Computer Science and Application 计算机科学与应用, 2026, 16(9), 169-181
Published Online September 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.169298

Hans汉斯

基于多文档异构结构图的检索增强生成方法

师培仁[1]，柴变芳[2]

[1]河北地质大学信息工程学院，河北 石家庄
[2]河北省地质环境感知与数据处理重点实验室，河北 石家庄

收稿日期：2026年8月7日；录用日期：2026年9月8日；发布日期：2026年9月17日

摘　要

针对财政制度多文档问答任务中证据分散、结构异构和生成依据不足等问题，提出一种基于多文档异构结构图的检索增强生成方法(Multi-Document Heterogeneous Structural Graph-Based Retrieval-Augmented Generation Method, MDHSG-RAG)。该方法通过结构模板构建统一层级图索引，按问题类型进行由粗到细地分层检索，并以证据链约束答案生成。实验结果表明，MDHSG-RAG在检索和生成阶段均优于对比方法，提高了财政制度问答的证据定位能力、答案准确性和可追溯性。

关键词

GraphRAG，多文档问答，异构文档，层级图索引

Retrieval-Augmented Generation Based on Multi-Document Heterogeneous Structure Graphs

Peiren Shi[1], Bianfang Chai[2]

[1]School of Information Engineering, Hebei Geo University, Shijiazhuang Hebei
[2]Hebei Key Laboratory of Geological Environment Sensing and Data Processing, Shijiazhuang Hebei

Received: August 7, 2026; accepted: September 8, 2026; published: September 17, 2026

Abstract

To address the problems of scattered evidence, heterogeneous structures, and insufficient generation grounding in multi-document question answering for fiscal regulations, this paper proposed a Multi-Document Heterogeneous Structural Graph-based Retrieval-Augmented Generation Method, called MDHSG-RAG. The method used structural templates to construct a unified hierarchical graph index. It

文章引用: 师培仁, 柴变芳. 基于多文档异构结构图的检索增强生成方法[J]. 计算机科学与应用, 2026, 16(9): 169-181.
DOI: 10.12677/csa.2026.169298

performed coarse-to-fine hierarchical retrieval according to question types and used evidence chains to constrain answer generation. Experimental results showed that MDHSG-RAG outperformed the comparison methods in both retrieval and generation stages. The method improves evidence localization, answer accuracy, and traceability in fiscal regulation question answering.

Keywords

GraphRAG, Multi-Document Question Answering, Heterogeneous Documents, Hierarchical Graph Indexing

1. 引言

大语言模型(Large Language Models, LLMs) [1]在智能问答、文本生成和语义理解等任务中取得了显著进展，但在财政、金融、法律等专业制度文档问答场景中，仅依靠参数化知识难以给出可靠答案。财政制度文档通常包含通知正文、制度条文、附件材料、清单表格、税则表和弱结构说明等多种内容形态，问题答案往往分布在多个文件、多个条款或表格字段中。例如，财政会计类问题“合同取得成本与合同履约成本分别是资产类会计科目还是成本类会计科目？”需要同时检索“合同取得成本”和“合同履约成本”的相关规定，并结合制度条文判断其会计科目属性，属于典型的多概念、多证据联合判断问题。此类任务不仅要求模型理解语义，还需要准确定位制度依据、保持证据来源可追溯，并避免在证据不足时生成超出依据范围的结论。因此，如何在多文档、异构结构和细粒度证据并存的条件下组织外部知识，成为财政制度问答中需要解决的关键问题。

检索增强生成(Retrieval-Augmented Generation, RAG) [2]通过在生成前检索外部知识，为缓解大模型事实性幻觉和领域知识不足提供了有效思路。传统RAG通常将文档切分为固定长度文本块，并采用BM25 (Best Matching 25) [3]或向量相似度检索召回相关片段，再将检索结果输入LLM生成答案。传统方法实现简单、效率较高，但知识表示阶段[4]文本块之间缺少显式结构关系，如财政文档中的文件层级、附件依附和表格字段结构，导致不能基于文档间和文档内的逻辑结构检索。为增强RAG知识表示的结构化表示能力和检索的结构化推理能力，研究者提出了一些图的检索增强生成(Graph-Based Retrieval-Augmented Generation, GraphRAG) [5]方法，将实体、关系、文本片段或文档单元组织为图结构。Microsoft GraphRAG [6]通过构建文档蕴含的知识图谱和实体的社区摘要组织知识，实现局部检索和全局检索；LightRAG [7]将利用大模型构建更准确的知识图谱，存储知识图结构、实体关系描述和文档块向量，支持文档快、实体、关系及混合检索；RAPTOR [8]通过递归聚类和摘要构建树状索引，检索阶段采用自顶向下查询与检索向量相关的节点，最后聚合检索的相关节点信息生成答案。这些方法对文本块的实体关系结构进行不同粒度的利用，但聚焦表示文档集合的语义结构，对文档中的内部结构表达不足，导致检索结果为孤立文本片段，限制大模型生成答案。

利用多文档结构实现问答的研究进一步关注文档内部精细表示和跨文档证据检索。KGP (Knowledge Graph Prompting) [9]将文档内部段落、页面和表格建模为图节点，并通过语义相似和结构从属关系连接节点，再利用基于LLM的图遍历代理在多个文档及其页面、段落和表格等结构之间交替检索与推理，逐步收集证据。KG-Retriever [10]通过知识图谱层和协作文档层构建层次化索引，以缓解跨文档检索中的信息

碎片化问题；MMRAG-DocQA [11] [12]则面向长文档问答构建层次化索引和多粒度检索机制，以连接跨页、跨模态证据。RoG [13] [14]基于多文档构建知识图谱，设计基于知识图谱中的关系路径推理策略，提升生成结果的可解释性。GRAG [15]从文本图中检索相关子图并保留拓扑信息用于大模型生成。G-Retriever [16]对构建的问题相关文档语义图进行嵌入表示，基于查询向量检索该图，并对检索种子节点进行扩展得到检索子图，最后对其进行序列化用于大模型生成答案。GNN-Ret [17]构建文档段落图，利用图神经网络学习感知上下文的段落表示，通过稠密检索段落子图，并利用 GNN 得到子图表示，进而将排序靠前的子图段落提供给生成模型。同时，HybridRAG [18] [19]将向量检索与知识图谱检索结合，用于复杂金融文档中的信息抽取和问答；UniOQA 将大语言模型与知识图谱问答流程结合，提升开放知识图谱问答能力；MBA-RAG [20]根据问题复杂度动态选择检索策略，平衡检索准确性与效率。然而，现有多文档检索方法大多针对问题构建相关文档图，不适合问题未知的问答任务。在文档图索引方面文档图结构单一，不适应富含多种结构的文档结构化表示。在检索策略上，不能利用逐步求精的检索过程形成问题到答案的证据链。

