

---

# 基于大语言模型与知识图谱的国际工程合同问答系统

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/40722.html>

*本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！*

基于大语言模型与知识图谱的国际工程合同问答系统。随着国际工程项目的持续扩展，建筑企业积累了海量的合同相关文本数据，如何从这些密集文本中有效管理和提取知识，对于减少知识流失、保障合同管理效率至关重要。大语言模型的出现为通过智能系统提升合同知识管理提供了前景广阔的途径。然而，幻觉、灵活性不足及可解释性欠缺等问题，往往削弱了从业者将这些模型应用于实际场景的信心。本研究旨在通过融合知识图谱与大语言模型，开发一个面向国际工程合同的知识型问答系统。该系统基于FIDIC 2022版黄皮书与NEC4合同条件中构建的领域知识图谱，利用大语言模型与知识图谱进行协同推理，使其能够结合隐性知识与外部资源回答复杂问题。实验结果表明，所用方法显著提升了模型在合同知识问答任务中的表现，在准确性与可解释性方面平均成功率均超过87%。该模型为国际工程企业提供了一个专业化的问答系统，有助于在合同管理中实现灵活的知识获取与任务导向分析，同时也为人工智能技术融入国际工程合同管理引入了一种新颖的框架。

## 1.引言

在国际工程项目中，合作方通过设计与谈判复杂的国际工程合同来缓解并管控交易风险(Chen等人, 2018)。有效的合同管理对于确保项目成功至关重要，它有助于控制成本超支、减少进度延误并降低索赔与争议的发生(Zheng等人, 2025)。在合同评审阶段，合同管理人员分析大量的合同文件，根据部门相关性对条款进行分类，并进行风险评估、不一致性检测及语义模糊性识别等相关活动。这些流程增强了对合同文件的理解，减少了错误与疏漏(Hassan等人, 2021)。在合同执行阶段，合同管理人员监控可能导致索赔的事件，开展调查、分析证据，并依据合同条款执行索赔程序或启动争议解决流程(Kalogeraki与Antoniou, 2024)。这些合同管理任务大多依赖人工完成，其效果在很大程度上取决于合同管理人员的认知能力，需要丰富的专业知识与专门技能(Candas与Tokdemir, 2022)。

国际工程企业人员的主要知识来源通常包括同事、企业经验及内部文档(Kivrak等人, 2008)。然而，与合同相关的内部文档通常复杂且篇幅庞大。人工从海量文本中筛选以定位具体答案是一项极其耗时的任务。国际工程项目具有固有的临时性与严格的时间限制特点，这阻碍了项目人员之间有效的知识共享(Ren等人, 2018)。此外，对于典型的项目预算而言，咨询外部专家往往成本高昂。虽然在线搜索能力与国际工程项目的的需求高度契合，但所获信息的可靠性与质量常常存疑，极大地限制了其实际效用。

为应对合同知识获取困难的挑战，越来越多的建筑组织正在采用先进的信息技术来优化合同相关

---

知识的管理(Mart í nez-Rojas等人, 2016)。通过利用技术工具提取并系统化组织相关知识, 这些组织能够有效共享知识, 减少操作错误的重复发生, 并最终提升项目绩效与成果(Banerjee等人, 2023)。

当前, 建筑工程合同领域的知识管理主要依赖于结构化知识库。然而, 这些知识库面临若干关键挑战, 例如数据碎片化、过度冗余以及缺乏评估知识表示质量的可靠指标, 这阻碍了知识库中信息检索的精确度(Cerovsek, 2011)。此外, 互操作性仍是一个关键障碍(Lu et al., 2015)。现有的知识管理系统往往缺乏支持基于自然语言的问答交互能力, 其有限的语义互操作性损害了知识检索的准确性、适应性和整体用户体验。提升信息检索系统——尤其是问答框架在精确度和互操作性方面的表现, 对于推进合同知识管理的效能至关重要。

知识掌握一直是人工智能系统演进的基本目标, 已有研究强调了知识表示与逻辑推理不可或缺的作用。在最为复杂的人工智能模型中, 大语言模型展现出卓越的语义理解和上下文学习能力, 使其能够执行复杂推理任务并生成连贯的文本输出(Zhao et al., 2023; Yu and Gong, 2024)。特别是, 大语言模型在革新信息检索系统和增强人机交互方面的变革潜力日益凸显, 使其成为强大的知识库(Fu et al., 2024)。这些模型减轻了通常与构建和维护专业知识库相关的高昂成本和管理负担, 同时通过动态、交互式的问答接口为知识处理与应用提供了更灵活的框架(Wang et al., 2023)。

在建筑工程合同管理领域, 大语言模型作为分析合同数据、获取领域专业知识和优化决策过程的工具展现出巨大潜力(Gao et al., 2025)。然而, 在这一专业领域实施大语言模型驱动的问答系统并非没有挑战。首先, 大语言模型容易产生幻觉现象, 即无法准确回忆或生成可验证的事实, 有时会输出错误或误导性信息(Bang et al., 2023)。此外, 大语言模型在知识范围上存在固有局限, 而重新训练这些模型以整合新信息所需的大量资源加剧了这一问题。虽然微调可以避免全模型重新训练, 但会引入过拟合风险, 可能导致灾难性遗忘的有害现象(Chen et al., 2020)。另外, 大语言模型的可解释性仍是一个重要关切点, 因为这些模型作为黑箱系统运行, 其参数中嵌入的知识不易被辨识(Danilevsky et al., 2020)。这种不透明性阻碍了对大语言模型决策过程的全面理解与控制, 从而限制了其在建筑工程合同管理等敏感领域的实际应用。

知识图谱作为一种精密、精确且可解释的知识表示框架应运而生。合同领域中的知识图谱通过结构化建模增强了合同知识的可验证性。然而, 其基于规则的本质固有地限制了问答体验的灵活性。因此, 认识到知识图谱能够全面可视化知识结构和推理路径, 研究提出将其作为缓解大语言模型局限性的有效方法(Pan et al., 2024)。为增强大语言模型, 研究探索了在预训练阶段融入知识图谱, 从而向这些模型注入领域特定知识(Shen et al., 2020)。随后的进展进一步优化了该方法, 通过促进从知识图谱检索相关实体和关系, 增强了大语言模型的输入提示, 引导其生成更具信息量和上下文准确的输出(Baek et al., 2023)。然而, 尽管存在这些混合方法, 大语言模型在基于知识图谱的直接推理过程中仍处于边缘地位。因此, 大语言模型固有的逐步推理能力在这些系统中未能得到充分利用, 从而限制了其进行更复杂、知识驱动推理的潜力。

