

# 基于大语言模型和 RAG 的舆情知识交互及推荐研究

马钰<sup>1</sup>, 任世明<sup>1</sup>

(1. 郑州升达经贸管理学院 信息工程学院 河南郑州 451191)

**摘要:** 随着社交媒体与网络舆论变革, 海量舆情信息呈碎片化、实时化及多源异构化等特点, 对公共决策及社会稳定具有重大影响。本文基于大语言模型与检索增强生成 (RAG) 技术, 设计并实现舆情知识交互系统。系统构建领域知识库, 利用 RAG 高效检索文档片段, 并将检索结果与用户提问输入 LLM, 生成准确而丰富的舆情解读; 同时嵌入推荐子模块, 依托提问语境主动拓展相关话题, 提升回答的启发性与可读性。本系统为构建高效可交互的舆情知识服务平台提供了新技术路径, 并为公共管理与舆情治理领域的信息服务提供了有力支持。

**关键词:** 大语言模型; RAG; 舆情知识交互

**基金项目:** 郑州升达经贸管理学院 2025 年度立项课题: 融合协同过滤与大语言模型的智能推荐系统研究 (编号: SD-YB2025018)。

DOI: doi.org/10.70693/rwsk.v1i5.1027

## 一、引言

### (一) 研究背景

近年来, 社交媒体与网络舆论环境发生了深刻变革, 传统信息传播方式逐渐被微博、微信、短视频等多层次、多渠道的碎片化信息所取代。与此同时, 舆情信息具有实时化特征, 短时间内海量数据涌现且来源多元, 涵盖新闻报道、公众评论、问答互动等异构内容。这些数据既蕴含丰富舆情动向, 又存在噪声信息与冗余信息并存的挑战, 使得精准、高效地获取关键信息成为急需突破的难题。因此, 亟需构建一套能够支持多层次、多维度深度查询与交互分析的舆情知识服务平台。

近年来, 大语言模型凭借其强大的语义理解能力, 为自然语言处理领域带来突破性进展。然而, 其在知识时效性、领域适配性等方面仍存在瓶颈。检索增强生成 (RAG) 技术可在模型内部融入外部知识库, 为解决上述问题提供了新思路。基于此, 本系统引入大语言模型 (LLMs) 与检索增强生成 (RAG) 技术, 旨在突破传统关键词检索与模板式问答的局限, 为舆情知识获取、交互及推荐提供一种更具智能化、可拓展性的技术路径。

### (二) 研究意义

本研究在理论层面丰富了舆情知识服务体系的技术视角, 将大语言模型与 RAG 检索机制有机结合, 为舆情信息处理提供了一种全新的范式。该范式既能发挥检索模块对异构文档的高效定位能力, 也能借助大语言模型在自然语言理解与生成上的优势, 实现精确且富有深度的舆情知识解读。此外, 研究中嵌入的推荐子模块, 可根据用户提问语境主动延展相关话题, 提升系统响应的启发性与可读性, 推动从单一问答向智能化交互式知识服务的转变。

在实践层面, 所提出的舆情知识交互系统可为相关企业或部门辅助决策支持, 为公共管理、媒体传播、社会治理等领域提供高效、智能的信息服务解决方案, 具有较高的应用价值与推广前景。