针对上述问题，提出一种基于多文档异构结构图的检索增强生成方法(Multi-Document Heterogeneous Structural Graph-based Retrieval-Augmented Generation Method, MDHSG-RAG)。该方法面向财政制度文档实现多类型文档结构表示和多粒度内容检索：索引阶段，利用结构模板动态识别文档结构，并构建统一层级图索引，使通知正文、制度条文、附件材料、表格清单和弱结构说明等内容能够映射到文档结构表示中；在检索阶段，依据问题类型选择相应结构化检索模板，然后自顶向下层级检索文档级、条款级、内容块级和表格字段级逐层缩小候选空间，并根据查询意图结合引用关系、附件依附关系和表格关联关系进行检索扩展；在生成阶段，将检索到的多粒度证据组织为带有来源路径的结构化上下文，通过证据链约束答案生成。该方法可提升异质文档问答多粒度证据定位能力、答案准确性和可追溯性。

2. MDHSG-RAG 方法框架

2.1. 整体框架

MDHSG-RAG 以财政文档为示例，构建多文档异质结构感知的层级文档结构图，实现自适应多粒度文档图检索，为大模型提供结构化文档证据。整体框架如图 1 所示。

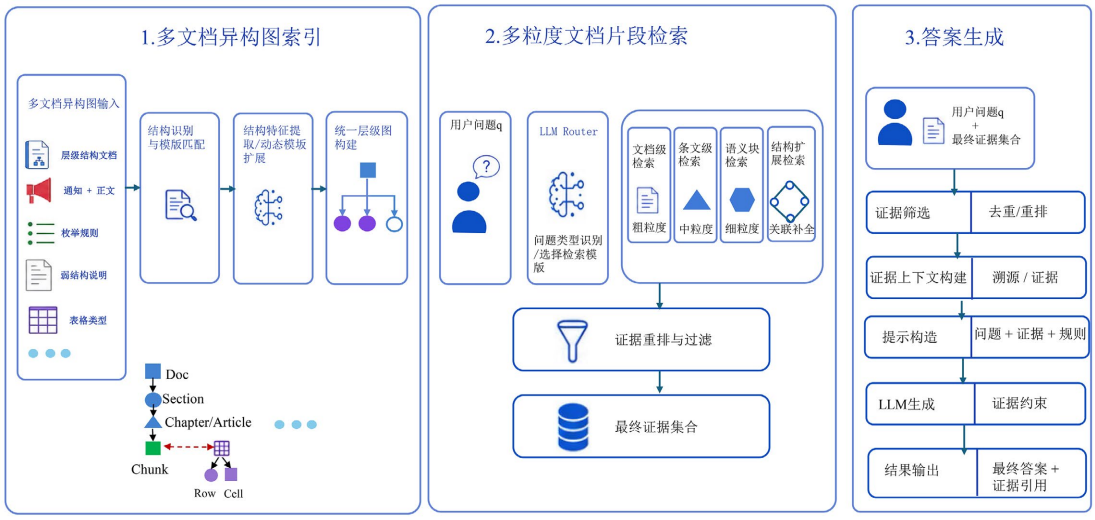

Figure 1. A retrieval-augmented generation framework based on multi-document heterogeneous structural graphs
图 1. 基于多文档异构结构图的检索增强生成方法框架

给定财政制度文档集合 $D=\{d_1,d_2,\cdots,d_n\}$ 和用户问题 q，MDHSG-RAG 其由图索引构建、分层图检索和证据约束生成三个阶段组成。图构建阶段对多结构文档进行结构识别、模板匹配和统一图构建，将异构文档映射为统一层级图结构 G；检索阶段依据问题类型选择检索模板，在文档级、条款级、内容块级和表格字段级逐层收缩候选空间，进一步沿图关系进行受控扩展，形成最终证据集合 S_q 成阶段则将证据集合组织为结构化上下文，并通过显式回答规则约束大语言模型生成最终答案。整体过程可表示为。

$$G=\text{Build}(D,\mathcal{T}) \tag{1}$$

$$S_q=\text{Retrieve}(q,G,\mathcal{T}) \tag{2}$$

$$y=\text{Generate}(q,S_q,\mathcal{R}) \tag{3}$$

其中，$\mathcal{T}$ 表示结构模板库 $G=\text{Build}(.)$ 表示多结构图索引构建过程，$\text{Retrieve}(.)$ 表示问题感知的分层图检索过程，$\text{Generate}(.)$ 表示证据约束下的答案生成过程，$\mathcal{R}$ 表示答案生成阶段的约束规则集合。

MDHSG-RAG 整体流程由图索引构建、分层图检索和证据约束生成三个阶段组成：图构建阶段负责将异构文档映射为统一层级图结构；检索阶段依据问题类型选择检索模板，并由粗到细定位证据；生成阶段则基于结构化证据上下文生成可追溯答案。

2.2. 多图结构索引构建

图索引构建过程如图 2 所示。其先识别文档结构类型，再根据已有模板或新增模板将不同类型文档表示为图结构。制度文件可能是章条型制度文件、通知正文、枚举规则、弱结构说明、附件材料和表格清单等结构。若采用统一切块方式，容易破坏条款边界、附件依附关系和表格字段结构。