为应对上述挑战, 本文提出一种面向国际工程合同知识的先进知识问答系统, 该系统融合知识图谱与大语言模型以支持协同推理。知识图谱通过构建基于图的结构来系统化组织与表征合同知识, 从而封装国际工程合同中的实体及复杂关系。借助知识图谱与大语言模型的协同增强, 该系统有效缓解了大语言模型在工程领域普遍存在的幻觉与灵活性不足等问题。此外, 知识图谱中嵌入的多跳实体与关系路径为大语言模型内的问题分析与解决提供了坚实基础与实质证据, 从而增强了系统逻辑推理过程的解释性。这种协同机制不仅为系统注入了广泛的领域知识, 还提升了其整体推理能力, 使其能够有效处理涉及合同的复杂多维问题。通过系统化构建与高效存储合同知识, 所提出的知识问答系统赋能合同人员精准快速地检索相关知识, 从而深化对合同条款的理解。

---

此外，它还能合同分析、审查及整体管理等关键任务提供辅助支持，最终助力组织优化合同工作流程，并依据具体项目需求提升决策效能。

本文结构如下:第2章概述工程领域知识管理、大语言模型、问答系统及知识图谱的理论基础。第3章阐述国际工程合同知识图谱的构建，以建立全面知识库。第4章通过已构建的知识图谱探索合同知识，并介绍一种融合大语言模型与知识图谱协同推理的知识问答系统。第5章进一步评估所提问答系统的多维能力。最后，第6章总结全文核心发现。

## 2.相关研究

### 2.1 工程领域知识管理

随着信息技术的持续进步及对数据驱动方法依赖性的日益增强，知识库已成为系统化存储与组织信息的高效手段。通过从多元来源提取知识并将其结构化为数据，这些知识库促进了集体归档，从而提升了可访问性、可扩展性与可共享性(Lacosta与Thomas, 2020)。本体作为语义网与知识图谱的基础框架，以其结构精巧性著称，已成为知识表示与传播领域构建知识库不可或缺的工具(Zhang等人, 2015)。其在工程领域的应用尤为显著，涵盖风险与安全管理等方面(Chen等人, 2025)。依据关联开放术语方法论开发的工程安全本体UNOCS，可在施工过程中用于利益相关方间共享安全知识(Speiser等人, 2025)。工程质量保证本体可作为初始本体，结合模型数据、历史质量数据及潜在不符合项，以识别与评估预期质量缺陷(Lü nig等人, 2025)。与此同时，针对工程专用查询语言的研究日益增多，强调知识管理不仅限于异构数据的表示与存储，更关键的是建立关联关系以实现查询、推理与决策等高级功能(Deng等人, 2022)。

知识库的构建主要依赖于历史知识源，用户通过针对特定数据结构定制的查询语言提取预定义内容，由于操作限制和内容适应性的局限，这种方法本质上难以满足国际建设项目的动态复杂知识需求(Liu et al., 2016)。相比之下，迁移学习利用先前获取的知识更高效地解决新颖但类似的挑战，而知识库与迁移学习技术的结合有助于选择适当的知识源并持续更新知识库，从而实现跨项目和项目内的动态知识转移与共享(Pan and Yang, 2010; Xu et al., 2022; Xu et al., 2024)。然而，迁移学习的有效应用需要针对目标项目知识需求进行专门定制的模型训练，这通常需要大量数据支持，在适用性方面存在潜在限制。在此背景下，问答系统作为自然语言处理(NLP)的重要领域，能够响应用户以自然语言表达的查询，相比信息检索系统提供更高的效率和用户友好性(Wen et al., 2022)。借助深度学习这一问答系统发展的关键方法，这些系统显著提升了检索准确性和操作效率，是对传统数据库系统的重大升级(Zhong et al., 2020)。此外，问答系统促进自然语言理解，并支持基于底层知识库的智能知识生成(Tian et al., 2023)。尽管有所进展，当前建设领域的基于知识的问答系统主要通过简单匹配技术处理事实性问题，非事实性问题仍需大量领域知识、深度推理和理解，这仍是重大挑战。

建设领域的知识库、信息检索系统和问答系统，凭借其知识表示、重用和分析能力，对促进整个建设生命周期中的信息交换、知识共享和决策制定具有积极作用。建立建设领域特定知识库和问答系统已成为提升建设企业管理能力的可行策略。

### 2.2 基于大语言模型的问答系统

完成问答任务的基础在于自然语言的表示学习，这是自然语言处理使计算机理解人类语言的过程(Chowdhary, 2020)。随着自然语言处理的进步，问答系统的结构已从传统的三阶段流程——包括问题分析、文档检索和答案提取——演变为更现代的检索器-

---

阅读器模型，其中检索器为给定查询检索相关文档，阅读器从这些文档推断最终答案(Zhu et al., 2021)。生成式预训练语言模型利用Transformer架构，能够在其参数中编码从大规模文本数据中提取的丰富知识，使其无需外部信息检索即可自主生成对查询的响应(Radford et al., 2019)。这些预训练模型在众多基准测试中表现出卓越性能，基于此类模型的免检索方法为问答系统开发带来了新范式(Roberts et al., 2020)。通过采用类似的基于Transformer的架构和预训练目标，当前的大语言模型大幅扩展了模型规模、数据规模和计算能力，从而增强了其容量(Kojima et al., 2022)。在问答任务中，具备上下文学习和提示工程的大语言模型在零样本和少样本学习场景中显著优于其他模型(Alomari, 2024)。

尽管大语言模型在开放领域问答任务中展现出巨大潜力，但其直接应用于受限领域问答任务常面临挑战，主要源于不同领域在对话风格、语言使用和任务目标上的差异(Wang et al., 2023)。一个关键问题是幻觉，这一现象因大语言模型的多功能性而加剧，导致该术语扩展为三种类型:输入冲突幻觉、上下文冲突幻觉和事实冲突幻觉(Zhang et al., 2023)。事实冲突幻觉发生在大语言模型生成与既定世界知识相矛盾的信息时，从而削弱了大语言模型生成内容的可靠性，并对其在专业领域的实际部署构成重大挑战，尤其与其他类型幻觉相比。当处理其专业领域之外的任务或训练语料库中的错误信息被内化时，大语言模型容易产生事实冲突幻觉(Ye et al., 2023)。例如，在问答任务中，大语言模型在缺乏相关知识时倾向于提供不完整或看似合理的答案，而非拒绝回答，这正体现了此问题(Adlakha et al., 2024)。

为提升大语言模型对领域特定应用的适应性，通常采用下游任务学习进行微调。然而，鉴于大语言模型通常包含数十亿参数，其适应专业领域需要海量高质量任务特定数据及大量计算资源，包括高性能GPU或TPU(Dettmers等人, 2024)。此外，大语言模型架构的复杂性使其易受灾难性遗忘(先前习得知识丢失)和过拟合(模型过度适应目标领域)等问题影响，进一步阻碍其在领域特定问答任务中的效能(Ling等人, 2023)。同时，由于非线性数据处理与庞大参数网络间错综复杂的交互作用，大语言模型常被描述为黑箱系统，难以追溯和解释输入如何生成特定输出(Luo与Specia, 2024)。这种透明度的缺失在模型可解释性方面构成重大挑战，削弱了用户对AI驱动问答系统的信任(Zhao等人, 2024)。鉴于这些挑战，亟需开发能融合专业领域知识并针对特定目标优化的领域专用大语言模型。

