

基于大语言模型的糖尿病智能问答系统

徐步翔, 张 婷

西京学院计算机学院, 陕西 西安

收稿日期: 2025年11月17日; 录用日期: 2025年12月19日; 发布日期: 2025年12月31日

摘要

为探索大语言模型在医学知识问答领域的应用, 满足糖尿病患者的健康管理知识需求, 本研究构建了糖尿病智能问答系统。方法上, 以权威糖尿病防治指南为核心, 结合爬虫采集的科普信息构建含32,104个实体、31,242个关系的本地知识图谱, 采用LangChain4J框架、RAG技术及前后端分离架构开发系统。结论为: 该系统可精准理解患者口语化提问并提供专业通俗的解答, 支持多轮对话, 还整合了健康记录、饮食推荐、用药提醒等功能, 在提高患者自我管理能力的同时, 也能缓解医疗资源紧张的问题, 具有一定的应用前景和社会价值。

关键词

糖尿病, 智能问答系统, 知识图谱, LangChain4J框架, 健康管理

Intelligent Diabetes Question Answering System Based on Large Language Models

Buxiang Xu, Ting Zhang

School of Computer Science, Xijing University, Xi'an Shaanxi

Received: November 17, 2025; accepted: December 19, 2025; published: December 31, 2025

Abstract

To explore the application of large language models (LLMs) in the field of medical knowledge question answering and meet the health management knowledge needs of diabetic patients, this study constructed an intelligent diabetes question answering system. Methodologically, guided by authoritative diabetes prevention and management guidelines and combined with popular science information collected via web crawlers, a local knowledge graph containing 32,104 entities and 31,242 relationships was constructed. The system was developed using the LangChain4J framework, RAG technology, and a front-end/back-end separation architecture. The conclusion is that this system

can precisely understand patients' colloquial queries and provide professional yet easy-to-understand answers, supporting multi-turn dialogues. It also integrates functions such as health record tracking, dietary recommendations, and medication reminders. By enhancing patients' self-management capabilities, it also helps alleviate the strain on medical resources, demonstrating certain application prospects and social value.

Keywords

Diabetes, Intelligent Question Answering System, Knowledge Graph, LangChain4J Framework, Health Management

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

糖尿病作为一种常见的慢性疾病，正日益成为全球性的公共卫生挑战。在中国，糖尿病患者数量庞大且呈上升趋势，已成为世界上糖尿病患者最多的国家之一。近 30 余年患病率呈爆发式增长：1980 年全国患病率仅 0.67%，2018~2019 年 18 岁以上人群患病率已达 11.9%，30 余年间增长超 17 倍，疾病负担持续加重[1]。糖尿病不仅给患者个人带来身体和心理上的痛苦，还对社会医疗资源造成了沉重负担。长期高血糖状态可引发多种严重的并发症，如心血管疾病、肾脏疾病、神经病变和视网膜病变等，这些并发症不仅会显著降低患者的生活质量，甚至可能危及生命。

在糖尿病的管理过程中，患者需要持续获取准确、全面的疾病知识，以辅助他们做出合理的治疗决策、调整生活方式并进行有效的自我监测。然而，糖尿病知识体系繁杂，涵盖了疾病病理、治疗方法、饮食运动、药物管理等多个方面，普通患者往往难以系统掌握。传统的获取知识途径，如医生咨询，受限于时间和空间，难以满足患者随时、频繁的咨询需求；而查阅医学书籍和网站，又面临信息过载、质量参差不齐以及难以理解专业术语等问题。因此，开发一个高效、智能的糖尿病问答系统具有重要的现实意义。

随着人工智能技术的飞速发展，特别是自然语言处理(NLP)、知识图谱和大语言模型(LLM)等技术的不断突破，为构建智能问答系统提供了强大的技术支持。知识图谱以结构化的方式组织和表示知识，能够清晰地展示实体之间的关系。检索增强生成简称 RAG (Retrieval-augmented generation)，它是一种深度学习模型架构，旨在通过结合生成模型与检索模型以实现在自然语言处理任务中的性能提升[2]。通过知识图谱检索增强技术，可以从海量的糖尿病知识中快速、准确地提取与用户问题相关的信息，提高答案的准确性和可靠性。LangChain4j 则是一个强大的 Java 框架，它简化了将大语言模型集成到 Java 应用程序中的过程，能够实现智能对话、文本生成等功能，为糖尿病问答系统赋予了更加灵活和智能的交互能力。