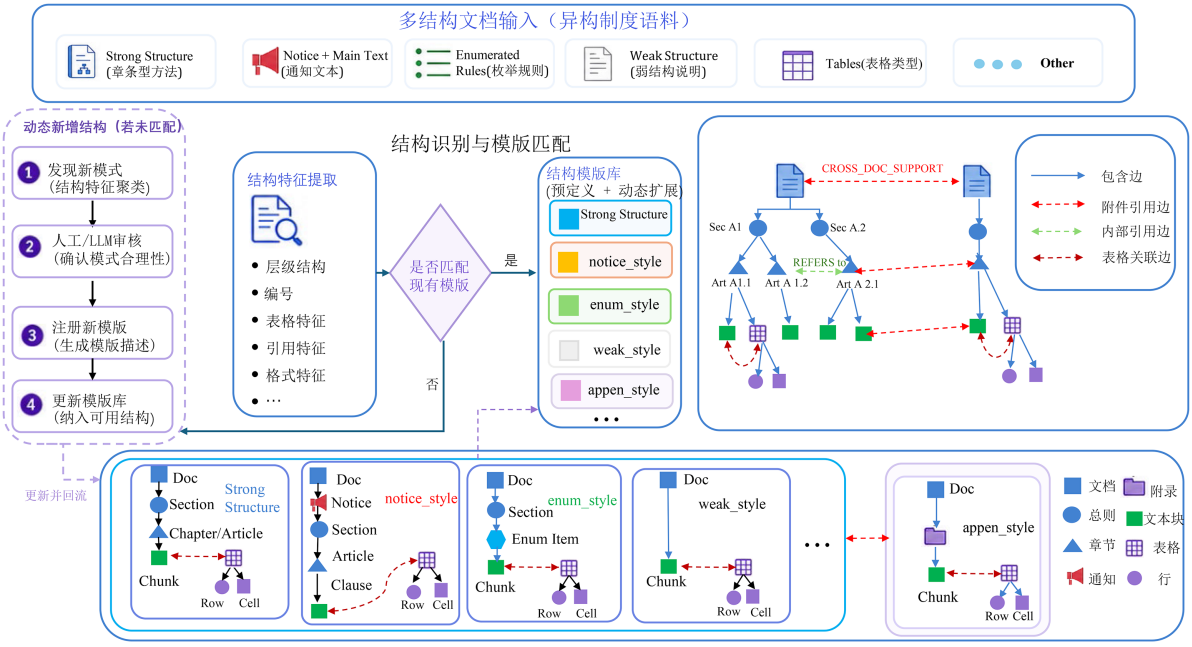

Figure 2. Construction diagram of multi-document heterogeneous graph
图 2. 多文档异质结构图构建

对于任意输入文档 d_i，首先采用规则解析与版式统计相结合的方式提取其结构特征。结构特征主要包括标题层级、编号模式、条文标识、附件标记、表格分布、引用表达、枚举格式和正文组织方式等。其

中，标题层级和编号模式通过正则规则识别，附件标记通过关键词和位置模式识别，表格特征通过表格检测和表头字段抽取获得，引用表达则通过“依据”“参照”“见附件”等句式模式进行识别首先提取其结构特征：

$$F_i = \text{ExtractFeature}(d_i) \tag{4}$$

进一步地，结构特征可表示为：

$$F_i = \{h_i, n_i, a_i, t_i, r_i, e_i, p_i\} \tag{5}$$

其中，h_i 表示标题层级特征，n_i 表示编号模式特征，a_i 表示附件标记特征，t_i 表示表格结构特征，r_i 表示引用表达特征，e_i 表示枚举格式特征，p_i 表示正文组织方式特征。

结构模板库 $\mathcal{T}$ 用于存储不同类型财政制度文档的结构解析模式。每个结构模板 $\tau \in \mathcal{T}$ 由模板类型、节点类型、边类型和解析规则组成，可表示为：

$$\tau = \{\text{type}, V_\tau, E_\tau, P_\tau\} \tag{6}$$

其中，type 表示模板类型，如章条型制度文件、通知正文型文件、附件清单型文件、表格规则型文件和弱结构说明型文件等；V_τ 表示该模板中的节点类型集合，E_τ 表示节点之间的结构边类型集合，P_τ 表示标题识别、条文切分、附件识别和表格解析等结构解析规则。随后，将 F_i 与结构模板库 $\mathcal{T}$ 中的模板进行匹配：

$$\tau_i = \arg\max_{\tau \in \mathcal{T}} \text{Sim}(F_i, \tau) \tag{7}$$

其中，$\text{Sim}(F_i, \tau)$ 表示文档结构特征与模板规则之间的匹配程度，由标题层级、编号模式、附件标记、表格结构、引用表达和正文组织方式等多类特征的加权匹配得分共同决定：

$$\text{Sim}(F_i, \tau) = \alpha_1 S_h + \alpha_2 S_n + \alpha_3 S_a + \alpha_4 S_t + \alpha_5 S_r + \alpha_6 S_p \tag{8}$$

若最大相似度低于阈值 θ，则说明现有模板无法稳定解析该文档，需要根据当前文档的结构特征生成新模板并更新模板库：

$$\mathcal{T} \leftarrow \mathcal{T} \cup \{\tau_{\text{new}}\}, \quad \text{if } \max_{\tau \in \mathcal{T}} \text{Sim}(F_i, \tau) < \theta \tag{9}$$

其中，F_i 包括标题层级、编号模式、表格特征、引用表达、附件标记、枚举格式和正文组织方式等信息。随后，将 F_i 与结构模板库 $\mathcal{T}$，进行匹配：

$$\tau_i = \arg\max_{\tau \in \mathcal{T}} \text{Sim}(F_i, \tau) \tag{10}$$

若最大相似度低于阈值 θ，则说明该文档结构无法被现有模板稳定解析，需要进入动态模板扩展流程：

$$\mathcal{T} \leftarrow \mathcal{T} \cup \{\tau_{\text{new}}\}, \quad \text{if } \max_{\tau \in \mathcal{T}} \text{Sim}(F_i, \tau) < \theta \tag{11}$$