知识图谱作为理想的知识建模方法，被认为与大语言模型存在互补关系:大语言模型具备优化知识图谱构建的能力，而知识图谱可利用显性知识指导大语言模型训练，增强其知识回忆与应用能力(Yang等人, 2024)。知识图谱对大语言模型的增强主要体现在可解释性、预训练和推理三个关键方面(Pan等人, 2024)。大语言模型的可解释性指理解并解释模型内部工作机制与决策过程的能力。LAMA作为开创性工作之一，通过预定义提示模板将知识图谱中的事实转化为完整陈述，并利用大语言模型预测缺失实体(Petroni等人, 2019)。该方法使得嵌入大语言模型中的知识得以评估。为提升大语言模型对实际知识的理解能力，研究者提出在预训练阶段融入知识图谱的方法。如ERNIE 3.0和K-BERT等模型将知识图谱整合到文本输入中，使模型在初始训练期能访问外部知识，从而提升对事实性知识的理解(Sun等人, 2021; Liu等人, 2020)。

现实世界知识的动态性构成重大挑战，现有方法无法在不重新训练大语言模型的情况下更新已整合知识。因此，研究者提出将知识空间与文本空间分离，并将知识融入推理过程(McCoy等人, 2019)。这通过检索增强生成技术实现，该技术结合外部知识库与信息检索及上下文学习，以提升大语言模型性能，从而避免任务特定的重新训练需求(Fan等人, 2024)。模块化检索增强生成包含针对特定场景定制的检索流程，直接搜索知识图谱等语料库，丰富系统知识语境并强化其知识验证能力(Gao等人, 2024)。Think-on-Graph提出紧密耦合的知识图谱增强大语言模型范式，使大语言模型能在知识图谱上迭代执行束搜索，识别最有希望的推理路径并返回最可能结果(Sun等人

---

，2024)。类似地，MindMap构建提示管道，使大语言模型通过整合隐性知识与检索的外部知识来理解知识图谱输入并进行推理(Wen等人，2023)。Graph RAG采用两阶段顺序方式，利用大语言模型从紧密关联的实体群构建基于实体的知识图谱和社群摘要。通过基于图谱检索文本索引，显著提升生成答案的全面性与多样性(Edge等人，2024)。

这些进展推动了法律、医疗、金融、工程等广泛领域的大语言模型与知识图谱驱动问答系统的发展。典型案例如水工结构安全管理问答的知识图谱增强大语言模型，以及集成领域特定多模态知识图谱的建筑工程管理问答大语言模型(Zhang等人，2025；Zhou等人，2025)。

凭借语言理解与生成能力，大语言模型已成为领域问答系统阅读器的首选模型。将上下文文档之外的领域特定知识与目标问题相结合，是增强问答系统的关键。针对大语言模型的领域知识表示与领域知识增强研究已展开，旨在提供多源多结构数据作为领域特定知识的补充。

### 2.3 施工合同知识图谱与知识表示

人类知识的表示与推理是人工智能领域的一个基础方向(Newell等人，1959)。知识图谱作为一种结构化框架，用于表示不同现实领域中的实体及其关系，促进实体间的互联，并使推理系统能够推导出新知识(Wang等人，2017)。从图结构角度看，知识图谱可视为一个图；当涉及语义时，则可视为本体与语义网的有机结合，包括基于本体的模式层和能准确、可解释地表示领域知识的数据层(Ji等人，2022)。知识图谱能表示不同形式的知识，包括事实型与非事实型知识。事实型知识通常涵盖具有完整语义的通用人类知识，可直接以三元组(主体、谓词、客体)形式从文本中提取(Ehrlinger和Wöβ，2016)。相比之下，非事实型知识涉及动态、事件驱动的信息，具有复杂关系，常通过事件知识图谱表示(Ding等人，2019)。这些图谱将事件和实体建模为节点，并通过边捕捉它们之间的复杂关系——如时间、因果和条件联系——从而能够追踪事件随时间的演变与逻辑(Knez和?itnik，2023)。一项基于BERT-Att-BiLSTM-CRF模型开发了针对大型水电项目建设安全隐患的知识图谱，涵盖九类实体和八类与安全隐患相关的关系(Yang等人，2025)。Shaw等人(2025)构建了一个基于本体的原型知识图谱，以支持建筑环境中的生命周期资产管理。

建设合同通常采用分层结构，合同知识嵌入在实体、事件及其复杂关系(如时间、条件和因果关系)中，并通过形式逻辑表示，但这种方法往往缺乏高效检索的可扩展性。知识图谱通过基于模式层的标准化概念，提取并组织合同中的实体、事件和关系，从而克服了这一限制，实现了对合同知识更高效的检索与推理(Aghaei等人，2022)。在开发知识图谱模式层的背景下，El-Diraby等人(2005)提出了一个建设知识领域本体，将概念分为七类:过程、产品、项目、参与者、资源、技术主题和系统。类似地，Niu和Issa(2015)设计了概念化与形式化的本体模型，以表示AIA A201合同索赔相关的特定知识。在知识图谱的数据层，Zheng等人(2023a)利用知识图谱的嵌套结构，基于NEC合同构建了本体层和嵌套图，并应用这些知识图谱来增强合同审查。尽管已有研究开始探索使用知识图谱提取合同中的复杂关系，但在上下文中表示这些复杂关系的规范化方法仍不完善。

## 3.知识图谱构建

工程合同中的知识可通过知识图谱有效捕捉，这些图谱表示合同中嵌入的实体和复杂关系。本节概述了国际工程合同知识图谱的开发过程，主要包括三个步骤。首先，收集与国际工程合同相关的数据，这些数据为后续步骤奠定基础；其次，创建定义合同关键概念及其关系的本体，作为知识图谱的模式层，确保标准化与结构清晰。最后，基于模式层提取实体和关系，构建知识图谱的数据层，并存储于Neo4j数据库中，以支持高效查询。

---

### 3.1 源数据:标准国际工程合同

数据是构建高质量知识图谱的基础，因此需要选取具有代表性的合同文本来构建知识图谱。为确保跨项目的适用性和普适性，专业机构已开发出国际工程合同标准化模板，并在行业内广泛应用(Elkhayat and Marzouk, 2022)。项目参与方可修改模板中的部分内容以满足特定需求，而其他条款保持不变(Choi et al., 2013)。常见模板包括土木工程师学会(ICE)发布的新工程合同(NEC)系列和国际咨询工程师联合会(FIDIC)合同系列。NEC与FIDIC具有不同的发展历程和主要业务领域。NEC系列合同的最新版本NEC4于2017年发布，提供了适用于各类建筑项目的法律框架和项目管理程序，在英国及英联邦国家广泛应用。FIDIC合同因其科学性、公平性和严谨性在全球范围内被广泛接受，并得到世界银行、亚洲开发银行等国际金融机构的认可和使用。2022年版FIDIC黄皮书是对2017年版的修订，适用于设计-建造(DB)合同模式。鉴于NEC4和FIDIC黄皮书在全球各类工程项目中的广泛应用，本研究将这两种代表性合同模板作为构建国际工程合同本体与知识图谱的基础。

