

融合 RAG 与大语言模型的建筑合同问答系统研究

李欢馨 贺佳音 徐佳艺 佟烁

中国矿业大学（北京） 北京 100083

【摘要】：针对建筑合同条款复杂、传统人工审核效率低下、纯大语言模型易产生知识幻觉等问题，本文提出一种融合检索增强生成（RAG）与大语言模型的建筑合同智能问答系统。系统构建包含 301 份多类别建筑合同的知识库，采用层级感知混合切片策略与 BGE-large-zh-v1.5 编码器实现文本向量化，基于 FAISS 建立高效语义检索模块，结合结构化提示工程与 DeepSeek 大模型完成合规化答案生成。实验结果表明，系统检索准确率达 96.64%，幻觉率仅 5.67%，有效提升了建筑合同问答的专业性、准确性与可解释性，为工程合同合规审核提供智能化技术支撑。

【关键词】：检索增强生成；建筑合同；智能问答；大语言模型

DOI:10.12417/2811-0528.26.14.072

1 引言

随着建筑业快速发展，工程规模与复杂度持续提高，建筑合同作为界定各方权责、规范履约的核心法律文件，其合规审核关乎项目风险防控与利益保障^[1]。合同问答作为快速获取条款信息、辅助合规判断的核心手段，其性能优劣直接决定合规审核效率与风险识别质量。

建筑合同条款繁杂、逻辑复杂，人工审核耗时费力^[2]；传统关键词检索局限于表层匹配，纯大语言模型存在知识滞后与幻觉问题，通用问答系统也缺乏建筑领域专业适配，难以满足合规审核需求。现有问答系统虽已实现语义理解与推理功能，但仍无法深挖条款深层语义、保障法律准确性，且未能融合行业规范^[3]。

针对上述问题，本研究以 RAG 技术为核心，并协同利用 LLM 的深度理解生成能力与外部知识库的精准检索能力，构建了一个合同智能问答系统，将大量合同文本中的结构化与非结构化知识，转化为一个可交互、可查询的动态知识体系，直接赋能于建筑行业业务决策、风险核查和履约管理等多个环节，解决了建筑合同复杂性高、传统 LLM 易产生知识幻觉等问题，为建筑行业合同合规审核的智能化升级提供技术支撑与实践参考。

2 相关研究

2.1 合同问答系统研究现状

早期合同问答研究多采用基于规则的关键词匹配，难以应对复杂条款关系。BERT 的诞生开启了法律 NLP “预训练+微调”新时代^[4]；Chalkidis 等提出的 Legal-BERT 在法律文本分

类任务上性能普遍优于通用 BERT^[5]。近年来，LLM 在合同信息抽取中展现出零样本泛化潜力^[6]，两阶段提示链方法在细粒度合同问答任务上取得更准确的结构化输出^[7]。总体而言，合同问答系统正沿着“数据—模型—范式”路径快速演进，但在大规模中文合同数据稀缺、超长文本幻觉控制等方面仍面临挑战。

2.2 RAG 技术

Amazou 等验证了 RAG 在合同法助手中的卓越表现，通过语义索引技术将外部法律文本集成至生成式模型中，有效规避模型幻觉^[8]。李圣飞等通过对工程管理规范进行精细化切片与向量化处理，实现了针对现场管理标准的专业问答^[9]。现有研究在跨章节风险因子精准关联方面仍有不足，亟须开发具备长文本处理能力与行业知识深度融合的 RAG 增强型审查架构。

3 系统设计

本系统采用基于检索增强生成（RAG）的技术框架，构建面向合同文本的智能问答系统，该技术通过协同利用 LLM 的深度理解生成能力与外部知识库的精准检索能力，实现对合同内容的“即问即答”。系统的核心流程分为精准检索和核心生成两个阶段：首先将用户的自然语言问题转化为查询向量，并在由合同文本经过切片、向量化后构建的专用知识库中进行语义相似度检索，锁定最相关的合同文本片段作为“证据”；随后将检索到的精确证据与原始问题一同作为提示，提交给 LLM 进行推理与整合，在预设的法理约束下，生成既具备法律准确性又符合行业语境的解答，具体框架如图 1 所示。

作者简介：李欢馨，女，山西省晋中市平遥县人，本科在读，主要研究方向：大数据管理与应用。贺佳音，女，湖南省湘乡市人，本科在读，主要研究方向：大数据管理与应用。徐佳艺，女，江苏省苏州市人，本科在读，主要研究方向：大数据管理与应用。佟烁，男，北京市人，本科在读，主要研究方向：大数据管理与应用。