在模板确定后，文档被解析为多粒度节点集合。根据不同文档结构，节点可包括文档节点、章节节点、条款节点、内容块节点、附件节点、表格节点、行节点和字段节点等。

边集合用于表达文档内部层级关系、正文与附件关系、条款引用关系、表格关联关系和跨文档支撑关系。单篇文档对应的结构图可记为：

$$G_i = (V_i, E_i) \tag{12}$$

对所有文档图进行合并，可得到全局层级图索引：

$$G = \bigcup_{i=1}^{n} G_i = (V, E) \tag{13}$$

图索引用于保持结构连接关系，MDHSG-RAG 同时建立稀疏向量索引和稠密向量索引。稀疏向量索引用于支持关键词匹配，稠密向量索引用于语义补召回。

2.3. 多粒度文档片段检索

检索过程如图 3 所示。其先通过问题路由识别问题类型，再选择相应检索模板，最后按照文档级、条款级、内容块级和表格字段级逐层缩小候选范围。

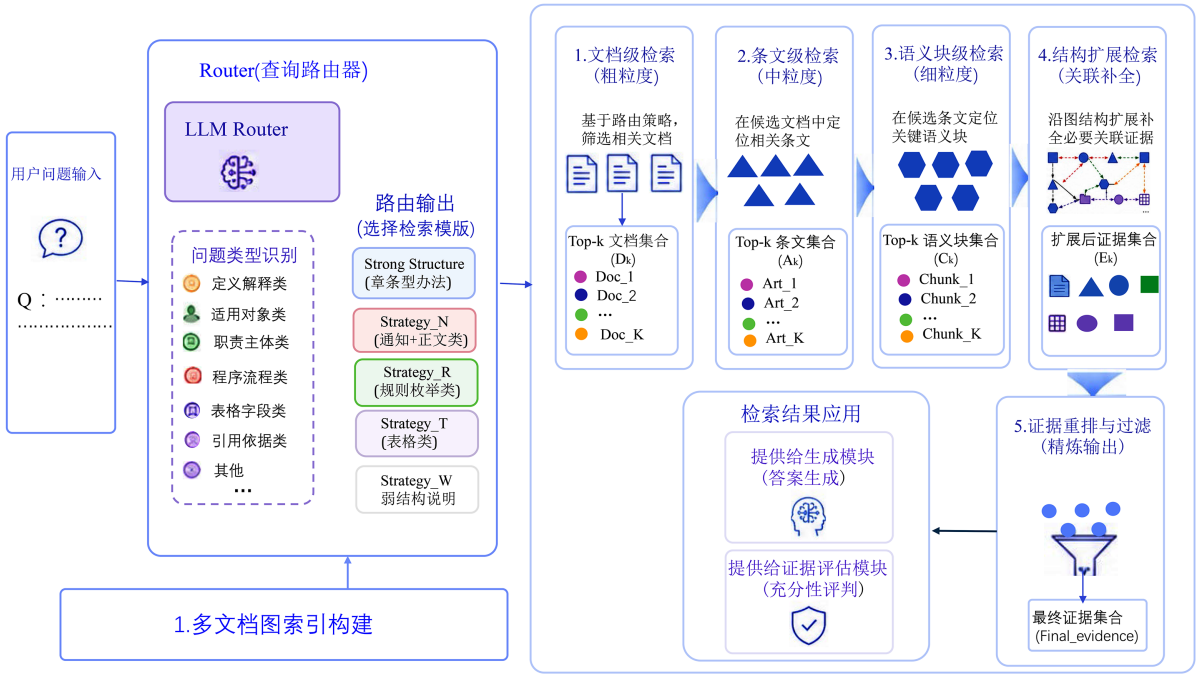

Figure 3. Multi-granularity document fragment retrieval diagram
图 3. 多粒度文档片段检索图

首先，给定用户问题 q 题路由器识别其问题类型：

$$r_q = \text{Router}(q) \tag{14}$$

其中，q 表示用户输入问题，$\text{Router}(.)$ 表示问题类型识别函数，r_q 表示问题 q 的路由结果，可对应定义解释类、适用对象类、职责主体类、程序流程类、表格字段类、引用依据类和弱结构说明类等。不同问题类型对应不同的检索模板和图扩展规则：

$$s_q = \text{SelectStrategy}(r_q, \mathcal{T}) \tag{15}$$

其中，$\mathcal{T}$ 表示结构模板库，存储不同文档结构和问题类型对应的检索模板；$\text{SelectStrategy}(.)$ 检索策略选择函数；s_q 表示问题 q 对应的结构化检索策略。该策略主要包括检索层级顺序、优先检索节点类型、可跟随的图边类型、图扩展深度和候选证据数量限制等。例如，表格字段类问题优先检索表格节点、行节点和字段节点；引用依据类问题优先检索条款节点并沿引用边扩展；附件类问题优先沿附件依附边扩展。

在文档级检索阶段，系统依据问题语义、标题匹配和结构模板约束，筛选候选文档集合 D_q：

$$D_q = \text{TopK}_{\text{doc}}(q, G, s_q) \tag{16}$$

其中，G 表示全局层级图索引，$\text{TopK}_{\text{doc}}(.)$ 表示文档级 Top-K 检索函数，D_q 表示与问题 q 相关的候选文

档集合。文档级检索的作用是先缩小全局搜索空间，避免直接在所有内容块或表格字段中进行大范围匹配。随后，在候选文档内部执行条款级检索：

$$A_q = \mathrm{TopK}_{\mathrm{art}}\left(q, D_q, s_q\right) \tag{17}$$

其中，其中，$\mathrm{TopK}_{\mathrm{art}}(.)$ 表示条款或规则单元级 Top-K 检索函数 A_q 表示候选条款或规则单元集合。对于无明显条文编号的弱结构文档，该层可退化为规则段落或说明性片段检索。之后，系统继续定位内容块或表格字段：

$$C_q = \mathrm{TopK}_{\mathrm{chunk}}\left(q, A_q, s_q\right) \tag{18}$$

$$T_q = \mathrm{TopK}_{\mathrm{table}}\left(q, A_q, s_q\right) \tag{19}$$

其中，$\mathrm{TopK}_{\mathrm{chunk}}(.)$ 表示内容块级 Top-K 检索函数，C_q 表示正文内容块证据集合；$\mathrm{TopK}_{\mathrm{table}}(.)$ 表示表格字段级 Top-K 检索函数 C_q 正文内容块证据，T_q 表示表格、行或字段级证据。对于表格字段类问题，检索过程会优先保留表格节点、行节点和字段节点；对于引用依据类问题，则优先保留条款节点及其引用边。得到初始证据集合后，进一步沿图关系执行受控扩展：

$$S_q^{t+1} = S_q^t \cup \mathrm{Expand}\left(S_q^t, E, s_q\right) \tag{20}$$

其中，S_q^t 表示第 t 轮扩展后的候选证据集合，表示第 $t+1$ 轮扩展后的候选证据集合；E 表示全局层级图索引中的边集合；$\mathrm{Expand}(.)$ 表示图关系扩展函数，用于根据检索策略 S_q 定边类型补充相关证据。扩展过程并非无约束多跳搜索，而是根据问题类型选择可跟随的边集合。设图索引中的边集合可表示为：

$$E = E^{\mathrm{refer}} \cup E^{\mathrm{attach}} \cup E^{\mathrm{table}} \cup E^{\mathrm{cross}} \tag{21}$$

