、姓、后缀、婚前姓氏、婚姻状况、种族、族裔、性别、出生地、地址、城市、州、县。

本实验旨在找出表’患者 A’和表’患者 B’中包含相同信息的表头，并考虑到映射时可能需要合并或拆分某些表头。理想的映射如表 1所示。

图 3 展示了不使用基于 COM-DB 模式时，使用 ChatGPT 的输入和输出。来自图标 ’FF’ 的消息是输入，而来自图形图标的消息则是 ChatGPT 的输出。输入包含两部分信息。第一部分用于说明情况，其中包含每个表的表头名称。请注意，此处仅提供表头，没有任何样本数据或数据类型。第二部分“识别

表 1 表“患者 A”和表“患者 B”的理想表头映射结果

患者 A	患者 B
Name	名
姓氏	LAST
出生日期	出生日期
出生地	出生地
地址	地址 城市 州 县
性别	性别

表 ’患者 A’ 和表 ’患者 B’ 中包含相同信息的表头。某些表头可能需要合并或拆分。”这是对 ChatGPT 要完成任务的说明。

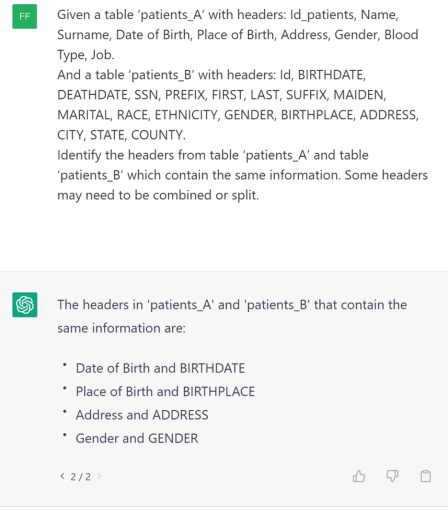


图 3 实验1：不使用基于 COM-DB 模式的表头映射

图 3 表明，ChatGPT 能够理解任务并在一定程度上完成任务。它正确匹配了 Date of Birth 和 BIRTHDATE、Place of Birth 和 BIRTHPLACE、Gender 和 GENDER。然而，它未能将 Name 与 FIRST、Surname 与 LAST 匹配。此外，患者 B 中的 ADDRESS 应与其他表头 CITY、STATE、COUNTY 一起使用。ChatGPT 并未注意到这一点。

图4展示了使用 ChatGPT 与基于 COM-DB 的模式时的输入和输出。除了图3中使用的输入外，本体模型信息还描述为：“在表‘患者 A’中，表头 Name 和 Surname 属于患者姓名这一上下文。在表‘患者 B’中，表头 ADDRESS、CITY、STATE 和 COUNTY 属于患者地址这一上下文。”

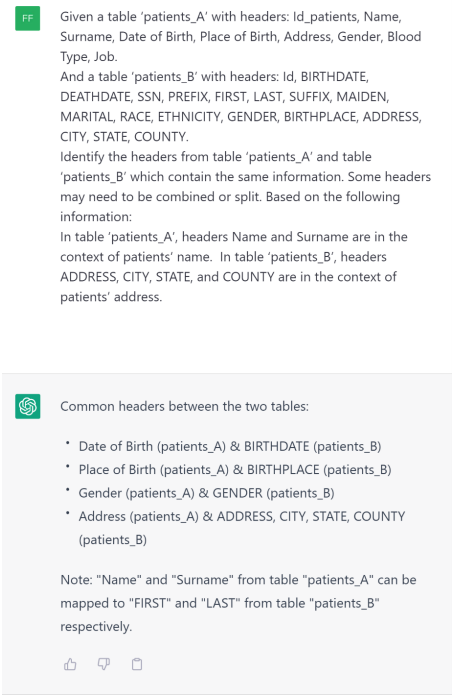


图 4 实验1：基于 COM-DB 模式的表头映射

图4 展示了 ChaGPT 的典型结果，并表明基于 COM-DB 的模式提升了语义集成的效果，因为 ChatGPT 已成功识别出所有预期的映射。

4.2 实验2：表连接

表连接是指生成一个新表或视图，将多个表中的数据合并起来。此过程需要实体关系模式，而这种模式没有常规的自然语言表示方式。为展示基于 COM-DB 的模式在表示实体关系方面的有效性，本实验使用 Synthea Alabama。该数据库包含 14 个表，如图 2 所示。本实验的目标是创建一条 SQL查询，生成护理计划列表，并附上相应医疗服务提供者和患者的身份信息。护理计划是医疗服务提供者（如医师）向患者提供的建议。SQL查询需要正确连接四个表：护理计划、医疗服务提供者、患者和就诊记录。就诊表在此发挥关键作用，因为它将患者表与护理计划表连接起来。这类信息通常包含在实体关系模式中。