### 3.2 本体开发

本体构建应遵循清晰性、一致性、可扩展性、最小编码偏差和最小本体承诺等原则(Gruber, 1995)。经典开发方法包括TOVE法、骨架法、IDEF5法、METHONTOLOGY法和七步法(Uschold and Grüninger, 1996)。基于METHONTOLOGY方法，结合骨架法与七步法的优势，设计出适用于国际工程合同的定制化本体开发方法。具体流程如图1所示。

---

## 图1. 本体开发流程

在与合同专家协作下，明确了本体应解决的能力问题，并确定了知识来源，包括标准化术语和非结构化的国际工程合同文本数据。从数据源提取相关术语并复用现有构建成果形成归属列表，确保构建过程的质量与效率(Niu and Issa, 2015)。厘清了概念间关系，包括构建类层次结构和识别概念间关联。附录图A1展示了国际工程合同本体的主次类别层次关系，如"环境"、"承诺"、"资源"、"参与者"、"过程"、"行动"和"产品"。

选用Protégé作为本体开发环境，建立初始本体层次结构，并将术语表中的关键术语作为实例输入Protégé。通过基础评估和迭代改进验证了本体的实用性与准确性。利用Protégé中的Hermit推理机检验本体一致性，确保其内在属性与断言无矛盾。继3.3节知识图谱构建后，通过在实例层面填充内容验证了本体框架。迭代开发的国际工程合同领域本体框架见附录表A1。

### 3.3 知识图谱构建

知识图谱是一种大型语义网络，其核心在于表示实体间的关系。构建知识图谱本质上是构建三元组，即由主语、谓语和宾语构成的结构。主语和宾语是知识图谱中的实体，而谓语描述主语与宾语之间的关系或属性。合同知识主要涉及非事实性知识，包含各种复杂关系，如条件关系、时间关系和因果关系。图2展示了NEC4合同条件第9条中的一句话，其中包含连接<项目经理、签发、终止证书>与<各方、执行、终止程序>的如果-那么关系。若不对此如果-那么关系进行建模，该条件关系无法通过离散的三元组表达。

尽管近期出现了关于非事实性知识建模的研究，如事件知识图谱和"三元组即节点"概念，但在合同语境中表征此类知识的规范化方法仍然缺失(Guan et al., 2022)。因此，在提取三元组前，需要构建能够表达复杂合同关系的方法。建立了合同分析方法的系统框架，合同条款三元组提取方法总结于表1。

特别是为了捕捉合同中蕴含的复杂关系，本研究定义了一组关系属性，这些属性保留了精细的交互逻辑，如表2所示。这些属性以边属性的形式存储，可在Cypher查询中直接过滤或匹配。该方法使得查询不仅限于"是什么"行为，还能检索其"在何种条件下、在什么时间内、有何例外情况以及出于何种目的"适用。

由于合同的复杂性，人工提取三元组的结果可能存在差异。本研究制定了实施规则以规范提取过程中的操作，详见附录表A2。如附录图A2所示，以NEC终止条款为例，演示了基于合同条款通过实体识别和关系提取形成三元组的具体流程："若任何一方希望终止承包商提供工程的义务，应通知项目经理及另一方，并说明终止理由的详细信息。"该流程以该条款的应用为起点展开。随后，对人工提取的三元组进行了数据格式优化，包括将中文标点替换为英文标点、关系名称首字母大写、属性标签用双引号括起，并核验节点属性的正确位置。

图2. NEC4合同条件中"假设-则"关系示例

表1. 国际工程合同知识图谱三元组提取方法

| Extraction step                          | Description  |
|--|--|
| Syntactic analysis of contract sentences | Analyze the contract clauses to identify their grammatical components, including conditional adverbial clauses, subjects, predicates, objects, and other key elements.                             |
| Extraction of key triples                | Extract the most critical triples, usually the subject-predicate-object structure of the main clause. Ensure the extracted triples can express the main actions and involved objects.              |
| Supplementing related Information        | Starting from the key triples, analyze the sentence's logical structure to find other related triples. Supplement relevant details such as prerequisites until all information is fully expressed. |
| Checking and explaining New Nodes        | Check whether the extracted triples introduce new concepts or objects, determine if further explanation or clarification of newly introduced nodes is needed to ensure the logical consistency.    |
| Ontological classification of Nodes      | Classify all extracted nodes according to their nature and relationships to ensure information conforms to the structured ontology, facilitating system organization and retrieval.                |

表2. 国际工程合同知识图谱的关系属性

| Relationship attribute | Description  |
|------------------------|--|
| Time                   | Temporal constraints or deadlines                              |
| Condition              | Triggering events or situational requirements                  |
| Exception              | Exceptions or alternative cases                                |
| Object                 | The entity or artefact that an action is performed upon        |
| Purpose                | Intended outcomes or goals                                     |
| Meaning                | Definitional clauses linking terms to their conceptual meaning |

最终，NEC4合同条款形成了1,559个三元组，而2022年版的FIDIC黄皮书则形成了2,234个三元组。这些三元组构成了国际工程合同知识图谱的基础数据，为后续的本体细化和知识图谱应用提供了支持。

通过知识图谱利用以三元结构化实体-关系路径表示的知识，其基础在于能够以图形化方式高效存储和检索。Neo4j被选为国际工程合同知识图谱的存储引擎。在初期人工提取过程中，结果以Excel格式呈现，随后通过Python批量导入Neo4j数据库。Neo4j Browser作为图数据库的可视化工具，为用户提供了直观的数据探索与分析界面。图3(a)展示了国际工程合同知识图谱的可视化结果，包含3,150个节点和3,812个关系。节点的不同颜色代表不同类型的实体，例如绿色节点表示"执行者"类型实体，蓝色节点表示"行为"类型实体。节点间的边表示实体间的关系，边的密度反映了关系的复杂程度。

通过对知识图谱的可视化分析，可以深入探索图谱中实体间的联系与作用。核心实体通常位于图中心位置，如"承包商"、"工程师"和"业主"，这反映了它们的重要性和复杂关系，如图3(b)所示。外围节点通常涉及合同条款的解释和补充信息，如图3(c)所示。此外，利用Neo4j的Cypher查询语言对构建的知识图谱进行简单查询和多种组合查询，确保了对图库中数据的灵活检索与分析。

#### 4. 问答系统构建

本研究提出了一种利用知识图谱增强大语言模型的新方法，使二者在国际工程合同问答系统中协同工作。该问答系统利用大语言模型作为阅读器，凭借其强大的语义理解和上下文学习能力，克服传统基于知识的问答系统对模板的依赖性和互操作性低的问题，同时基于知识图谱提供的领域知识，缓解大语言模型的幻觉、僵化和黑箱问题，从而增强问答系统的深度推理能力和可解释性。本研究提出的问答系统可分为三个部分：