其中，E^{refer} 表示引用边集合，用于连接存在制度引用、条款引用或文件依据关系的节点；E^{attach} 表示附件依附边集合，用于连接正文与附件、附表或清单节点；E^{table} 表示表格关联边集合，用于连接内容块、表格、行和字段节点；E^{cross} 表示跨文档支撑边集合，用于连接不同文档之间存在补充说明、上位依据或规则支撑关系的节点。

扩展过程并非无约束多跳搜索，而是根据问题类型可跟随扩展的边集合。例如，引用依据类问题主要沿 E^{refer}，附件类问题主要沿 E^{attach} 扩展，表格类问题主要沿 E^{table}，跨文件依据类问题则沿 E^{cross}。为防止引入过多噪声，扩展深度和候选数量均受到限制。

候选证据扩展后，进行去重、排序和过滤。设任意候选证据为 e_j，其综合得分定义为：

$$\mathrm{Score}\left(e_j, q\right) = \lambda_1 \mathrm{Sim}\left(e_j, q\right) + \lambda_2 \mathrm{Match}\left(e_j, q\right) + \lambda_3 \mathrm{Level}\left(e_j\right) + \lambda_4 \mathrm{Path}\left(e_j\right) + \lambda_5 \mathrm{Source}\left(e_j\right) \tag{22}$$

其中，e_j 表示第 j 个候选证据，q 表示用户问题，$\mathrm{Score}\left(e_j, q\right)$ 表示候选证据 e_j 与问题 q 的综合相关性得分；$\mathrm{Sim}\left(e_j, q\right)$ 候选证据与问题之间的语义相似度，可由嵌入向量余弦相似度计算；$\mathrm{Match}\left(e_j, q\right)$ 表示关键词、实体、条文编号或表格字段名的匹配得分；$\mathrm{Level}\left(e_j\right)$ 表示证据所在结构层级与问题需求之间的匹配程度，例如表格字段类问题更偏向表格或字段级证据，引用依据类问题更偏向条款级证据；$\mathrm{Path}\left(e_j\right)$ 表示证据链路径完整度，用于衡量证据是否保留了从文档、章节、条款到内容块或表格字段的完整来源路径；$\mathrm{Source}\left(e_j\right)$ 表示证据来源可靠性或结构位置权重，例如正式制度文件、上位依据文件和核心条款可赋予更高权重；$\lambda_1 \sim \lambda_5$ 为各评分项的权重系数，用于控制语义相似度、关键词匹配、层级匹配、路径完整度和来源可靠性在综合排序中的影响。最终证据集合可表示为：

$$S_q = \mathrm{TopK}\left(\mathrm{Rank}\left(\mathrm{Filter}\left(S_q^{\mathrm{T}}\right)\right)\right) \tag{23}$$

每条证据均保留来源路径：

$$\text{path}\left(e_j\right)=\text{Doc}\rightarrow\text{Section}\rightarrow\text{Article}\rightarrow\text{Chunk/Table}\rightarrow\text{Row/Cell} \tag{24}$$

经过该阶段处理后，输出的不是孤立文本片段，而是带有文档来源、结构层级、证据类型和关联路径的结构化证据集合，为后续生成阶段提供可追溯上下文。

2.4. 答案生成

答案生成模块将检索阶段返回的结构化证据转化为大语言模型可利用的受控上下文，并限制模型在给定证据范围内生成答案。先对证据进行筛选、编号、结构化组织，再构造显式约束提示，使最终答案同时包含回答内容、证据引用、充分性标记和缺失信息说明。

在此模块中，所有证据项被组织为结构化证据上下文，保留证据编号和来源路径，使模型能够在答案中引用具体证据。随后，将用户问题、结构化证据上下文和回答规则共同构造为提示：

$$P_q=\text{Prompt}\left(q,C_q,\mathcal{R}\right) \tag{25}$$

其中，$\mathcal{R}$ 三类约束：只能基于给定证据回答；必须输出对应证据引用；当证据不足时，应说明缺失信息并避免给出确定性结论。为控制生成边界，引入证据充分性判断函数：

$$\delta_q=\text{Sufficiency}\left(q,C_q\right) \tag{26}$$

其中，$\delta_q\in\{0,1\}$，当 $\delta_q=1$ 表示当前证据足以支撑问题回答；当 $\delta_q=0$ 表示证据存在缺失或无法支撑完整结论。最终答案生成过程可表示为：

$$y=\begin{cases}M\left(P_q\right), & \delta_q=1\\ \text{Cautious}\left(M\left(P_q\right)\right), & \delta_q=0\end{cases} \tag{27}$$

模型输出统一表示为：

$$O_q=\left(\text{answer},\text{citations},\text{is\_sufficient},\text{missing\_info}\right) \tag{28}$$

其中，answer 表示最终答案，citations 表示引用证据编号，is_sufficient 表示证据是否充分，missing_info 表示证据不足时的缺失信息说明。

通过该生成机制，MDHSG-RAG 将答案生成过程限制在结构化证据范围内。生成模型不再依赖自由补全或外部常识扩写，而是在明确证据链约束下完成回答。对于证据充分的问题，系统输出带引用的确定性答案；对于证据不足的问题，则输出审慎说明，从而降低财政制度问答中的事实偏移风险，并提升答案的忠实性、可解释性和可追溯性。