图 5 展示了不使用基于 COM-DB 模式时 ChatGPT 的输入和输出。输入包含两部分信息。第一部分是

按字母顺序说明所有表及其包含的表头。与实验1类似，此处仅提供表头，不包含任何样本数据或数据类型。第二部分“创建一条 SQL查询，以生成护理计划列表，并附上相应医疗服务提供者和患者的身份信息。”是对 ChatGPT 待完成任务的说明。

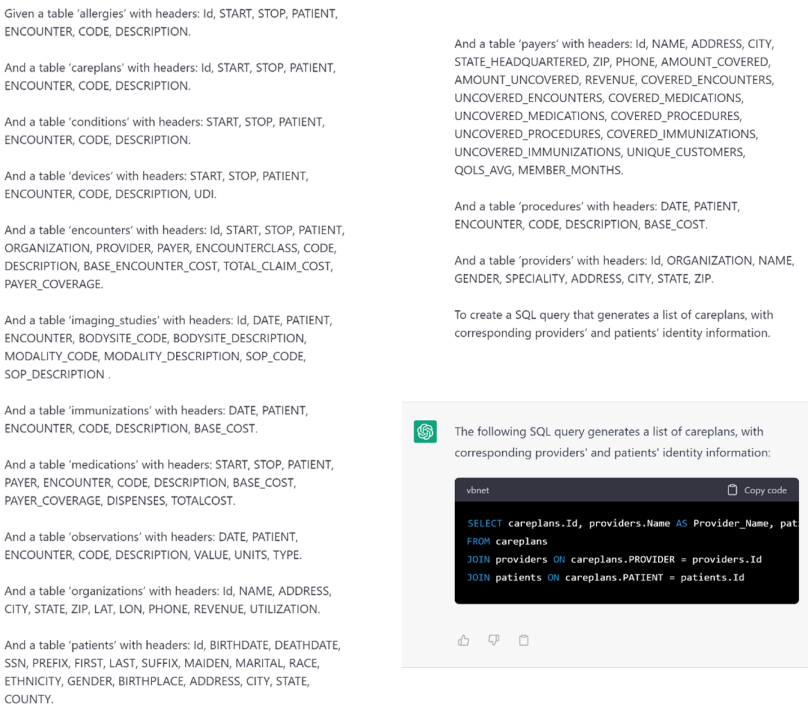


图5 实验2：在没有 COM-DB 模式的情况下，从多个表生成新视图。对话分为左右两栏。

通过在医院数据库中执行 SQL查询，验证 ChatGPT 的输出。结果如图6所示。从结果可以看出，该查询因错误“no such column: careplans.PROVIDER”而无法执行。正如前文所述，此错误的根本原因是缺少就诊表。

图7展示了使用基于 COM-DB 模式的 ChatGPT 的输入和输出，该模式说明了每个表的上下文信息。在这种情况下，基于 COM-DB 的模式描述了表之间的关系。图8通过在医院数据库中执行 SQL 查询对其进行验证。结果表明，ChatGPT 成功生成了查询，并得到了正确的视图。