## 二、文献综述

知识交互问答系统旨在实现用户与系统之间的动态对话与内容探索。传统的问答系统多基于静态关键词检索

**作者简介:** 马钰 (1997—), 男, 河南郑州人, 郑州升达经贸管理学院助教, 硕士研究生, 研究方向: 自然语言处理、舆情分析。

任世明 (2003—), 男, 河南商丘人, 郑州升达经贸管理学院本科生。

和预设模板,缺乏深度交互与智能引导。近年来,基于大模型技术的智能问答逐渐成为研究热点,大语言模型是一种基于深度学习技术的自然语言处理模型,拥有数十亿甚至上千亿个参数,不仅能够理解和生成自然语言,还能显著提升问答系统的准确性和效率<sup>[1]</sup>。借助 LLM 技术,研究者提出了多种问答交互策略:一是基于提示学习的问答系统,如方全等人为提升知识问答系统的性能,提出了一种新的基于组合上下文提示的大型语言模型领域知识问答方法<sup>[2]</sup>;二是基于知识图谱的问答系统,如李锦等人为解决在陶瓷收藏及鉴赏领域研究人员获取相关知识时面临知识匮乏和知识难有效统一表达的问题,开发了融合前沿技术的五六七瓷知识图谱问答系统<sup>[3]</sup>;三是基于检索增强生成的问答系统,如杨诗卓等人融合 RAG 检索增强与 LoRA 微调等技术,设计了一种基于 Qwen 2.5 大语言模型的智能医疗问答系统<sup>[4]</sup>;四是基于特定领域知识库的问答系统,如李明等人构建了中医古籍方剂数据挖掘与知识问答系统<sup>[5]</sup>,黄冰等人构建了古生物学领域中知识问答系统<sup>[6]</sup>。

本文针对舆情领域大语言模型知识更新滞后、领域适配性不足及个性化推荐服务能力受限的瓶颈问题,设计了一种基于 RagFlow 的舆情信息智能问答系统。通过大语言模型、RAG 检索增强算法、构建舆情领域专用问答数据集等方法,实现动态知识融合与轻量化领域适配的协同优化,使用户能够通过自然语言询问舆情事件相关信息,如事件背景、演变路径和关联主体等。

### 三、相关技术研究

#### (一) 大语言模型

基于 Transformer 架构的大语言模型 (Large Language Model, LLM) 是通过海量文本数据进行无监督预训练,以学习语言的统计规律和语义表示的人工智能模型<sup>[7]</sup>。在预训练阶段,模型使用自回归或自编码目标,例如 GPT 系列采用下一词预测任务, BERT 系列采用掩码语言模型任务,通过多层自注意力机制捕捉长程依赖。随后,使用者可通过微调 (fine-tuning)、提示工程 (prompt engineering) 或少样本学习 (few-shot learning) 等技术,将通用语言知识迁移到特定任务,如文本分类、问答生成、摘要撰写和机器翻译等。得益于模型规模和数据量的持续增长, LLM 展现了强大的自然语言理解与生成能力,并在零样本或少样本场景中表现出色。

#### (二) 检索增强生成技术

检索增强生成 (Retrieval-Augmented Generation, RAG) 是一种将大规模文本检索与生成模型结合的混合框架,旨在提升生成结果的准确性和知识覆盖度。RAG 主要由两部分组成:一是检索模块 (Retriever),通常基于倒排索引或向量检索,从预先构建的文档库中选取与查询最相关的若干文本片段;二是生成模块 (Generator),将检索到的上下文与原始输入一起作为条件,通过大型生成模型 (如 GPT) 进行回答或文本生成<sup>[8]</sup>。相比于纯生成式模型, RAG 能够借助外部知识库补充最新信息,显著提升事实性和一致性。

### 四、基于大语言模型和 RAG 的舆情知识交互及推荐系统方法研究

#### (一) 系统总体架构介绍

本研究提出了一种基于 RAGFlow 引擎的舆情解读与推荐系统,旨在实现对海量、多源舆情信息的高效检索、准确解读及主动推荐。系统整体架构可划分为四个核心模块:数据层、检索增强层、推理层和推荐层。

数据层通过定期抓取文献库、社交平台、新闻网站和论坛等渠道文本,并进行清洗、分词、去重和标准化,存储于向量化数据库中;检索增强层利用 RAGFlow 引擎构建向量查询,召回相似度最高的文档片段,并结合 TF-IDF 与余弦相似度进行二次排序与去冗余,最终将精炼片段与用户提问拼接为推理输入;推理层通过 API 调用大语言模型 (如 DeepSeek 等),基于结构化 Prompt 生成要点概括、趋势研判与风险评估,并对输出结果进行格式校验与术语规范化;推荐层通过意图识别和对话上下文分析,调用检索增强流程获取相关背景资料或典型案例,结合话题相关性和启发价值生成推荐项,以列表形式呈现。各模块协同构成从数据摄取到智能输出的闭环流程,为舆情监测与决策提供了高效、可靠且可扩展的技术方案。