1) 三元组向量数据库开发。该组件利用词嵌入技术将知识图谱中的实体和关系嵌入到向量数据库

---

中，便于将用户查询与知识图谱中的实体和关系进行匹配。

2) 合同知识探索。此过程对用户查询进行实体识别和关系抽取，并通过余弦相似度比较定位初始实体。确认初始实体后，启动知识探索过程，在知识图谱中挖掘针对特定问题的推理路径，从而构建合同子图。

3) 答案生成。此过程采用设计的混合提示模板，使大语言模型能够理解并利用用户查询和合同子图进行推理，最终生成最终答案。

系统的架构设计、模块功能及其协同工作流程如图4所示。

图3. 国际工程合同知识图谱可视化及部分节点示例

图4. 整体构建过程

#### 4.1 三元组向量数据库

词嵌入是一种将词汇表中的单词或短语映射到实数向量的技术(Mikolov等人，2013a)。通过将词汇映射到低维连续向量空间，每个单词或短语被表示为实数域上的一个向量。词嵌入是自然语言处理下游任务的重要基础，它将自然语言的语义信息嵌入到数值向量中，使计算机能更好地理解 and 处理文本数据。自然语言处理领域的许多革命性成果都源于词嵌入的发展，例如Word2Vec、GloVe和BERT(Mikolov等人，2013b；Pennington等人，2014；Devlin等人，2019)。

利用词嵌入技术将知识图谱中的实体和关系嵌入到向量数据库中，便于将用户查询问题与知识图谱中的实体和关系进行匹配，从而为用户提供准确且相关的答案。

在此背景下，本研究特别关注词嵌入模型的余弦相似度性能，这是衡量两个语义向量间相似度的重要指标。所选用的词嵌入模型为multi-QA-mpnet-base-dot-v1模型。该模型是基于Transformer架构的sentence-transformers模型，能将句子和段落映射至768维稠密向量空间。它专为语义搜索任务设计，并已在包含2.15亿(问题，答案)对的多元数据源上进行训练。该模型的优势在于其捕捉文本上下文信息的能力，通过多层自注意力机制和全连接网络有效提取输入文本的语义信息，生成相应的语义向量。此外，multi-QA-mpnet-base-dot-v1模型支持多种损失函数和优化方法，使其能适应不同任务和数据特性，提升模型性能与泛化能力。该模型强大的语义表征能力和广泛的语料

---

训练为构建向量数据库提供了可靠的基础支撑。

## 4.2 合同知识探索与推理子图构建

在开始知识探索前，必须确定指导探索过程的核心实体与初始实体——这些是推理路径的起始节点。给定查询时，通过提示大语言模型进行实体识别与关系抽取，以提取核心实体、关系及属性，并通过向量嵌入将查询实体、关系和属性与知识图谱中的实体、关系及属性进行链接，从而在知识图谱上定位初始实体集合。具体而言，使用multi-QA-mpnet-base-dot-v1模型对查询实体、关系和属性进行编码，获得查询嵌入向量。接着，计算每个查询实体、关系、属性与三元组向量数据库中每个实体、关系、属性之间的余弦相似度。通过比较这些余弦相似度，为每个查询实体选取三元组向量数据库中余弦相似度最高的实体作为最近邻，形成用于知识探索与推理子图构建的初始实体集合。

在确认初始实体后，启动知识探索过程以揭示知识图谱中针对特定问题的推理路径，从而构建合同知识子图。该探索采用基于实体关键词的知识图谱信息检索过程，涵盖基于路径的探索(Wen等人，2023)和基于邻域的检索(Soman等人，2024)两种方式：前者利用通过多跳关系路径连接起始与终止节点，后者聚焦于直接相邻的关系与实体。在整个过程中，各阶段均应用剪枝与排序策略以提升效率与相关性。本研究的知识探索主要由这两种方法构成：基于路径的探索和基于邻域的探索。基于路径的探索主要关注由多个实体和关系构成的扩展推理链的遍历，为合同子图提供深度推理逻辑，从而增强问答系统的整体推理能力与可解释性。相比之下，基于邻域的探索侧重于通过聚类与初始节点紧密关联的邻域实体来扩展推理网络，通过整合密集领域特定知识来丰富合同子图，从而提升问答系统的准确性及其总结能力。

基于路径的探索是一种关系驱动的方法，旨在探索知识图谱内初始实体间的关联连接，以构建用于问答的深度合同子图。该过程首先从初始实体集合中选择一个初始实体作为探索的第一个起始节点，其余初始实体则作为候选实体构成候选实体集合。依次选取候选实体集合中的每个实体作为目标节点，从起始节点向外发起实体-关系探索，以在预定义跳数(记为 $K$ )内识别连接起始节点与目标节点的所有可能路径，并存储这些路径。若起始节点与目标节点间未发现路径，则搜索下一个目标节点。理论上，此过程持续进行直至提取出以初始实体为起始节点、各候选实体为目标节点的所有路径，从而完成对第一个初始实体的知识探索。一旦某初始实体被用作起始节点，它将在后续探索中被排除在候选实体集合外，以避免生成冗余路径，从而优化探索效率。随后对每个初始实体依次执行上述探索过程，生成连接任意两个初始实体的推理子图。通过这两阶段的知识探索，知识图谱中初始实体间所有潜在的关联路径得以全面提取。

基于邻域的探索是一种实体驱动的方法，它探索知识图谱中初始实体的相邻关系，旨在构建面向广度的合同子图以用于问答。在此过程中，初始实体集中的每个实体进行一跳探索，识别通过单一关系链接与初始实体直接相连的相邻实体，并生成连接初始实体与相邻实体的所有路径。这种探索获取与初始实体相关的所有上下文三元组，生成捕获每个实体长度为一级关系的推理子图。然而，某些核心节点——如承包商、工程师和客户——往往拥有大量相邻节点，直接将其选为起始节点可能导致与查询的相关性降低，使知识探索效率低下且意义有限。因此，在知识探索中特意排除具有广泛关系连接的核心节点，以提升基于邻域探索的整体效率，并确保生成的合同子图与查询保持相关。

由于向量数据库中三元组数量庞大，通过知识探索生成的合同子图可能包含非必要或语义重复的实体和关系以及冗余路径，这可能导致合同子图超出大语言模型的最大窗口限制，降低探索效率并增加计算成本。为解决此问题，需要对合同子图进行剪枝。剪枝过程首先识别知识图谱中关联

---

三元组超过预设阈值的初始节点实体，然后将这些三元组分组为待剪枝的候选集。接着，对与同一初始实体关联的每组三元组进行余弦相似度比较，并丢弃余弦相似度高于指定阈值的三元组。通过合并剩余的子图，获得最终的合同子图，确保其既全面又简洁，从而为后续处理提供坚实基础。

上述知识探索策略及合同子图的后续精炼，为推理和生成的后期阶段提供了关键支持，促进了对合同知识的更深入理解和更有效应用。基于路径和基于邻域的探索方法的优势在于能够有效整合用户查询中的关键信息，构建兼顾深度与广度的相关合同子图。这种方法确保问答系统获得更精确、更相关的输入，最终提升系统性能及整体应用效果。