本研究旨在综合运用知识图谱检索增强和 LangChain4j 等技术，构建一个功能强大、性能优越的糖尿病问答系统。该系统能够理解患者用自然语言提出的问题，利用知识图谱中的结构化知识和大语言模型的语言理解与生成能力，为患者提供准确、详细且易于理解的回答。同时，系统还提供健康记录、用药提醒以及饮食推荐等功能，为糖尿病患者的健康管理提供全方位、个性化的支持，在提高患者自我管理能力的同时，也能缓解医疗资源紧张的问题，具有一定的应用前景和社会价值。

2. 相关工作

2.1. 知识图谱构建

在信息爆炸与数据驱动的智能时代,海量非结构化、半结构化数据的快速涌现,对信息的有效组织、深度理解与高效利用提出了严峻挑战。传统数据处理方式多依赖关键词匹配或线性存储,难以捕捉数据背后隐藏的语义关联,导致信息检索的精准度不足、知识挖掘的深度有限。在此背景下,知识图谱(Knowledge Graph)作为一种结构化的语义知识库技术应运而生,知识图谱是2012年Google知识图谱发布后首次确定其概念,本质上是在语义网的基础上发展而来的知识库,以结构化的形式对真实世界中的概念、实体以及实体间的关系展开描述。知识图谱能够将某一领域分散、非结构化信息转化为结构化、可理解的知识库,实现数据与知识的结合[3]。其核心是以“实体-关系-实体”的三元组为基本单元,将现实世界中的概念、实体及其复杂关联进行形式化表达与可视化存储,打破了孤立数据间的信息壁垒。知识图谱不仅实现了数据的结构化组织,更通过语义网络的构建赋予数据“理解”能力,支持基于关联关系的推理与知识发现,为机器模拟人类认知过程提供了重要的技术支撑。

研究首先通过网络爬虫技术,定向采集互联网中公开可获取的糖尿病相关知识与实用信息;同时,选取经中国医学组织认证的糖尿病防治指南、权威医学教科书等核心文献作为知识来源,确保数据的专业性与可靠性。在此基础上,借助自然语言处理相关技术,从多源数据中精准提取疾病、症状、药物、检查项目、治疗方案等医疗实体,识别实体间的关联关系及实体属性信息,并对提取的知识进行规范化处理与质量校验。最终,将结构化的三元组知识存储于Neo4j图数据库中,形成逻辑清晰、关联紧密的糖尿病领域知识网络,为后续智能问答、临床辅助决策等应用提供坚实的知识支撑。

2.2. 糖尿病智能问答系统设计

为给糖尿病患者提供专业化、一体化的健康管理支持,本研究设计开发了一款融合多技术的糖尿病智能问答系统。系统核心聚焦两大需求:一是通过自然语言交互功能,让患者以日常表述咨询糖尿病相关的疾病知识、诊疗疑问、护理要点等,依托LangChain4J实现精准的语义理解与问答匹配,同时结合构建的糖尿病本地知识图谱,其内容存储在Neo4J中,使用检索增强提高和确保解答的专业性与准确性;二是整合实用化健康管理模块,涵盖个人健康数据记录、个性化用药提醒、科学饮食推荐等功能,全方位辅助患者规范疾病管理流程。技术架构上,系统采用前后端分离模式,前端基于Vue框架,搭配ElementUI组件库快速构建直观友好的交互界面,并通过Axios实现与后端的高效数据通信;后端采用SpringBoot+SpringMVC+MyBatis架构组合,结合MySQL数据库完成业务逻辑处理与数据持久化存储,保障系统运行的稳定性与扩展性。糖尿病问答系统的架构如图1所示。