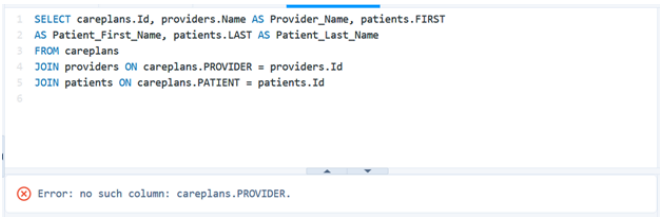


图 6 实验2：不使用基于 COM-DB 模式的 SQL 查询结果

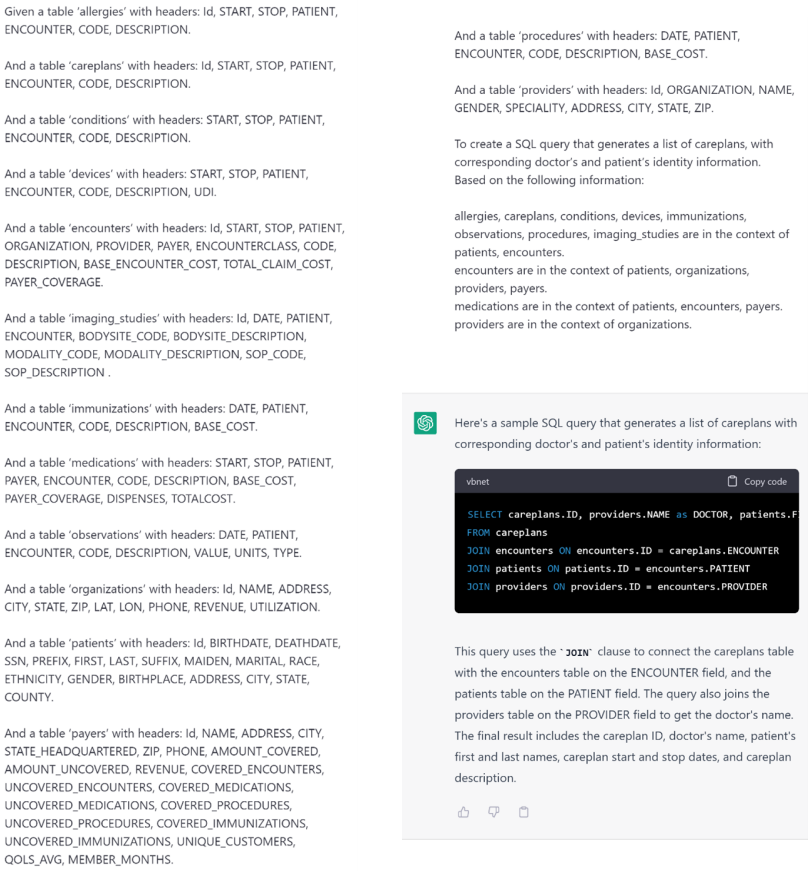


图 7 实验2：使用基于 COM-DB 的模式从多个表生成新视图。对话从左到右分为两列。

5 讨论

实验结果表明，使用基于 COM-DB 的模式时，ChatGPT 在语义集成和表连接任务方面的表现均更好。本体模型提供的上下文信息有助于 ChatGPT 更好地完成这些任务。该研究表明，

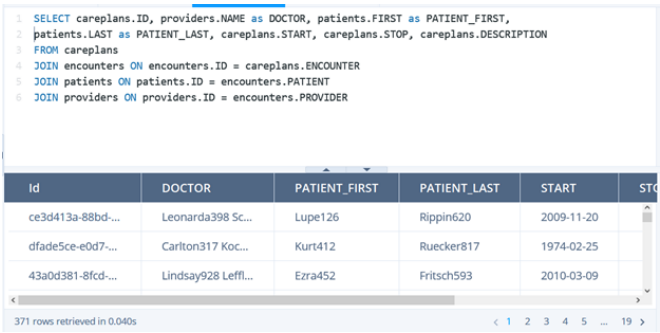


图 8 实验2：基于 COM-DB 模式的 SQL 查询结果

COM-DB 中的“context-of” 构造能够捕获传统本体模型可能未包含的某些上下文信息。这些上下文信息决定了概念之间的关系，有助于消除概念间的歧义，并提高数据库集成成功的可能性。

此外，COM-DB 能够在不影响隐私保护的前提下自动执行数据库操作。与现有的人工智能驱动的数据库操作不同，后者要求人工智能算法访问数据库中的所有数据，而 COM-DB 仅根据表结构生成模式，而非依据其内容。该模式不包含任何隐私信息，因此可以安全地与 ChatGPT 等第三方服务共享。将模式发送给 ChatGPT 后，隐私泄露风险会显著降低，因为 ChatGPT 无需访问敏感数据即可执行自动化数据库操作。这使企业和机构能够利用人工智能和自动化的强大能力来简化运营、提高效率，同时不牺牲客户数据的隐私和安全。借助 COM-DB，企业可以放心地确保数据安全并受到保护，同时仍能享受自动化数据库操作带来的益处。

6 结论

本文探讨了 ChatGPT 在数据库管理领域的应用，并着重指出使用自然语言处理执行数据库查询所面临的挑战。我们的研究通过开发一套用于以自然语言表示数据库语义的语法，提出了一种解决方案。这些称为 COM-DB 的语法使 ChatGPT 能够执行与数据库管理相关的任务，例如语义集成和表连接。我们的案例研究表明，与不使用此类表示的情况相比，在数据库管理中使用语义表示能够带来更精准的结果，并减少常见错误。

我们的研究旨在通过提出一种将数据库模式转换为自然语言形式的新方法，为数据库管理领域作出贡献，从而为 ChatGPT 开辟新的应用场景。

这种方法有望带来显著益处，包括加快数据库管理、降低对领域知识的要求，并通过无需访问实际数据的自动化数据库操作增强隐私保护。

未来的工作包括扩大我们方法的适用范围，以支持更复杂的数据库操作，并在更大型的数据库上进行测试。此外，我们打算研究将其他自然语言处理模型纳入数据库管理的可行性，并探索结合不同模型以增强其能力的可能性。

