**开题报告**

|  |  |
| --- | --- |
| 开题论文题目 | 结合知识图谱的检索增强生成算法研究与应用 |
| 研究背景 | 低空经济依托低空空域，以有人驾驶和无人驾驶航空器的低空飞行活动为牵引，广泛涵盖物流配送、精准农业、空中交通、城市管理、低空旅游等种应用场景，是影响并带动相关领域融合发展的综合性经济形态[1]。低空经济不仅拓宽了传统航空业的边界，也为社会经济注入了新的活力与创新动力，其蓬勃发展的背后，是政策法规体系的日益完善。国家层面的政策扶持，如《通用航空装备创新应用实施方案（2024—2030年）》的出台，为行业的规范化、法制化发展奠定了坚实基础。除了中央层面在大力部署，各省市也在加速布局低空经济。据不完全统计，截至2024年9月，全国已有27个省、自治区、直辖市将低空经济写进了政府工作报告，发布了合计174份低空经济相关政策文件。  现有的低空经济语料多为非结构化文本，各地区发布的低空经济政策文件在结构和内容上相似，术语高度统一。这些语料篇幅较长、专业性强，理解和应用较为困难。此外，语料分布零散，缺乏统一管理，增加了信息获取和利用的难度。在此背景下，基于大规模预训练语言模型（Large Language Model, 大模型）技术，展现出了在低空经济领域的广泛应用潜力。大模型通过对海量数据和大量参数的训练，能够高效完成自然语言生成任务，并且在开放域自然语言理解方面具有较强的能力[2]。因此，它在智能问答等领域得到了广泛应用[3-5]。然而，由于缺乏针对特定领域的训练，大模型在生成特定领域内容时可能缺乏准确性，甚至会产生所谓的“幻觉事实”。  为了解决这一问题，检索增强生成（Retrieval Augmented Generation, RAG）[22]技术引入了外部知识库，模型可以通过先检索相关知识后再生成回答，从而提升生成结果的准确性。传统RAG方法通常使用向量库存储知识，但这种方式缺乏知识的结构化信息。相比之下，知识图谱通过挖掘、分析和展示知识及其相互关系，能够更加智能地组织和表示知识。结合知识图谱的智能问答系统，通过挖掘实体间的直接关联，可以推理出潜在关系，不再是简单的文档排序，而是通过语义分析反馈更接近用户需求的答案[6]。在航天领域，基于知识图谱的飞机电源系统故障诊断技术，已经实现了对领域内实体和关系数据的全面整合，为故障知识搜索、问答和推理提供了丰富的知识基础[7]。目前，国内外学者在知识图谱问答领域已经取得了许多先进性研究成果[6-8]，但在低空经济领域的专项研究尚显不足，难以满足该领域对深度、精准信息的需求。 |
| 研  究  目  标  与  意义 | 研究目标：  本研究提出了一种结合知识图谱的检索增强生成算法，通过融合大模型与知识图谱技术，优化传统基于向量库的检索增强生成算法。构建一个面向低空经济领域的语料分析系统（Low Altitude Economy-GPT, LAE-GPT），通过结合大模型的强大自然语言处理能力与知识图谱的结构化知识表示能力，该系统能够实现低空经济领域信息的深度挖掘与精准分析，为行业用户提供专业、准确的信息服务与支持，推动低空经济领域的智能化发展。  研究意义：  1.优化传统检索增强生成技术：  提出了三级检索机制，包括细粒度检索、粗粒度检索和网络搜索，以充分利用知识图谱的结构化优势，避免上下文冗长和信息冗余的问题。这种基于知识图谱的检索增强生成方法显著优化了传统RAG技术的性能，特别是在面向复杂的垂直领域时，能够更好地处理多维度和多层次的知识结构。通过提高检索的精准性和效率，该系统有助于用户快速获取所需信息，减少信息检索的时间和成本。  2.为低空经济领域提供信息服务与决策支持：  通过结合大模型的强大自然语言处理能力与知识图谱的结构化知识表示能力，LAE-GPT系统能够实现低空经济领域信息的深度挖掘与精准分析，为行业用户提供专业、准确的信息服务与支持。此外，LAE-GPT系统能够生成详尽专业的低空经济分析报告，深入剖析低空经济的行业发展、政策支持、产业链布局及其面临的挑战等方面。该报告附上了参考的真实信息来源，实现了知识溯源功能，确保了报告的权威性和可信度。通过为学术界、政策制定者及产业从业者提供有力的参考依据，该系统能够更好地为他们在决策和研究过程中提供帮助，促进低空经济领域的可持续发展。 |