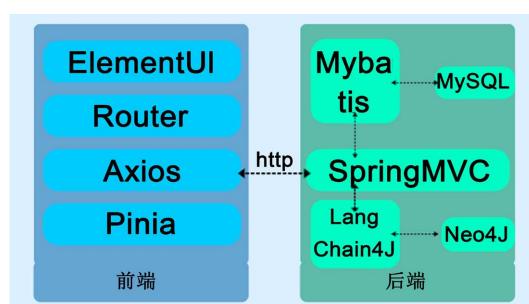


Figure 1. Architecture of the diabetes question-answering system

图1. 糖尿病问答系统的架构

3. 糖尿病知识图谱构建

3.1. 数据获取

糖尿病知识图谱的构建以“权威核心 + 补充拓展”的多源数据采集策略为基础, 确保知识的专业性与完整性。权威知识来源聚焦中华医学会糖尿病分学会发布的系列重磅指南, 包括《中国糖尿病防治指南(2024 版)》《中国 2 型糖尿病防治指南(2020 版)》《中国 1 型糖尿病诊治指南(2021 版)》及《中国糖尿病肾脏病防治指南(2021 版)》, 这些指南涵盖糖尿病分型诊断、用药规范、血糖管理、并发症防治等核心内容, 为知识图谱提供了兼具科学性与临床指导性的核心知识支撑。补充知识则通过网络爬虫技术获取, 具体采用 JavaScript 语言结合 Puppeteer 无头浏览器爬虫框架实现——Puppeteer 作为谷歌官方推出的高级爬虫工具, 支持模拟真实浏览器的渲染与交互行为, 能够有效突破动态网页的数据爬取限制, 避免因页面 JavaScript 动态加载导致的关键信息缺失。爬取过程中, 首先明确糖尿病相关知识的目标数据源, 包括正规医疗科普平台、三甲医院糖尿病专科官网、权威健康资讯网站等, 通过编写针对性的爬虫脚本设定爬取规则; 随后利用 Puppeteer 启动无头浏览器模拟用户访问行为, 自动加载目标网页并解析 DOM 结构, 定向提取糖尿病科普文章、患者护理要点、饮食运动建议、常见疑问解答等补充信息; 同时通过设置合理的请求频率、添加请求头伪装、处理 Cookie 与 Session 等方式规避反爬虫机制, 确保爬取过程的稳定性与合规性; 最后对爬取的原始数据进行初步去重、去噪与格式规整, 与权威指南提取的核心知识形成互补, 为后续知识抽取、融合及图谱构建筑牢数据基础。

3.2. 构建过程以及结果

知识图谱构建先明确核心实体与关系类型, 再通过结构化数据提取、非结构化文本解析等方式采集信息, 经实体消歧、关系校验清洗数据, 最后按图结构存储实体、属性及关联, 形成结构化知识网络。构建流程图如图 2 所示。

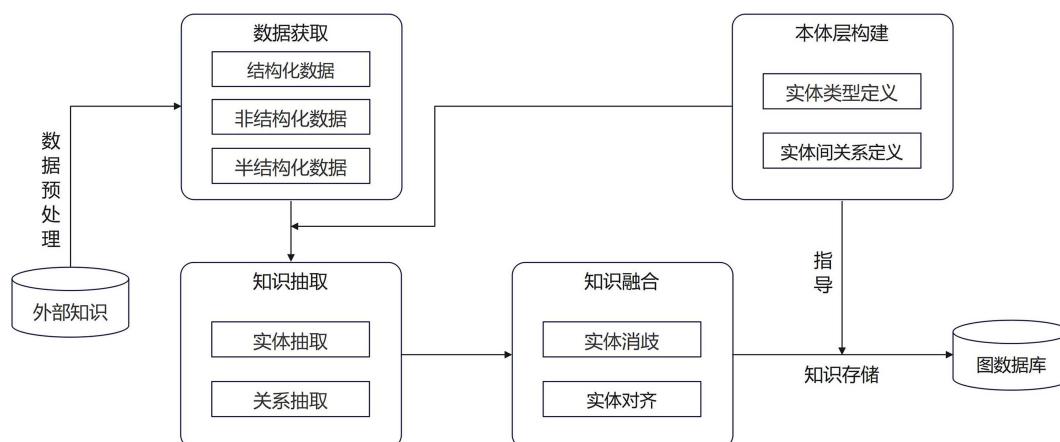


Figure 2. Flowchart of the knowledge graph construction process

图 2. 知识图谱构建流程图

糖尿病知识图谱的构建先完成概念层设计, 明确核心实体, 包含疾病、科室、症状、检查项目、药物、食物等; 实体型, 涵盖疾病类型、症状、并发症、治疗方法、风险因素、检查方法、病理生理; 实体关系, 包含属于、适宜使用、禁忌使用、需做检查、伴随症状、推荐药物、疾病 - 症状。奠定图谱的结构与语义逻辑基础; 随后进入知识抽取与存储阶段, 采用基于深度学习的方法进行命名实体识别, 将关