3. 实验与分析

为验证多文档结构感知异构语料检索增强生成方法(MDHSG-RAG)在财政制度文档问答任务中的有效性，实验从检索阶段和答案生成阶段两个方面展开。检索阶段考察不同方法对目标文档、条文、条款及细粒度证据的定位能力；生成阶段评估系统在获得检索证据后生成的答案的正确性、忠实性和证据利用程度。

3.1. 数据集

语料主要来源于财政领域公开政策文件、财政部门咨询答复、地方财政政策问答及相关业务说明材料，共 157 篇文档。其内容覆盖政府采购、政府会计、非税收入、行政事业性资产管理、差旅会议经费、城市基础设施配套费等多个财政业务主题。

根据文档层级将语料划分为文档、条文、条款、内容块、表格行列和字段等不同粒度单元。经处理后，财政文档语料共形成 18966 个候选内容块，用于后续索引构建和检索实验。

原始问答数据主要包括财政部咨询留言答复和河北省财政政策问答两部分，其中财政部咨询留言包含“留言编号、答复单位、回复时间、留言问题、留言回复”等字段，河北省财政政策问答采用“问－答”形式组织。原始收集问答记录共 206 条，处理后得到 206 条问答集。每个问答均经过人工核验，标注内容包括用户问题、标准答案、目标文档、条文或条款位置、关键证据片段以及证据所属粒度。根据问题所需证据类型，评测样本主要分为四类：

a) 结构定位类问题。主要考察系统对目标文件、条文编号和制度位置的定位能力。

b) 政策解释类问题。主要考察系统对政策含义、适用条件和执行要求的理解能力。

c) 依据追溯类问题。主要考察系统对上位文件、引用条款和跨文件规则支撑的检索能力。

d) 表格字段类问题。主要考察系统对表格、清单、收费标准和字段说明等半结构化内容的定位能力。

每个问答样本均标注文档级、条文级、条款级和严格证据级信息。其中，文档级标注用于计算文档召回率(Document Recall@K, Doc R@K)，即前 k 个检索结果中是否包含目标文档；条文和条款级标注用于计算 Article R@K 与 Clause R@K，分别衡量前 k 个结果中是否命中正确条文或条款；严格证据级标注对应具体正文片段或表格字段，用于计算 Strict Evidence R@K，以评价系统是否命中可直接支撑答案的细粒度证据。通过多粒度标注方式，评测集能够同时评价系统的文档定位能力、条文条款定位能力和细粒度证据召回能力。

3.2. 对比方法

选取四类同类主流方法和提出方法对比：

a) Naive RAG：该方法采用固定长度文本块作为检索单元，利用中文文本嵌入模型将文本块和用户问题编码为向量，并根据余弦相似度召回相关片段。

b) 基于 BM25 的检索增强生成方法(BM25-Based Retrieval-Augmented Generation, BM25-RAG)：BM25 是一种经典的稀疏检索方法，并非完整的 RAG 框架。本文将原始文档切分为文本块并构建词项索引，在查询阶段基于词频、逆文档频率和文本块长度归一化计算相关性得分，返回排名靠前的文本块。随后，将检索结果与用户问题拼接后输入大语言模型生成答案。该方法主要依赖关键词匹配，用于评估稀疏检索方法在文档问答中的表现。

c) LightRAG：该方法通过大语言模型抽取实体和关系，构建实体关系图，并采用局部—全局双层检索机制召回相关上下文。实验中将其作为代表性 GraphRAG 基线，用于比较普通实体关系图在财政文档问答中的效果。

d) KGP：该方法基于多文档问答图遍历框架进行适配，使用财政问答数据训练 T5 图遍历代理，并在候选段落图上进行证据选择。该方法用于评估多文档图遍历式检索在财政数据集上的表现。

3.3. 评价指标

检索阶段采用文档召回率@1 (Doc R@1)、文档召回率@5 (Doc R@5)、条文召回率@5 (Article R@5)、条款召回率@5 (Clause R@5)、文档平均倒数排名(Doc MRR)、条款平均倒数排名(Clause MRR)和严格证据召回率@5 (Strict Evidence R@5)作为评价指标。其中，R@1 和 R@5 用于衡量目标文本对象是否出现在前 1 个或前 5 个检索结果中。

采用答案准确率(Answer Accuracy)、忠实度(Faithfulness)、证据覆盖率(Evidence Coverage)和 ROUGE-L 进行综合评估。Answer Accuracy 衡量生成答案是否正确回答问题；Faithfulness 衡量答案是否忠实于给

定证据，是否存在超出证据范围的内容；Evidence Coverage 衡量答案是否覆盖标准答案所需的关键证据；ROUGE-L 用于衡量生成答案与参考答案之间的最长公共子序列相似度。

3.4. 参数设置

Naive RAG 和 BM25-RAG 均采用扁平文本块作为基本检索单元；LightRAG 采用其官方流程从文本块中抽取实体和关系，并构建实体关系图；KGP 使用财政领域问答训练样本进行适配，以获得面向财政多文档问答的图遍历检索能力；MDHSG-RAG 则按照文档结构模板将财政文档解析为文档、条文、条款、内容块、附件、表格和字段等多粒度结构单元，并在此基础上构建统一层级图索引。

在文档切分策略方面，均采用统一的固定长度切分方式。文本块大小设置为 300 字符，重叠长度设置为 50 字符，用于构建向量索引、BM25 词项索引和 LightRAG 的基础文本输入。MDHSG-RAG 不采用单纯固定长度切分，而是依据财政文档的原生结构进行规则化切分。

涉及向量检索的所有方法均采用 text2vec-base-chinese 作为嵌入模型，向量维度统一设置为 768。检索阶段统一设置候选返回规模 Top-k 为 20。为避免生成阶段输入过长，最终用于答案生成的证据数量统一限制为 Top-k = 8，即每种方法均选取得分较高的 8 条证据作为生成上下文。所有方法在答案生成阶段均采用 DeepSeek 作为统一生成模型。为减少生成随机性对实验结果的影响，生成过程采用相同的提示模板框架和一致的解码设置。KGP 基线使用财政领域训练问答样本对 T5 模型进行适配训练，使其具备面向财政多文档问答的候选证据选择能力。