| 国内外研究现状 | 1. 大模型技术  在近年来的自然语言处理（NLP）领域，大模型已经成为研究和应用的热点，以其卓越的性能和广泛的应用前景，推动了NLP技术的快速发展。大模型是指包含数千亿（或更多）参数的语言模型[9]，在海量文本数据[10]上进行训练，例如GPT-3 [11]、PaLM [12]、Galacica [13] 和LLaMA [14]。 具体来说，大模型建立在Transformer架构[15]之上，其中多头注意力层堆叠在非常深的神经网络中。大模型通常采用两阶段学习策略：预训练（Pre-training）和微调（Fine-tuning）。预训练阶段，模型在大量无标注文本上进行学习，获取语言的通用表示；微调阶段，则针对特定任务的数据集调整模型参数，以适应具体的应用场景。  大模型在自然语言处理领域的应用广泛，包括但不限于文本分类、问答系统、机器翻译、文本摘要等任务。在复杂问答任务中，大型模型能够通过理解上下文和长距离依赖关系，提供准确的答案[16]。在文本生成任务中，它们能够生成连贯、相关且语法正确的文本，展现出强大的创造性写作能力[11]。  研究发现，大模型的性能随着模型规模的增大而显著提升，即所谓的“规模效应”。但大模型同时面临着一些挑战。首先是缺乏领域专业知识：大模型通常是在大规模通用语料库上训练的，这些语料库涵盖了广泛的主题，但往往不够深入或具体于某一特定领域。当处理特定领域的文本时，模型可能无法理解或准确解释该领域的专业术语、缩写、行业惯例或复杂概念[17]。如果进行模型的微调，又需要耗费昂贵的资源和时间[18]。其次，大模型在生成过程可能面临幻觉问题，生成不准确或者不一致的信息[19]，特别是在使用思维链（CoT）等提示技术时，幻觉可能会逐渐累积[20]。另外，大模型通常缺乏最新信息，因为存储在模型参数中的知识是静态的，并且不会随着时间的推移进行同步更新[20]，这限制了它们在一些快速变化的领域的应用。  2. 检索增强生成技术  为了解决上述提出的问题，检索增强生成（Retrieval Augmented Generation, RAG）[22]技术被提出。RAG是将传统的信息检索和生成模型相结合的一种方法。一个典型的RAG系统包括：知识库包含大量结构化或非结构化的知识数据，可以是文本文档、数据库、网页等。将知识库中的文本进行切分，嵌入到向量库中，之后检索模块利用信息检索技术，从向量库中检索到与用户查询相关的Top-K个文档或段落。最后，大模型基于用户输入和检索到的相关信息生成回答。  早期的RAG方法主要依赖于静态和有限的语料库进行检索，但随着技术的发展，研究者们开始探索使用大规模网络搜索作为扩展来增强检索结果。这种方法不仅能够返回更多样化的信息，还能够提高检索的准确性和时效性[23]。GaoY等[24]研究了RAG范式的进展，包括朴素RAG、高级RAG和模块化RAG。ZhaoP等[25]把RAG基础范式分为四个类：基于查询的RAG、基于潜在表示的RAG、基于Logit的RAG和投机性RAG。在RAG技术的发展过程中，研究者们提出了多种优化策略。在输入增强方面，HyDE [26]使用查询转换的方式，利用query生成一个伪文档进行检索，从而使检索到的结果更加准确。RQ-RAG[27]提供查询重写、分解和消岐功能来增强模型。BlendedRAG[28]使用语义搜索和基于混合查询的检索器来提高RAG准确性。Self-RAG[29]通过使用经过训练的生成器根据不同指令下的检索词符来确定是否进行检索，并评估检索到文档的相关性和质量。CRAG[30]训练一个轻量级检索器评估检索到文档的质量并返回置信度，根据不同置信度进行不同的只是检索操作。REAR[31]提出了一种基于双粒度相关性融合和抗噪训练的改进训练方法,可以通过有效地感知检索到的文档的相关性来更好地利用外部知识。一些学者把RAG和微调进行结合，研究[32]说明了如何以端到端的方式微调整个检索增强生成（RAG）架构。RAFT[33]提出检索增强微调训练策略，训练模型忽略无关文档并直接引用相关文档内容来回答问题，在多个数据集上取得显著效果。尽管RAG技术在提高大模型生成质量方面取得了显著进展，但仍面临一些挑战。首先，传统的RAG模型在处理文本时，往往局限于捕捉基于语义相似性的信息，而难以捕捉到仅通过语义相似性无法全面体现的结构化关系知识。其次，当RAG模型将多个文本片段串联作为生成提示时，常常会出现内容上的重复叙述，导致生成的文本上下文变得冗长复杂，降低了信息的有效性和阅读体验。另外，RAG缺乏全局视角，仅能通过检索到的上下文信息生成回答，在一些需要广泛上下文理解和综合分析的任务时难以达到理想的效果[34]。  3. 知识图谱技术  知识图谱本质上是语义网络（Semantic Network）的知识库，不仅包含实体、关系和属性信息，还涵盖实体与关系间的语义关联。知识图谱由节点、边和属性三部分构成：节点表示现实世界或抽象概念中的实体，如个人、企业、事件等；边则描述节点间的关系或属性；属性则提供关于实体的额外信息，如创建日期或所属年份等。  知识图谱的构建是一个系统性过程，通常涉及从结构化、半结构化或非结构化数据中进行知识抽取、融合与推理，并通过多次迭代逐步补全知识。