系抽取转化为关系分类任务以应对文本语义复杂、实体嵌套等问题，再通过基于规则的实体对齐方法完成知识融合，解决不同数据源实体重复或表述不一的问题；最终将提取并整合后的 32,104 个实体、31,242 个关系以 RDF 三元组形式，存储于 Neo4j 图数据库中，形成结构化的糖尿病知识图谱。构建的部分结果如图 3 所示。

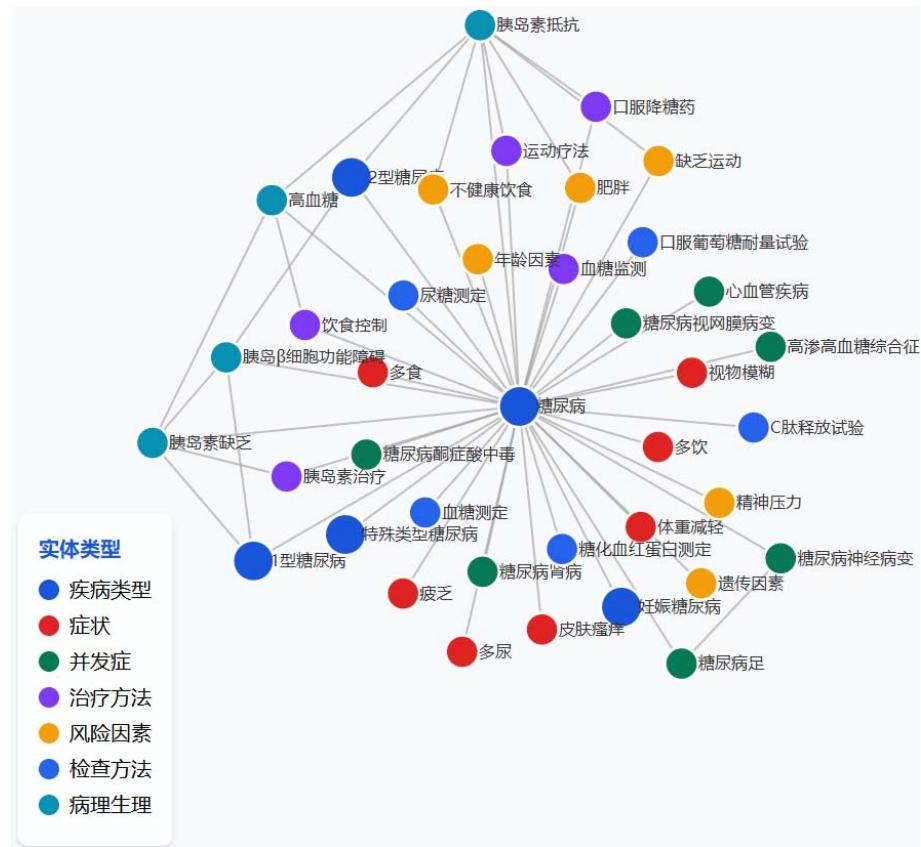


Figure 3. Knowledge graph of diabetes

图 3. 糖尿病知识图谱

4. 糖尿病问答系统的设计与实现

4.1. 问答系统功能设计

糖尿病智能问答系统以患者核心需求为导向，构建了“智能交互 + 多元管理”的一体化功能体系，其核心智能问答模块具备强大的自然语言处理能力，能够精准捕捉患者通过日常口语化表达提出的糖尿病相关疑问，无论是疾病认知、症状咨询、治疗困惑还是健康管理疑问，系统均能依托构建的糖尿病领域知识图谱，该图谱整合了疾病机制、诊疗规范、用药标准、饮食营养等多维度结构化专业知识，结合大语言模型的上下文理解、逻辑分析及自然语言生成优势，经过多层级语义解析与知识匹配，为患者输出兼具专业性、准确性与通俗性的详细解答，既保证医学知识的严谨性，又通过通俗化表述降低患者理解难度，同时支持多轮对话交互，满足患者深层次、连贯性的咨询需求。为进一步拓展系统实用性，适配患者全流程健康管理场景，系统还整合了五大核心辅助功能模块：分类浏览模块围绕患者高频咨询方向，划分为基础知识、饮食管理、运动指南、用药知识及并发症预防五大细分板块，每个板块均梳理了该领域的核心知识点与常见问题，支持患者按分类快速检索目标信息，实现高效信息获取；健康记录模