3.5. 检索阶段结果分析

检索阶段主要用于评价不同方法能否准确定位财政制度问答所需的文档来源、条文位置、条款位置和细粒度证据。结果如表 1 所示。

Table 1. Performance comparison in the retrieval stage
表 1. 检索阶段结果分析

Approaches	Doc R@1	Article R@5	Clause R@5	Doc MRR	Clause MRR	Strict @5
Naive RAG	34.04	61.70	23.40	27.66	46.10	14.15
BM25-RAG	40.43	65.96	36.17	42.55	50.77	19.61
LightRAG	23.40	25.53	8.51	6.38	25.32	5.11
KGP	46.81	61.70	36.17	40.43	53.55	21.60
MDHSG-RAG	61.70	87.23	74.47	76.60	70.78	49.47

从表 1 可以看出，MDHSG-RAG 在各项检索指标上均取得最优结果。其中，Doc R@1 和 Doc R@5 分别达到 61.70%和 87.23%，说明该方法能够更准确地定位相关财政文件。与 Naive RAG 和 BM25 相比，MDHSG-RAG 不仅利用语义相似度和关键词匹配，还引入文档层级结构和结构模板约束，从而减少了相似政策文件之间的误召回。

在条文和条款级检索方面，MDHSG-RAG 的 Article R@5 和 Clause R@5 分别达到 74.47%和 76.60%，明显高于其他方法。这表明财政制度问答不能仅停留在文档级召回，还需要进一步定位到具体条文、条款或规则单元。Naive RAG 和 BM25 主要基于扁平文本块检索，缺少从文档到条款的结构化收缩过程；KGP 虽然引入图遍历机制，但其节点主要围绕通用段落图构建，对财政文档原生层级结构的利用仍然有限。

在严格证据召回方面，MDHSG-RAG 的 Strict Evidence R@5 达到 49.47%，显著高于 BM25 的 19.61%、

Naive RAG 的 14.15%和 LightRAG 的 5.11%。该结果说明，MDHSG-RAG 通过“文档级–条文级–条款级–内容块级–表格字段级”的逐层检索机制，能够将候选范围逐步收缩到可直接支撑答案的细粒度证据单元。

LightRAG 在本实验中的表现相对较弱，主要原因在于其图结构以实体和关系为核心，更适合表达概念关联和主题关系，但难以稳定表示财政制度文档中的章节层级、条款边界、附件依附和表格字段。因此，在需要精确定位制度依据的任务中，普通实体关系图容易召回语义相关但证据粒度不够精确的内容。

总而言之，MDHSG-RAG 的优势不只是提高文本召回率，同时结构感知图索引、问题类型识别和受控图扩展，实现了从文档到细粒度证据的逐层定位，为后续答案生成提供了更可靠的证据基础。

3.6. 生成阶段结果分析

生成阶段进一步考察不同方法在获得检索证据后能否生成正确、忠实且证据充分的答案。所有方法均采用相同生成模型，以减少模型能力差异对结果的影响。实验结果如表 2 所示：

Table 2. Performance comparison in the generation stage
表 2. 生成阶段结果

Approaches	Answer Accuracy	Faithfulness	Evidence Coverage	ROUGE-L
Naive RAG	40.64	41.91	29.15	27.97
BM25-RAG	32.34	37.87	24.89	28.41
LightRAG	11.28	15.53	7.87	22.94
KGP	35.11	59.57	34.36	30.91
MDHSG-RAG	54.47	57.63	37.45	26.11

MDHSG-RAG 的 Answer Accuracy 和 Evidence Coverage 分别达到 54.47%和 37.45%，均优于其他对比方法。该结果表明，基于层级结构组织的证据链能够为大语言模型提供更完整的上下文支撑。相比仅返回扁平文本块的 Naive RAG 和 BM25-RAG，MDHSG-RAG 保留了证据的文档来源、条文位置和内容层级，有助于减少无依据推断，并提高答案对关键制度依据的覆盖程度。

在 Faithfulness 指标上，KGP 取得 59.57%，略高于 MDHSG-RAG 的 57.63%。这说明 KGP 在部分样本中生成结果较为保守，但其 Answer Accuracy 和 Evidence Coverage 低于 MDHSG-RAG，关键内容覆盖仍存在不足。ROUGE-L 方面，KGP 和 BM25-RAG 得分较高，主要原因是其输出与参考答案具有更高的词面重合度。由于财政制度问答允许同义转述，ROUGE-L 更适合作为辅助指标，答案准确率、忠实度和证据覆盖率更能反映实际生成质量。

LightRAG 的生成效果相对较弱，主要受到检索阶段严格证据召回不足的影响。其实体关系图能够提供语义相关信息，但对条文边界、附件关系和表格字段等结构特征利用有限。总体来看，MDHSG-RAG 在保持较高忠实度的同时，取得了更优的答案准确率和证据覆盖率，说明结构感知图索引、分层检索和证据约束生成能够提升财政制度问答的综合效果。

4. 结论

针对财政制度文档问答中证据分散、结构异构、细粒度依据难以定位以及生成结果可追溯性不足等问题，本文提出了一种基于多文档异构结构图的检索增强生成方法(MDHSG-RAG)。该方法以制度文档的原生结构为基础，通过结构识别、模板匹配和动态模板扩展，将不同类型的文件统一组织为层级图索引。在检索阶段，根据问题类型选择相应模板，按照由粗到细的方式逐层缩小候选范围，并结合引用关系、

附件依附关系和表格关联关系补充证据；在生成阶段，将多粒度证据组织为包含来源路径的结构化上下文，以约束答案生成范围。实验结果表明，结构感知图索引与分层检索机制能够提升细粒度证据定位能力，基于证据链的上下文组织方式有助于提高答案准确性和证据覆盖程度。

当前方法仍存在一定的改进空间。其结构解析效果在一定程度上受到文档格式规范性的影响，对高度非结构化文本的适应能力仍需进一步验证。同时，模板维护以及图索引构建与检索会带来一定的维护和计算成本，在知识库规模持续扩大时，其运行效率仍有进一步优化的空间。

后续研究将重点从两个方面展开：一是研究结构识别、模板归纳与增量更新方法，提高系统对非标准格式文档及新增文档类型的自动适配能力，降低模板维护成本；二是通过增量索引、候选子图剪枝和关系缓存等方式优化检索效率，并进一步完善长表格、多附件和跨文件规则的关联表示与证据充分性判别机制，以提升复杂财政制度问答的准确性、稳定性与可追溯性。

参考文献

[1] 曹荣荣, 柳林, 于艳东, 等. 融合知识图谱的大语言模型研究综述[J]. 计算机应用研究, 2025, 42(8): 2255-2266.