#### 4.3 答案生成

知识探索后，利用合同子图提示大语言模型进行协同推理，生成最终输出。创建结构化的提示模板，包含系统消息、用户消息和AI消息。这些消息通过LangChain技术集成，形成完整的提示。系统消息提供必要的背景信息和指导，用户消息呈现查询以引导推理过程。AI消息将用户查询与推理子图结合，提供必要的推理材料。随后将AI消息传递给大语言模型，由其处理并生成详细、准确的推理结果，最终向用户交付答案。

#### 4.4 实际应用

终止条款在合同中具有关键意义。NEC4合同第9条终止概述了22类终止原因、4种终止程序、4类应付金额及其对应关系。以下部分通过合同终止场景全面演示集成知识图谱与大语言模型的推理过程，并将其输出与独立大语言模型的结果进行对比。本演示采用GPT-4作为大语言模型，具体工作流程如图5所示。

用户查询为在NEC4合同中，若业主因承包商实施腐败行为终止合同，终止时业主应支付哪些金额？。GPT-4从用户查询中提取的信息包括若业主因腐败行为终止合同和终止时业主应支付的金额。通过向量数据库匹配获取知识图谱中的本地化属性和实体，包括若承包商实施腐败行为.....、Reason22、应付金额和业主。随后对知识图谱进行探索，其中实线代表基于路径的知识探索，虚线表示基于邻域的知识探索。提示模板使GPT-4能够通过用户对用户查询和检索到的合同子图进行联合推理，生成最终答案。

对于未经知识图谱增强的GPT-4模型，其回应通常始于列举一般性考量并提供解释。这些内容通常涵盖已完成工作的付款、费用报销、禁止为腐败行为支付款项、损害赔偿、罚款或罚金、扣留付款以及争议解决等方面。此外，GPT-4明确强调腐败行为可能带来的法律后果——包括刑事指控、民事处罚以及被禁止参与未来合同——并建议业主咨询法律顾问以确保遵守适用的法律法规。然而，需要注意的是，尽管GPT-4的回应表现出全面性，但其广泛的范围可能使它们对合同新手或管理者而言过于笼统。此类回应往往缺乏针对性，未能解决具体的实际问题。相比之下，由知识图谱增强的大语言模型生成的回应则表现出更高的精确度。通过成功检索关于因承包商腐败而终止合同后的应付款项及相关终止程序的具体合同条款，该模型准确回答了关于应付金额的查询。关键的是，它避免了包含问题范围之外的无关内容。

---

## 图5. 知识图谱提示的GPT-4与未提示的GPT-4回答应用示例

大语言模型在特定领域问答任务中的准确性和可解释性通常有限。相比之下，我们提出的知识图谱增强大语言模型方法通过鼓励大语言模型将自身知识与检索到的推理子图相结合，从而提升性能。这种协作推理过程为回答问题提供了明确且全面的推理路径，进而提高了问答系统的准确性、全面性和可解释性。

### 5. 实验与评估

通过一系列需要复杂推理和领域知识的任务，对国际工程合同问答系统的性能进行了评估，并与基线模型进行了比较。

#### 5.1 数据集

在问答数据集中，问题通常根据预期回答的类型进行分类，一般区分为事实型问题与非事实型问题(Cortes等人，2022)。事实型问题直接且基于事实，通常需要以短短语或句子的形式提供简要答案。相比之下，非事实型问题是开放式的，需要复杂而详尽的回答，例如假设型问题、因果问题、确认性问题以及总结性问题(Mishra和Jain，2016)。非事实型问题通常具有更高的复杂性，需要高级的自然语言理解和推理能力，因此成为评估的重点。尽管存在多个用于非事实型问答系统的公共数据集，如SQuAD、MS MARCO、WikiQA以及Natural Questions语料库，但这些数据集主要针对开放领域问答任务设计，且基于与文本段落相关的问题回答(Kwiatkowski等人，2019)。这些数据集均不适用于我们的实验。因此，我们创建了自己的数据集，专门针对国际工程合同知识问答主题定制。

本研究以2017版FIDIC黄皮书和NEC4合同条件为主要数据源建立评估数据集，手动整理了一套结合多种问题类型的任务。具体而言，通过单项选择题设计了一项客观问答任务，要求模型从四个

---

选项中选择唯一最合适的答案。该任务旨在评估模型对国际工程合同中基本要素的掌握程度——包括主体、客体和情境因素——涵盖NEC4的所有核心条款以及2022版FIDIC黄皮书的所有一级条款。作为补充，一项主观问答任务采用开放式问题，要求提供准确、专业且全面的回答，以模拟真实的合同管理场景。该任务评估模型对国际工程合同知识的综合掌握能力，涵盖NEC4核心条款、可选条款及成本构成，以及2022版FIDIC黄皮书的所有一级条款。本数据集旨在同时反映条款层面的完整性和认知层面的多样性。2017版FIDIC黄皮书的每个一级条款以及NEC4的每个核心条款至少对应两个问题，确保没有条款被孤立评估。对于涉及索赔和变更等关键问题的条款，开发了约五个不同的问题以捕捉解释和应用的不同维度。通过结合客观和主观问题形式，数据集有效评估了基础知识和高级推理技能，支持对模型在实际合同场景中性能的全面评估。

所有问题由两位具有国际工程合同学术背景的工程管理专业硕士研究生制定，并严格遵循源文件。每个问题根据三个标准——清晰性(无歧义)、相关性(与具体条款的关联)以及实用性(反映真实合同管理需求)——进行严格评估，按1-5分评分。仅在所有标准得分均高于4分的问题被选中。最终整理的测试集包含158个问题，包括58个客观选择题和100个主观开放式问题。

## 5.2 评估指标

以往建筑领域的问答系统主要强调正确性作为关键评估标准，常用指标包括准确率、精确率、召回率和F1分数。对于有明确正确答案的评估任务，如选择题和填空题，准确率作为量化指标能够清晰比较模型性能；因此，本研究采用准确率作为客观问答任务的评估方法。对于主观问答任务，单一的正确性指标不足以全面评估其在问答任务中的深度推理、可解释性和总结能力。因此，本研究提出一套多维评估指标，涵盖输出质量的多个方面，聚焦于准确性、简洁性、全面性和可解释性，以更全面地评估模型性能。四项评估指标的定义和标准如下：

- 1) 准确性:模型的回答应与国际建筑合同条款和行业标准一致，无事实错误或误导性内容。
- 2) 简洁性:模型的回答应以最少必要信息解决问题，减少冗余细节。
- 3) 全面性:模型的回答应涵盖问题的所有核心方面，无关键信息遗漏。
- 4) 可解释性:模型的回答应呈现清晰的逻辑链和支持证据，使用户更容易理解推理过程。

为细化这些指标并减少评估偏差，本研究进一步为每个维度制定了详细的评估标准(如表3所示)，确保评估过程的一致性。

## 5.3 评估方法

问答任务中大型语言模型主要有两种评估方法:自动评估和人工评估(Chang等人, 2024)。自动评估通常依赖标准指标和工具，如ROUGE和BLEU，通过比较模型输出与参考答案来评估性能。然而，由于国际建筑合同问答任务中的问题多为开放式且缺乏固定参考答案，自动评估不完全适用于此场景。在自动评估不足的情况下，人工评估——即评估者对模型回答提供反馈——更贴近实际应用场景。但人工评估可能资源密集、成本高且耗时，也容易因评估者的文化和个人差异导致不一致，需要其具备专业领域知识。