块包含血糖记录与饮食记录两大功能, 其中血糖记录支持用户自主录入空腹血糖, 系统通过折线图可视化图表动态展示血糖变化趋势, 帮助患者与医护人员直观掌握血糖控制情况, 饮食记录则允许用户自定义添加每日三餐及加餐的饮食内容、摄入量等信息, 实现饮食情况的全面追踪与回顾; 饮食推荐模块聚焦糖尿病饮食管理的关键需求, 一方面通过 AI 个性化饮食推荐功能, 基于患者的血糖水平、体质特征、饮食偏好、并发症情况等个体差异, 利用算法模型生成定制化饮食方案, 涵盖每日热量摄入、营养素配比、食材选择及烹饪方式建议, 另一方面提供食物选择指南, 明确界定适合糖尿病患者的推荐食物与需要严格限制的食物, 为患者日常饮食选择提供清晰、可操作的参考依据; 用药提醒模块则支持用户详细记录所用药物的名称、剂型、剂量、用药频次、服用时间等关键信息, 同时提供灵活的用药提醒设置功能, 用户可根据自身用药方案自定义提醒时间、提醒方式, 系统将在设定时间自动触发提醒, 有效避免漏服、错服等问题, 助力患者养成规律用药的良好习惯。此外, 智能问答模块还配备了直观便捷的智能问答对话框, 支持文字输入、语音输入等多种交互方式, 进一步提升患者使用便捷性, 各功能模块相互协同、数据互通, 全方位覆盖糖尿病患者从知识获取、病情监测、饮食指导到用药管理的全流程健康需求, 为患者提供科学、高效、个性化的健康管理解决方案。系统的功能模块图如图 4 所示。