总之，我们的研究展示了自然语言处理模型在数据库管理领域的应用潜力，为交互和操作数据库提供了一种新方式。通过结合 ChatGPT 的强大能力与我们的 COM-DB 语法，我们证明了可以使用自然语言执行复杂的数据库操作，从而为简化数据库管理、提高生产力提供了一种新方法。

致谢。 本研究获得加拿大自然科学与工程研究委员会联盟资助，项目编号为 #ALLRP 555161 - 20。

参考文献

[1] Thorp, H.H.：《ChatGPT 很有趣，但不能作为作者》。美国科学促进会（2023）[2] Bais, H., Machkour, M., Koutti, L.：用于查询数据库的通用自然语言接口模型。《智能系统与应用国际期刊》**8**(2), 35（2016）[3] Sawant, A., Raina, R., Patil, A., Pardeshi, A.：通过语音将自然语言指令转换为 SQL 查询的人工智能模型。载于：《物理学杂志：会议系列》，第 2273 卷，页码 012014（2022）。IOP 出版社[4] Ghosh, P.K., Dey, S., Sengupta, S.：根据自然语言查询自动生成 SQL 查询。《国际计算机应用杂志》**975**, 8887（2014）[5] Nagare, P., Indhe, S., Sabale, D., Thorat, G., Chaturvedi, P.：根据自然语言查询自动生成 SQL 查询。《国际研究应用科学与工程技术期刊》**4**, 1589–1591（2017）[6] Kombade, C., More, M., Pujari, A., Patil, S.：使用带有一些缩写的自然语言处理生成 SQL。《应用科学与工程技术研究国际期刊》**8**(5), 1046–1048（2020）

[7]Volz, J., Bizer, C., Gaedke, M., Kobilarov, G.：Silk——面向数据网络的链接发现框架。Ldow **538**, 53（2009）[8] Ngomo, A.-C.N., Auer, S.：Limes——一种面向数据网络的大规模链接发现高效方法。integration **15(3)**（2011）[9] Suchanek, F.M., Abiteboul, S., Senellart, P.：Paris：关系、实例和模式的概率对齐。arXiv 预印本 arXiv:1111.7164（2011）[10] Urbani, J., Kotoulas, S., Maassen, J., Van Harmelen, F., Bal, H.：Webpie：一种使用 MapReduce 的网络规模并行推理引擎。《网络语义学杂志》**10**, 59–75（2012）[11] B¨ohm, C., De Melo, G., Naumann, F., Weikum, G.：Linda：分布式数据网络规模实体匹配。载于：第 21 届 ACM 信息与知识管理国际会议论文集，页码 2104–2108（2012）[12] Morales, C., Collarana, D., Vidal, M.-E., Auer, S.：Matetee：一种基于翻译嵌入的知识图谱语义相似度度量。载于：Web工程：第 17 届国际会议（ICWE 2017），意大利罗马，2017 年 6 月 5—8 日，会议论文集 17，页码 246–263（2017）。施普林格[13] Ristoski, P., Paulheim, H.：RDF2Vec：用于数据挖掘的 RDF图嵌入。载于：语义网——ISWC 2016：第 15 届国际语义网会议，神户，日本，2016 年 10 月 17—21 日，会议论文集，第一部分 15，页码 498–514（2016）。施普林格[14]Lantzaki, C., Papadakos, P., Analyti, A., Tzitzikas, Y.：使用签名进行感知半径的近似空白节点匹配。《知识与信息系统》**50**, 505–542（2017）[15] van Dis, E.A., Bollen, J., Zuidema, W., van Rooij, R., Bockting, C.L.：ChatGPT：研究的五项优先事项。《自然》**614(7947)**, 224–226（2023）[16]Humphreys, G.W., Sui, J.：注意控制与自我：自我注意网络（SAN）。《认知神经科学》**7(1-4)**, 5–17（2016）[17] Lin, W., Babyn, P., Yan, Y., Zhang, W.：现代计算机时代的本体。企业信息系统（2023 年，已投稿）[18]Walonoski, J., Kramer, M., Nichols, J., Quina, A., Moesel, C., Hall, D., Duffett, C., Dube, K., Gallagher, T., McLachlan, S.：Synthea：生成合成患者和合成电子医疗记录的一种方法、流程和软件机制。《美国医学信息学会杂志》**25(3)**, 230–238（2018）

[19]Silvestri, S., Esposito, A., Gargiulo, F., Sicuranza, M., Ciampi, M., De Pietro, G.：用于电子健康记录数据提取与分析的大数据架构。载于：2019年IEEE世界服务大会（SERVICES），第2642卷，第283–288页（2019）。IEEE