表3. 评估指标详情

| Dimension         | Explanation   |
|-------------------|---|
| Accuracy          | alignment with the question's core requirements<br>complete and correct addressing of key elements<br>use of precise terminology avoiding ambiguity |
| Conciseness       | employment of concise syntax for direct response delivery<br>exclusion of irrelevant content<br>elimination of redundant information                |
| Comprehensiveness | coverage of all pertinent contractual clauses<br>inclusion of typical applications and edge cases   |
| Interpretability  | provision of evidential support and logical reasoning<br>structured step-by-step explanation  |

随着大语言模型的持续进步，其在许多任务中展现出超越人工标注的潜力(Gilardi等人，2023年)。一些研究提出LLM-as-a-Judge方法作为传统人工评估的有前景的替代方案，减少人类参与需求并提供解释(Zheng等人，2023b年)。尽管LLM-as-a-Judge方法的使用日益增多，但由于存在多种偏见(如位置偏见、冗长偏见和情感偏见)，对大语言模型可靠性的担忧依然存在(Ye等人，2024年)。位置偏见指大语言模型倾向于偏好特定位置的回答，且随着潜在答案数量增加，这种偏见更为明显。本研究中的评估仅涉及回答的成对比较，受位置偏见影响较小。冗长偏见指某些模型表现出偏好与回答长度正相关的现象。本研究的模型评估指标包含简洁性和全面性，能在一定程度上缓解回答长度对评估结果的影响。情感偏见指大语言模型倾向于选择不含情感元素的内容。本研究的问答系统和模型较少涉及基于情感分析的问题，情感偏见影响甚微。Chen等人(2024年)探究了人类评估者与大型语言模型评估者的多重偏见，发现所有评估者都存在显著偏见，但具体倾向存在差异。

基于这些考量，本研究采用LLM-as-a-Judge方法进行主观任务评估。GPT-4在学术界和工业界广受认可，因其能提供与人工评估高度吻合的准确评价，常被用于评估模型生成回答的质量(Huang等人，2024年)。本研究特别采用成对比较方法:针对每个合同相关问题，基线模型与知识图谱增强模型分别生成回答，随后由GPT-4O根据评估指标细则进行评判，并计算知识图谱增强模型的胜率、平局率和败率。这种全面多维度的评估为国际工程合同问答系统的性能提供了宝贵洞见，并为模型优化和实际应用提供了关键指导。

#### 5.4 基线模型

为全面系统地评估国际工程合同问答系统，本研究选取了一系列大语言模型作为基线模型，包括GPT-4、GPT-3.5和GLM-4。基线模型与知识图谱增强模型分别回答测试问题，通过结果对比验证模型有效性(OpenAI等人，2024年)。

- 1) GPT-4由OpenAI于2023年发布，主要应用于自然语言生成、翻译、文本摘要等领域。其创新点在于更深入的理解与生成能力，在多项自然语言处理任务中表现卓越。
- 2) GPT-3.5是GPT-3的改进版本，由OpenAI于2022年发布。该模型在处理复杂自然语言任务时表现优异，尤其在对话系统和文本生成任务中。
- 3) GLM-4是智谱AI于2024年推出的新一代基础模型，在语言理解、文本生成及问答系统等多项任务中均表现出色。

## 5.5 评估结果

表4展示了三个基线模型及其对应知识图谱增强模型在客观任务中的表现。

在基线模型的对比分析中，GPT-4在客观任务上表现出最强性能，准确率达到81.03%。此外，对比基线模型与其知识图谱增强版本时，增强模型在客观任务性能上均呈现持续提升，在所评估的基线模型中准确率平均提高约5%。

表4. 客观任务实验评估结果

| Model   | Original LLM | LLM with KG |
|---------|--------------|-------------|
|         | Accuracy (%) |             |
| GPT-3.5 | 67.24        | 72.41       |
| GPT-4   | 81.03        | 86.21       |
| GLM-4   | 65.52        | 70.69       |

如图6所示的主观任务总体评估结果表明，利用知识图谱增强大语言模型的问答系统在多数能力维度上表现出色，在准确性与可解释性方面均显著超越基线模型。

## 讨论

以下是根据主观任务实验结果得出的主要观察与结论。

知识图谱通过条款级知识与结构化推理提升答案准确性。我们的知识图谱提供了针对国际工程合同定制的细粒度领域知识，使大语言模型在生成答案时能够参考明确条款内容、结构化关系及跨文档依赖。这种结构使模型能定位更精确的答案来源并避免事实性错误，从而推动观测到的准确性提升。例如，经知识图谱增强的GPT-3.5在准确性上相比未增强版本取得了97%的胜率，这归因于知识图谱能够补充基础模型有限的领域覆盖度，并提供包含复杂关系的推理信息链。不同大语言模型间的效果差异可能源于其基础模型训练数据与架构的差别。GPT-3.5与GLM4的基础训练数据量相对较小，因此集成知识图谱后的提升更为明显；而GPT-4本身已具备较高的知识覆盖度，故提升幅度相对较小。与对知识图谱效益的普遍预期不同，本研究在高度专业领域提供了可量化的证据，展示了结构化条款级知识如何直接支持正确答案生成，尤其对训练语料较小的模型效果显著。

模型回答更简洁，但覆盖度提升有限。例如，经知识图谱增强的GPT-4在简洁性上达到96%的胜率，优于其他知识图谱增强模型。然而其完整性胜率降至32%，表明相比未增强的原始GPT-4有所下降。这种权衡可能反映了GPT-4本身较强的基线覆盖能力：知识图谱通过过滤冗余内容帮助精简回答，但可能无意中遗漏某些相关细节。相比之下，GPT-3.5与GLM-4因领域知识较弱，在简洁性与内容丰富度方面均有更大提升空间，故从知识图谱增强中获益更均衡。这些发现凸显了大语言模型输出中简洁性与完整性之间的普遍张力。我们的知识图谱设计优先考虑实际需求——在应用型合同场景中侧重清晰简洁以支持快速理解。虽然这一侧重可能略微降低信息完整度，但提升了实际应用中的整体可用性与回答效率。

---

## 图6. 主观任务实验评估结果

知识图谱显著增强模型可解释性。尽管提示工程能影响输出清晰度，但其缺乏嵌入潜在领域逻辑或支持条款溯源能力。相比之下，我们的知识图谱增强系统通过将合同知识构建为显式逻辑路径，显著提升了可解释性。例如，经知识图谱增强的GPT-3.5在可解释性上取得100%胜率，所有评估模型均观察到类似提升。这一改进源于知识图谱能够提供清晰系统的逻辑关系与多热知识推理路径。知识图谱以结构化方式组织并呈现国际工程合同知识，使大语言模型在回答问题时能遵循明确推理路径、定位相关条款并理解合同要素间关系。这种结构化信息不仅帮助模型更精准定位相关知识节点，还清晰展示各知识节点间的关系与逻辑联系。此外，知识图谱支持在答案生成时直接引用具体条款及相关文档，提供可溯源性，从而增强透明度与用户信任。