基金项目：中国矿业大学（北京）大学生创新训练项目资助（校级项目编号：202505013）；中央高校基本科研业务费专项资金资助。

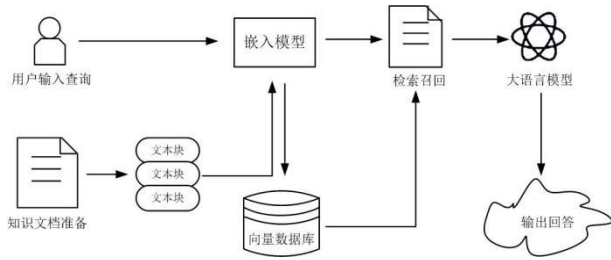


图1 建筑合同问答系统总体框架

3.1 合同语料语义嵌入与向量库构建

系统的实现首先需要将非结构化的合同原文转化为可计算、可匹配的向量化知识库。本研究从政府部门网站以及招标信息网站等搜集真实项目合同，涵盖工程勘察合同、建筑工程设计合同、施工总承包合同等13种类别，共计301份，并针对建筑合同篇幅长、结构层次复杂的特点，采用层级感知的混合切片策略进行文本分割。具体而言，以合同原文的自然层级边界为主要切分锚点，将每个完整的“条”作为独立切片单元，对于内容超过512 token的超长条款则采用滑动窗口重叠切分，确保跨窗口的语义连续性。随后选用BGE-large-zh-v1.5作为编码器，将每个切片转化为1024维的高维特征向量，同时保留合同中的逻辑关系和专业术语。编码完成后，采用FAISS（Facebook AI相似性搜索库）构建向量索引，建立高效的近似最近邻检索机制，为后续实时语义检索奠定基础。

3.2 精准检索

用户输入问题后，系统使用BGE-large-zh-v1.5将问题转化为查询向量，在知识库中进行余弦相似度计算，按相关性得分筛选Top-5高价值文本片段，剔除重复信息后形成生成解答的核心事实依据，例如，当用户询问“承包人在接收竣工付款证书后，还能否就工程接收前发生的事件提出索赔？”时，系统通过检索技术定位到知识库中关于索赔期限和工程接收的关联条款，为后续生成解答提供事实依据，从源头规避大语言模型的“知识幻觉”问题。为提升检索的准确性，针对建筑合同跨章节关联强的特点，系统采用跨章节检索机制，能够处理涉及多个条款交叉引用或需要总结归纳的问题，确保依据的完整性。

3.3 智能生成

系统遵循“法理约束+结构化输出”原则，构建结构化提示工程模板，包含三个核心部分：①回答原则：忠实原文、要素完备，严禁幻觉；②输出格式：简要结论、分点分析并引用条款编号、归纳总结；③参考范例。调用DeepSeek大模型，采用“结论—依据—总结”输出范式生成合规化解答，确保解答既合规又贴合业务实际。

4 系统实现与测试

4.1 数据收集与预处理

从政府部门网站及招标信息网站收集301份合同，涵盖工程勘察合同、施工总承包合同、材料买卖合同等13类；同时爬取现行有效的建筑行业法律法规，并邀请法律专家编写专家问答，涵盖勘察文件审查责任、索赔期限界定及逾期赔偿标准等关键履约环节。对原始文本进行清洗，过滤标题性内容和无效文本，将合同文本按段落拆分为独立条款；使用BGE-large-zh-v1.5进行向量化处理，转化为1024维特征向量，存储于FAISS向量数据库。表1为合同收集数据统计。

表1 合同收集数据

一级分类	二级分类	样本个数
A. 工程前期与咨询服务类	A1.工程勘察合同	27
	A2.建设工程设计合同	25
	A3.建设工程咨询合同	25
	A4.建设工程监理合同	25
B.工程承包与施工类	B1.施工总承包合同	34
	B2.专业分包合同	11
	B3.劳务分包合同	26
	B4.建筑装饰装修合同	25
C.材料设备采购与租赁类	C1.建筑材料/设备买卖合同	32
	C2.大型机械设备租赁合同	12
	C3.周转材料租赁合同	9
D.特殊类型	D1.BT/BOT/PPP项目合同	33
	D2.联合体协议	17