构建方式可以分为自顶向下和自底向上两种：自顶向下方法利用现有结构化知识库进行抽取与构建，而自底向上方法则从非结构化文本中提取高置信度三元组进行构建[35]。  近年来，知识图谱的概念已广泛渗透至多个行业领域，催生了基于知识图谱的智能问答系统的兴起与快速发展。在金融、医疗健康、旅游服务、农业科技以及电子商务等垂直行业内，针对知识图谱应用的研究层出不穷。例如，研究[36]开发的基于疾病知识图谱的智能问答系统，研究[37]构建的面向电商领域的知识图谱问答平台，均展示了知识图谱问答在具体场景下的应用潜力。  基于知识图谱的智能问答系统的构建策略多样，有的侧重于结合知识图谱中的丰富数据与深度学习技术，共同构建出强大的问答能力；有的则充分利用知识图谱的推理机制，深入解析用户提问背后的意图；还有的系统通过将用户问题与知识图谱中的三元组信息编码至统一的向量空间内，利用空间内的相似度计算，高效地检索并呈现与用户问题最为接近的答案。这些方法不仅提升了问答系统的准确性和效率，也进一步推动了知识图谱技术在实际应用中的深化与拓展。  参考文献：  [1] 刘先江，宋丹，徐政. 以低空经济发力打造新质生产力发展新引擎[J/OL]. 北京航空航天大学学报（社会科学版）, （2024-07-15）[2024-09-23]. https://doi.org/10.13766/j.bhsk.1008-2204.2024.0545.  LIU X J, SONG D, XU Z. Using low altitude economy to build a new Engine for the development of new quality productive forces[J/OL]. Journal of Beijing University of Aeronautics and Astronautics(Social Sciences Edition),（2024-07-15）[2024-09-23]. https://doi.org/10.13766/j.bhsk.1008-2204.2024.0545.  [2] 车万翔, 窦志成, 冯岩松, 等. 大模型时代的自然语言处理:挑战、机遇与发展[J]. 中国科学:信息科学, 2023, 53(09):1645-1687.  CHE W X, DOU Z C, FENG Y S, et al. Towards a comprehensive understanding of the impact of large language models on natural language processing: challenges, opportunities and future directions[J]. Sci Sin Inform, 2023, 53:1645–1687(in Chinese).  [3] 张春红, 杜龙飞, 朱新宁, 等. 基于大语言模型的教育问答系统研究[J]. 北京邮电大学学报(社会科学版), 2023, 25(06):79-88.  ZHANG C H, DU L F, ZHU X N, et al. Educational question-answering systems based on large language model[J]. Journal of Beijing University of Posts and Telecommunications(Social Sciences Edition), 2023, 25(06):79-88(in Chinese).  [4] 王昀, 胡珉, 塔娜, 等. 大语言模型及其在政务领域的应用[J]. 清华大学学报(自然科学版), 2024, 64(04):649-658.  WANG Y, HU M, TA N, et al. Large language models and their application in government affairs[J]. Journal of Tsinghua University(Science and Technology), 2024, 64(04):649-658(in Chinese).  [5] 王婷, 王娜, 崔运鹏, 等. 基于人工智能大模型技术的果蔬农技知识智能问答系统[J]. 智慧农业(中英文), 2023, 5(04):105-116.  