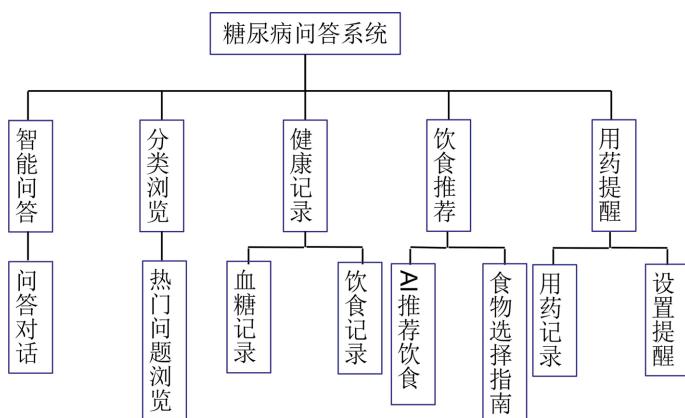


Figure 4. Diagram of the system's functional modules

图 4. 系统功能模块图

4.2. 问答系统前端界面设计

糖尿病智能问答系统前端界面基于 Tailwind CSS 构建, 采用以蓝色为主色调、橙色为辅助色的设计体系, 通过自定义配置扩展了颜色、字体及阴影样式, 确保视觉风格统一且专业, 同时适配用于移动端与桌面端不同设备; 界面整体布局包含顶部固定导航栏, 左侧展示系统 Logo 与名称, 右侧集成智能问答、分类浏览、健康记录等核心功能入口, 移动端设备自动切换为汉堡菜单式折叠导航、渐变欢迎横幅区域, 采用从主色调到信息色的渐变背景, 包含系统欢迎语及快速提问、健康小贴士两个快捷操作按钮、左侧边栏, 展示热门问题列表, 每个问题项配有辅助色图标, hover 时呈现背景色与文字色的平滑过渡效果、中间主内容区, 默认展示智能问答模块, 包含带语音输入与搜索按钮的问题输入框、可折叠的相关问题提示区、用户与智能助手的问答展示区及初始状态提示图, 同时支持切换至分类浏览模块——按基础知识、饮食管理等类别展示带摘要的问题卡片, 卡片采用悬浮动画增强交互体验, 以及健康记录模块。集成基于 Chart.js 的血糖数据图表、血糖记录表格与饮食记录列表; 界面通过自定义 CSS 实现了多项细节优化, 包括玻璃态效果、文字阴影、300 ms 统一过渡动画、自定义滚动条, 滑块采用主色调并支持 hover 变色、打字机效果, 回答区域文字动态输出时的光标闪烁、卡片悬浮时的上移效果、渐变背景及脉冲动

画等, 同时借助 Font Awesome 图标丰富视觉表达, 使整体界面既符合糖尿病健康服务的专业定位, 又通过动态交互元素提升了用户使用体验。系统界面图如图 5 所示。



Figure 5. Illustration of the system's user interface
图 5. 系统界面图

4.3. 问答系统后端数据库设计

糖尿病智能问答系统后端数据库采用关系型数据库设计, 涵盖用户管理、健康记录、饮食推荐和知识库四大模块, 具体包含 users (用户基本信息)、user_profiles (用户详细资料)、blood_glucose_records (血糖记录)、blood_pressure_records (血压记录)、weight_records (体重记录)、medication_records (用药记录)、exercise_records (运动记录)、diet_records (饮食记录)、diet_plans (饮食计划)、diet_plan_items (饮食计划项目)、food_library (食物库)、knowledge_categories (知识分类)、knowledge_items (知识条目)、frequently_asked_questions (常见问题)、system_notifications (系统通知)、user_reminders (用户提醒)等表, 各表通过主键与外键建立关联, 同时在 users 表的 email 和 username 字段、blood_glucose_records 表的 user_id 和 measure_time 字段等关键字段上设计索引以提高查询性能, 并采取密码加密、敏感数据保护、访问控制等数据安全与隐私保护措施, 以及分区表、读写分离、缓存策略等数据库优化建议, 为系统提供全面、安全、高效的数据存储解决方案。数据库表的部分字段内容以及数据库关系图如图 6 所示。

5. 结论

本研究以解决糖尿病患者健康管理知识需求迫切、传统知识获取途径受限及社会医疗资源紧张的问题为导向, 运用 LangChain4J 框架、知识图谱检索增强(RAG)技术、Neo4j 图数据库、Vue 与 Tailwind CSS 前端开发工具、SpringBoot + SpringMVC + MyBatis 后端架构、MySQL 数据库, 以及 JavaScript 结合 Puppeteer 无头浏览器爬虫框架等工具与技术, 首先通过“权威核心 + 补充拓展”的策略构建糖尿病本地知识图谱——以中华医学会糖尿病分学会发布的系列糖尿病防治指南为权威核心知识来源, 以爬虫技术从正规医疗平台采集科普信息为补充, 经自然语言处理技术提取 32,104 个实体和 31,242 个关系, 清洗后以 RDF 三元组形式存储于 Neo4j 数据库; 接着基于该知识图谱与 LangChain4J 框架搭建了糖尿病智能问答系统, 系统采用前后端分离模式, 前端实现多设备适配的交互界面并支持文字、语音输入, 后端保障业务逻辑稳定运行与数据安全, 最终完成了具备“智能交互 + 多元管理”一体化功能的系统开发,