4.2 大语言模型测试

构建80个典型查询问题的测试集，涵盖合同效力与主体资格、价款与支付、工期与履行期限、质量与责任、文件解释与冲突解决、风险控制与建议六大类别，对比DeepSeek、Qwen-3.5和GLM-5三个大语言模型。实验结果（表2）显示，DeepSeek在检索准确率（96.64%）、答案准确率（95.14%）、风险识别率（93.58%）上均表现优异，幻觉率最低（5.67%）。DeepSeek在风险识别率上较Qwen-3.5高出32.75个百分点，较GLM-5高出40个百分点，对合同条款中隐含的法律合规风险具有更强的敏感性，最终选定为系统生成模型底座。

表2 大语言模型测试结果

	检索准确率	答案准确率	风险识别率	幻觉率
DeepSeek	96.64%	95.14%	93.58%	5.67%
Qwen-3.5	83.35%	86.10%	60.83%	10.83%
GLM-5	86.50%	86.00%	53.58%	9.67%

4.3 系统展示

系统采用简洁明了的界面设置,用户以自然语言输入业务疑问,系统采用结构化输出范式:首先给出明确结论,随后分点论述并引用合同原文具体条款,最后归纳总结。以典型问答为例:当用户询问“承包人在接收竣工付款证书后,还能否就工程接收前发生的事件提出索赔?”时,系统通过解析合同第19条“提出索赔的期限”与第10条“验收和工程接收”等相关条款,得出否定性结论,有效规避了因错过索赔时效而造成的重大经济损失,体现了系统对复杂业务场景的深度理解与精准推理能力。图2展示了部分对话内容。



图2 合同问答系统界面展示

参考文献:

[1] 李陈.面向建筑工程合同风险智能问答系统研究[D].苏州科技大学,2023.

[2] 陈锦锋.《数字建筑发展白皮书》发布赋能城市智能生态[N].通信信息报,2022-04-06(007).

[3] 闫悦,郭晓然,王铁君,等.问答系统研究综述[J].计算机系统应用,2023,32(08):1-18.

[4] Devlin J,Chang M W,Lee K,et al.BERT:Pre-training of Deep Bidirectional Transformers for Language Understanding[C]//Proceedings of NAACL-HLT 2019.ACL,2019.

[5] Chalkidis I,Fergadiotis M,Malakasiotis P,et al.LEGAL-BERT:The Muppets Straight Out of Law School[C]//Findings of EMNLP 2020.ACL,2020.

[6] Haque A,Singh M P.Extracting Norms from Contracts Via ChatGPT:Opportunities and Challenges[C]//COINE 2024.Springer,2025.

[7] Roegiest A,Chitta R.Answering Questions in Stages:Prompt Chaining for Contract QA[EB/OL].arXiv:2410.12840,2024.

[8] Amazou Y,Tayalati F,Mensouri H,et al.Accurate AI Assistance in Contract Law Using Retrieval-Augmented Generation[J].IJACSA,2025,16(2).

[9] 李圣飞,卢昱杰,陈晓莹,等.基于检索增强生成的建筑工程管理知识问答模型实现[J].科学技术与工程,2025,25(25):10840-10849.

从系统的运行测试结果来看,该智能问答模块展现了高度的交互性与准确性。通过这种直观的问答交互,不仅显著压缩了人工查阅与理解合同的时间成本,更通过对底层事实的精准溯源,有效规避了传统生成式模型容易产生的幻觉风险,为企业合同履行过程中的风险动态监测与管理决策提供了可靠、高效的智能化工具。

5 总结与展望

本研究围绕建筑合同的高效检索与精准问答需求,构建了基于检索增强生成的智能问答系统,核心贡献体现在三个方面:其一,构建了涵盖合同语料库、法律法规文件库、专家问答数据集的三层数据体系,解决了通用模型在建筑合同领域语义适配不足的痛点;其二,设计“BGE 语义检索+LLM 生成+专家问答对优化”协同方案,通过向量检索实现条款级精准定位,结合法律法规文件库作为合规约束抑制模型“幻觉”,并引入“结论—依据—总结”三段式输出范式增强回答可信度;其三,研发的智能问答系统体现了高度的交互性和准确性,为建筑企业合同管理的智能化转型提供工具支持。然而,本研究仍具有局限性:一是数据规模较小(301份),其结论的泛化能力尚需在大规模语料上进行验证;二是知识动态更新机制仍需优化,知识库难以及时反映政策变动。针对以上局限,本研究将从扩充数据集、增加对比模型、构建知识动态更新机制等入手,进一步提升系统的学术支撑能力与用户体验,助力建筑行业合规治理数字化转型。