#### (二) 舆情知识交互系统构建及系统处理流程

在本系统中,当用户发出舆情咨询请求后,首先将用户的自然语言提问传入 RAGFlow 引擎。RAGFlow 根据预先构建的多源舆情知识库 (包含舆情文献、网络舆情文本、专家报告等异构数据),通过 RAGFlow 引擎结合 BM25 与向量检索快速定位相关文档片段;这些片段即以“知识提示” (Knowledge Prompt) 的形式,与用户原始问题一并送入大语言模型 (如 DeepSeek) 的接口。模型在接收到用户问题与检索到的高置信度片段后,首先对信息进行融合与语义对齐,生成准确且富有深度的舆情解读;与此同时,系统内嵌的推荐子模块会基于当前对话上下文主动识别潜在相关话题,补充提示或衍生问题,从而引导模型输出更多启发性见解。最终,用户将获

严谨又可读性强的舆情分析答复，整个流程实现了“用户询问→RAG 检索→LLM 生成→推荐扩展→结果呈现”的闭环交互。系统处理流程图如图 1 所示：

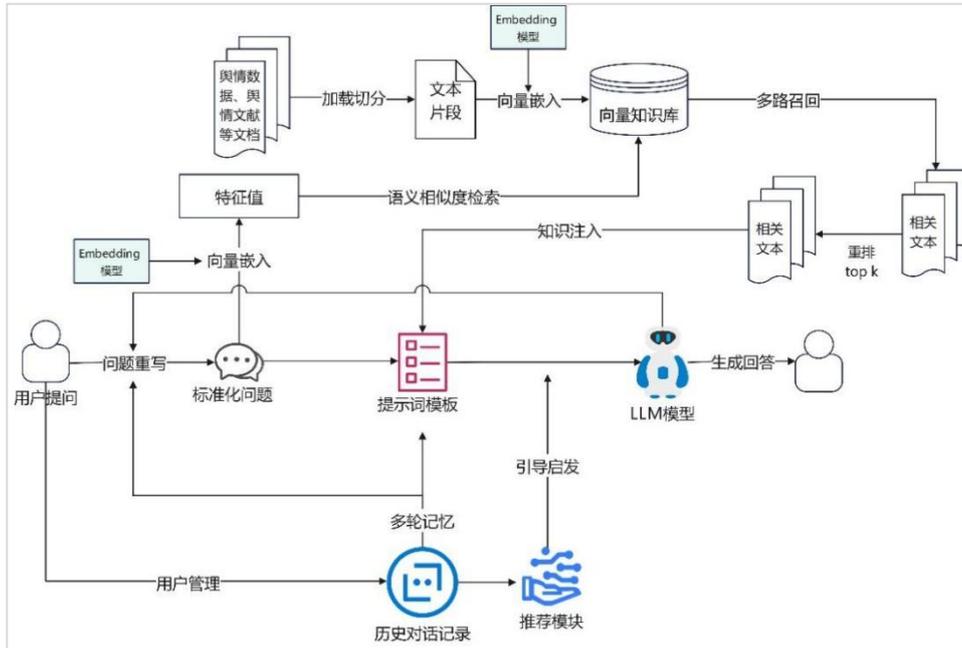


图 1 舆情知识交互系统处理流程图

### 五、实证分析

为验证所设计的舆情知识交互与推荐系统的实用性，本文组织了多轮模拟问答与真实用户评估。本系统在构建的舆情知识库的基础上，调用 DeepSeek-R1 大模型进行实例测试。测试结果表明，本系统平均响应时长仅为 1~2 秒，远超传统人工舆情分析（通常需 30 秒~1 分钟），显著减少了用户等待时间；在处理常见舆情概念和热点事件细节时，能够从知识库中提供既准确又深入的解读，并基于上下文主动补充背景资料。推荐子模块在用户提问后，会自动拓展相关话题，例如，当用户询问“舆情是什么？”时，系统不仅阐述了舆情的定义，还阐述了舆情的演化过程及舆情的复杂性，如图 2 所示。此外，在面向实际业务人员的盲测反馈中，87%的参与者认可系统生成内容的深度与可读性，83%的参与者表示推荐话题切中需求并具有启发性。总体来看，该舆情知识交互及推荐系统在系统性能、解读深度和用户体验三方面均表现出色，显著提升了舆情分析效率与决策支持能力，更能满足用户在实时舆情监测与分析中的多样化需求。