[2] Lewis, P., Perez, E., Piktus, A., *et al*. (2020) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Proceedings of the* 34*th International Conference on Neural Information Processing Systems*, Vancouver, 6-12 December 2020, 9459-9474.

[3] Robertson, S. and Zaragoza, H. (2009) The Probabilistic Relevance Framework: BM25 and Beyond. *Foundations and Trends® in Information Retrieval*, **4**, 1-174. https://doi.org/10.1561/1500000019

[4] Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., *et al*. (2020) Dense Passage Retrieval for Open-Domain Question Answering. *Proceedings of the* 2020 *Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), Online, 16-20 November 2020, 6769-6781. https://doi.org/10.18653/v1/2020.emnlp-main.550

[5] Peng, B., Zhu, Y., Liu, Y., *et al*. (2024) Graph Retrieval-Augmented Generation: A Survey. *ACM Transactions on Information Systems*, **44**, 1-52.

[6] Edge, D., Trinh, H., Cheng, N., *et al*. (2024) From Local to Global: A GraphRAG Approach to Query-Focused Summarization. arXiv:2404.16130.

[7] Guo, Z., Xia, L., Yu, Y., Ao, T. and Huang, C. (2025) LightRAG: Simple and Fast Retrieval-Augmented Generation. *Findings of the Association for Computational Linguistics*: *EMNLP* 2025, Suzhou, 4-9 November 2025, 10746-10761. https://doi.org/10.18653/v1/2025.findings-emnlp.568

[8] Sarthi, P., Abdullah, S., Tuli, A., *et al*. (2024) RAPTOR: Recursive Abstractive Processing for Tree-Organized Retrieval. *Proceedings of the* 12*th International Conference on Learning Representations*, Vienna, 7-11 May 2024, 32628-32649.

[9] Wang, Y., Lipka, N., Rossi, R.A., Siu, A., Zhang, R. and Derr, T. (2024) Knowledge Graph Prompting for Multi-Document Question Answering. *Proceedings of the AAAI Conference on Artificial Intelligence*, **38**, 19206-19214. https://doi.org/10.1609/aaai.v38i17.29889

[10] Chen, W., Bai, T., Su, J., Luan, J., Liu, W. and Shi, C. (2025) Kg-Retriever: Efficient Knowledge Indexing for Retrieval-Augmented Large Language Models. 2025 *IEEE International Conference on Knowledge Graph* (*ICKG*), Limassol, 13-14 November 2025, 27-34. https://doi.org/10.1109/ickg66886.2025.00011

[11] Gong, Z., Huang, Y. and Mai, C. (2025) MMRAG-DocQA: A Multi-Modal Retrieval-Augmented Generation Method for Document Question-Answering with Hierarchical Index and Multi-Granularity Retrieval. arXiv:2508.00579.

[12] Xiang, Z., Wu, C., Zhang, Q., *et al*. (2026) When to Use Graphs in RAG: A Comprehensive Analysis for Graph Retrieval-Augmented Generation. *Proceedings of the* 14*th International Conference on Learning Representations*, Brazil, 23-27 April 2026.

[13] 晋艳峰, 黄海来, 林沿铮, 等. 基于知识表示学习的 KBQA 答案推理重排序算法[J]. 计算机应用研究, 2024, 41(7): 1983-1991.

[14] Luo, L., Li, Y.F., Haffari, G. and Pan, S. (2024) Reasoning on Graphs: Faithful and Interpretable Large Language Model Reasoning. *Proceedings of the* 12*th International Conference on Learning Representations*, Vienna, 7-11 May 2024.

[15] Hu, Y., Lei, Z., Zhang, Z., Pan, B., Ling, C. and Zhao, L. (2025) GRAG: Graph Retrieval-Augmented Generation. *Findings of the Association for Computational Linguistics*: *NAACL* 2025, Albuquerque, 29 April-4 May 2025, 4145-4157. https://doi.org/10.18653/v1/2025.findings-naacl.232

[16] He, X., Tian, Y., Sun, Y., Chawla, N., Laurent, T., Lecun, Y., *et al*. (2024) G-Retriever: Retrieval-Augmented Generation

for Textual Graph Understanding and Question Answering. *Advances in Neural Information Processing Systems* 37, Vancouver, 10-15 December 2024, 132876-132907. https://doi.org/10.52202/079017-4224

[17] Li, Z., Guo, Q., Shao, J., Song, L., Bian, J., Zhang, J., *et al*. (2025) Graph Neural Network Enhanced Retrieval for Question Answering of Large Language Models. *Proceedings of the* 2025 *Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics*: *Human Language Technologies* (Volume 1: Long Papers), Albuquerque, 29 April-4 May 2025, 6612-6633. https://doi.org/10.18653/v1/2025.naacl-long.337

[18] Sarmah, B., Mehta, D., Hall, B., Rao, R., Patel, S. and Pasquali, S. (2024) Hybridrag: Integrating Knowledge Graphs and Vector Retrieval Augmented Generation for Efficient Information Extraction. *Proceedings of the* 5*th ACM International Conference on AI in Finance*, Brooklyn, 14-17 November 2024, 608-616. https://doi.org/10.1145/3677052.3698671

[19] Li, Z., Deng, L., Liu, H., *et al*. (2024) UniOQA: A Unified Framework for Knowledge Graph Question Answering with Large Language Models. arXiv:2406.02110.

[20] Tang, X., Gao, Q., Li, J., *et al*. (2025) MBA-RAG: A Bandit Approach for Adaptive Retrieval-Augmented Generation through Question Complexity. *Proceedings of the* 31*st International Conference on Computational Linguistics*, Abu Dhabi, 19-24 January 2025, 3248-3254.