## 结论

本研究通过融合知识图谱与大语言模型两项新兴技术，开发了面向国际工程合同的智能问答系统，旨在促进合同知识管理并支持国际工程人员学习合同相关知识。提出了国际工程合同本体框架，并基于NEC4与2022版FIDIC黄皮书合同提取的3000余组三元组构建了本体知识图谱。该知识图谱融合了复杂的条件、时间与因果关系，能完整捕捉合同文本的语义语境。此结构结合合同本体视角，确保了合同知识的全面表征。由知识图谱与大语言模型驱动的问答系统支持自然语言交互，增强对查询的理解并生成相关且贴合具体情境的答案。

本研究相较于现有问答系统展现出三大关键优势。首先，通过采用大语言模型作为问答系统的阅读器，该系统依托庞大的参数化知识库，充分利用其先进的语义理解与文本生成能力。这使得系统能够精准理解复杂查询、进行推理，并生成流畅且语境贴切的回答。其次，集成从知识图谱衍生的推理子图，为系统提供了密集领域专业知识与复杂逻辑推理链的坚实基础，显著增强了其准确全面回答复杂问题并减少幻觉的能力。这些显式推理路径不仅提升了系统的认知水平，还为其底层决策过程提供了关键透明度，缓解了黑盒模型通常伴随的不透明性问题。因此，该特性确保了系统在实际合同管理场景中的适用性，增强了输出结果的可解释性与可靠性。最后，系统的灵活性是另一显著优势。合同知识可通过知识图谱高效、低成本地更新，并能轻松与各类大语言模型集成。评估结果表明，所提出的问答系统在四个核心维度——准确性、简洁性、全面性与可解

---

释性上均优于基线模型。

基于本研究的洞察，所提出的系统构建了一个综合性框架，用于开发先进智能工具，赋能国际建设工程合同人员通过人工智能技术高效检索与吸收合同知识。该系统支持动态、灵活的合同相关查询与学习流程，同时为合同管理提供任务辅助分析支持，契合项目特定需求并与组织目标保持一致。通过协助合同人员在合同审查期间进行风险评估、需求分析与合规验证等关键活动，系统显著减少了时间与精力投入，从而提升了审查流程的效率、精确度与整体质量。此外，通过优化合同知识的利用，它有助于改进管理实践与项目成果。该方法不仅简化了国际建设工程合同部门内合同相关信息的管理，还确保了合同知识的高质量提取、表征与存储，促进了组织内部知识的复用与共享。最终，它在推动国际建筑企业建立稳健、自适应的知识管理体系方面发挥着关键作用。

尽管本研究在国际工程合同领域取得了进展，但仍存在若干局限。首先，主要局限在于数据源与评估设计的范围受限。具体而言，知识图谱与评估数据集仅基于NEC4合同与FIDIC 2017黄皮书构建，覆盖的合同类型与地区范围狭窄。这导致合同类型覆盖有限，未能纳入涵盖更广泛项目类型与地区的协议，如AIA或JCT合同。数据集主要聚焦于建设工程合同文件的核心要素——特别是通用条件——而排除了工程量清单、变更指令等补充文件，尤其是真实合同文件。未来研究将致力于通过两个关键维度扩展合同数据范围以应对这些局限。横向层面，将考虑纳入多样化的合同类型——如EPC交钥匙、AIA与JCT合同——以及地区标准范本，如中国的建设工程合同示范文本。纵向层面，研究将通过融入实际项目合同条件及相关文件——包括合同数据表、合同协议及其他附属文件——来提升条款粒度。这将增强模型的推理能力，尤其针对非标准条款。此外，当前研究未包含与检索增强生成基线的直接比较。这是有意为之的设计选择，旨在单独评估通过知识图谱整合结构化知识的效果，避免传统文本检索方法的干扰。然而，与RAG流程——特别是使用相同语料库进行非结构化检索的流程——进行全面比较，对于进一步验证本方法的优势至关重要。因此，未来工作将包含与代表性RAG系统的基准测试，以评估其在准确性、简洁性、全面性与可解释性方面的性能。

其次，本研究中的知识图谱通过人工三元组抽取构建，这一过程既耗时又费力。此外，该方法严重缺乏对合同执行期间可能发生的动态条款修改的支持。为解决这些局限，未来研究将探索集成自然语言处理(NLP)技术——如实体识别和关系抽取——以实现自动化知识图谱构建。同时，我们建议开发一个结构化抽取模块，利用大语言模型进行合同修订识别，生成知识图谱更新的结构化建议，例如节点添加和关系变更。展望更远，我们设想超越静态表示，通过使用图神经网络(GNNs)将知识图谱框架演化为上下文敏感结构。这可能使大语言模型能够动态交互和操作知识结构。然而，这一方向仍超出当前工作范围，被提议作为未来研究的主题。

第三，当前研究仅聚焦于英语合同文本来构建知识图谱和设计问答系统，使用了基于英语训练的嵌入模型multi-QA-mpnet-base-dot-v1。尽管该方法在单语语境中验证了方法论的有效性，但面对国际工程项目的多语言现实——如法语、阿拉伯语或其他语言的合同，以及多语言用户查询——系统表现出显著的语言局限性。其处理非英语合同文本或用户查询的有效性无法保证，目前依赖于翻译工具或适配多语言模型。为解决这一局限，未来工作将用多语言嵌入模型替代单语模型，实现跨语言合同文本映射到统一语义空间，以达成跨语言实体对齐。此外，知识图谱中的实体和关系将标注多语言标签，以确保翻译一致性。

最后，尽管本研究开发的问答模型主要关注知识问答系统的架构设计——为专业人士提供合同知识学习服务——但由于缺乏对实际执行合同的问答测试验证，尚未扩展到具体合同管理任务，如合同风险识别或合规审查。未来研究可通过开发风险评估任务的专用模型，增强所提技术的实际

---

适用性。此类系统将利用基于真实合同构建的知识图谱，结合大语言模型，建立任务特定的问答模型，能够通过查询风险审查检查点识别合同风险；依据企业标准评估风险等级；并提供可操作的谈判建议和实施措施。这一方法将显著提升该技术在实际合同管理中的效用。（来源：EngineeringJournals微信公众号）

相关论文信息：<https://journal.hep.com.cn/fem/EN/10.1007/s42524-026-4237-0>

特别声明：本文转载仅仅是出于传播信息的需要，并不意味着代表本网站观点或证实其内容的真实性；如其他媒体、网站或个人从本网站转载使用，须保留本网站注明的“来源”，并自负版权等法律责任；作者如果不希望被转载或者联系转载稿费事宜，请与我们联系。

作者：Yongcheng FU 来源：《工程·管理》

更多科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发