图 2 舆情知识交互系统交互界面

## 六、结语

本文基于大语言模型与检索增强生成 (RAG) 技术, 构建了一个集知识交互与话题推荐于一体的舆情分析系统。系统通过领域知识库的构建与高效检索, 能够在用户提出问题时快速筛选出相关文档片段, 并结合大语言模型生成具有深度解读价值的舆情分析; 同时, 推荐模块依托提问语境主动拓展关联话题, 提升了回答的启发性与可读性。实验结果表明, 该系统在回答准确性和用户满意度方面均优于传统单一生成或检索方法。未来研究可进一步丰富知识库的多模态信息源, 探索更高效的检索策略, 以及引入用户反馈机制以实现个性化优化, 从而构建更智能、更贴合实际需求的舆情交互与推荐平台。

### 参考文献:

- [1] 任海玉, 刘建平, 王健, 等. 基于大语言模型的智能问答系统研究综述[J]. 计算机工程与应用, 2025, 61(07): 1-24.
- [2] 方全, 张金龙, 王冰倩, 等. 基于组合上下文提示的大型语言模型领域知识问答研究[J/OL]. 计算机科学, 1-13.
- [3] 李锦, 刘陶, 邱望仁. 融合前沿技术的五七瓷知识图谱问答系统创新开发[J/OL]. 计算机应用与软件, 1-8.
- [4] 杨诗卓, 周燕玲, 解翔杰. 融合 RAG 检索增强与 LoRA 微调的医疗问答系统优化研究[J/OL]. 江西科学, 1-8.
- [5] 李明, 罗晓兰, 朱邦贤. 中医古籍方剂数据挖掘与知识问答系统构建[J]. 图书馆论坛, 2025, 45(04): 49-59.
- [6] 黄冰. 大语言模型在古生物学中的应用初探——以基于 RAG 的知识问答系统为例[J]. 古生物学报, 2025, 64(01): 100-112.
- [7] 赵正平. 人工智能大语言模型和 AI 芯片的新进展[J]. 微纳电子技术, 2025, 62(03): 7-37.
- [8] 刘雪颖, 云静, 李博, 等. 基于大型语言模型的检索增强生成综述[J/OL]. 计算机工程与应用, 1-31.

# Research on Interactive Public Sentiment Knowledge and Recommendation Based on Large Language Models and RAG

Yu Ma<sup>1</sup>, Shiming Ren<sup>1</sup>

<sup>1</sup>*School of Information Engineering, Zhengzhou Shengda University, Zhengzhou, Henan 451191, China*

**Abstract:** With the advent of social media and shifts in online public opinion, massive volumes of sentiment information have become fragmented, real-time, and multi-source heterogeneous, exerting profound effects on public decision-making, corporate reputation, and social stability. This paper, leveraging large language models and retrieval-augmented generation (RAG) techniques, designs and implements an interactive knowledge system for public sentiment. The system constructs a domain-specific knowledge base, efficiently retrieves document snippets via RAG, and feeds both retrieved results and user queries into an LLM to produce accurate and comprehensive sentiment analyses. At the same time, it embeds a recommendation submodule that proactively expands related topics based on the query context, thereby enhancing the insightfulness and readability of the responses. The proposed system offers a novel technical approach for building an efficient, interactive platform for sentiment knowledge services and provides robust support for information services in the fields of public administration and opinion governance.

**Key Words:** Large Language Model; RAG; Interactive Public Opinion Knowledge