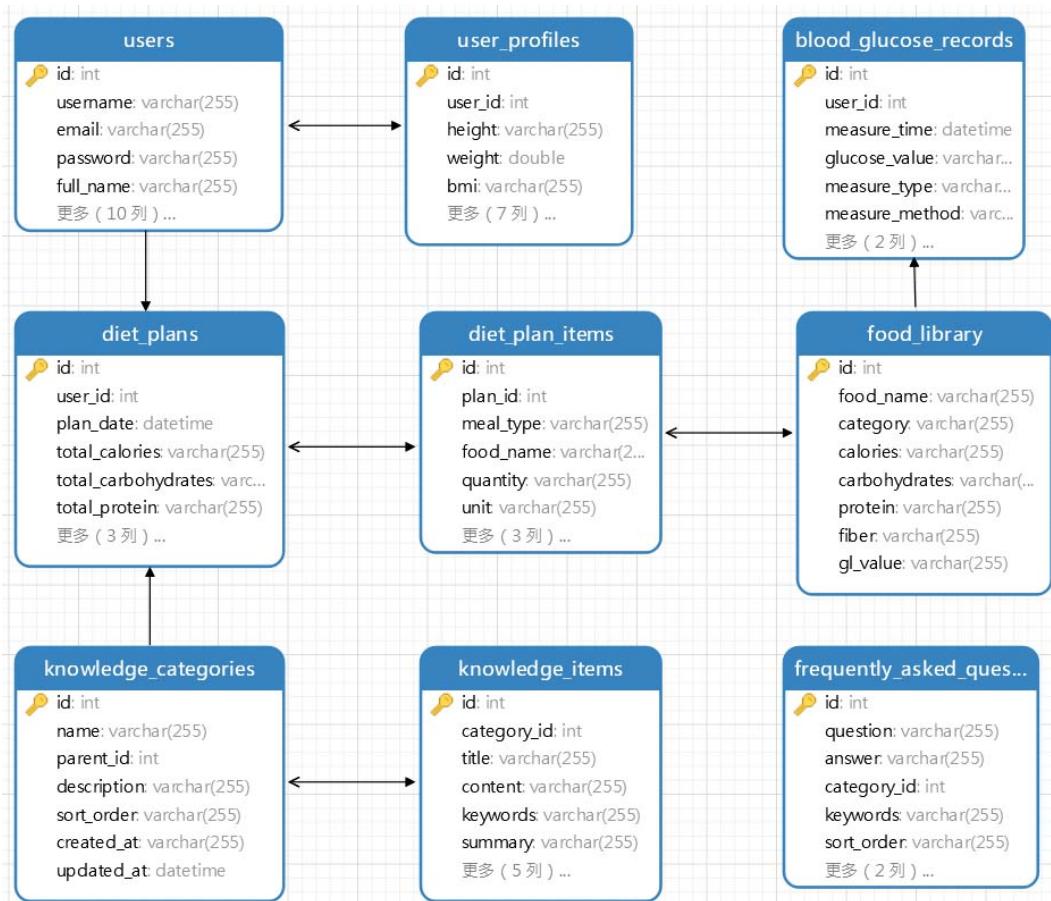


Figure 6. Entity-relationship diagram of the database
图 6. 数据库关系图

其中智能问答模块可精准理解患者口语化提问并输出专业且通俗的解答，还支持多轮对话，辅助功能模块涵盖知识分类浏览、血糖与饮食等健康数据记录、AI 个性化饮食推荐及用药提醒等。该研究的应用价值在于，一方面为糖尿病患者提供了随时获取准确疾病知识、科学管理健康的便捷工具，有助于提升患者自我管理能力，减轻其身体与心理负担；另一方面通过分流部分患者咨询需求，在一定程度上缓解了医疗资源紧张的局面，为大语言模型与知识图谱技术在医学知识问答领域的应用提供了可行的技术方案与实践参考，具有一定的实际应用前景和社会价值。

参考文献

- [1] 中国糖尿病防治指南(2024 版) (节选一) [J].中国实用内科杂志, 2025, 45(10): 864-874+892.
- [2] 王文湖, 韦昌法. 基于大语言模型和知识库的阿尔茨海默病智能问答系统构建研究[J]. 世界科学技术-中医药现代化, 2025, 27(3): 856-866.
- [3] 吴晓臣, 周英凤, 阮玉叶, 等. 妊娠期糖尿病膳食知识图谱的构建研究[J/OL]. 护士进修杂志, 1-13.
<https://link.cnki.net/urlid/52.1063.R.20251111.1605.005>, 2025-11-17.