WANG T, WANG N, CUI Y P, et al. Agricultural technology knowledge intelligent question-answering system based on large language model[J]. Smart Agriculture, 2023, 5(4): 105-116(in Chinese).  [6] 王智悦, 于清, 王楠, 等. 基于知识图谱的智能问答研究综述[J]. 计算机工程与应用, 2020, 56(23):1-11.  WANG Z Y, YU Q, WANG N, et al. Survey of intelligent question answering research based on knowledge graph[J]. Computer Engineering and Applications, 2020, 56(23):1-11(in Chinese).  [7] 聂同攀, 曾继炎, 程玉杰, 等. 面向飞机电源系统故障诊断的知识图谱构建技术及应用[J]. 航空学报, 2022,43(8):625499-625499.  NIE T P, ZENG J Y, CHENG Y J, et al. Knowledge graph construction technology and its application in aircraft power system fault diagnosis[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(8): 625499-625499 (in Chinese).  [8] 姚元杰, 龚毅光, 刘佳, 等. 基于深度学习的智能问答系统综述[J]. 计算机系统应用, 2023, 32(04):1-15.  YAO Y J, GONG Y G, LIU J, et al. Survey on intelligent question answering system based on deep learning[J]. Computer Systems & Applications, 2023, 32(4):1−15(in Chinese).  [9] 张金营, 王天堃, 么长英, 等. 基于大语言模型的电力知识库智能问答系统构建与评价[J/OL]. 计算机科学, （2024-05-28）[2024-09-23]. http://kns.cnki.net/kcms/detail/50.1075.TP.20240528.0931.002.html.  ZHANG J Y, WANG T K, YAO C Y, et al. Construction and evaluation of intelligent question answering system for electric power knowledge base based on large language model[J/OL]. Computer Science, (2024-05-28)[2024-09-23]. http://kns.cnki.net/kcms/detail/50.1075.TP.20240528.0931.002.html.  [10] GAO J, LIN C Y. Introduction to the special issue on statistical language modeling[J]. ACM Transactions on Asian Language Information Processing (TALIP), 2004, 3(2): 87-93.  [11] SHANAHAN M. Talking about large language models[J]. Communications of the ACM, 2024, 67(2): 68-79.  [12] BROWN T B. Language models are few-shot learners[J]. arXiv preprint arXiv:2005.14165, 2020.  [13] CHOWDHERY A, NARANG S, DEVLIN J, et al. Palm: Scaling language modeling with pathways[J]. Journal of Machine Learning Research, 2023, 24(240): 1-113.  [14] TAYLOR R, KARDAS M, CUCURULL G, et al. Galactica: A large language model for science[J]. arXiv preprint arXiv:2211.09085, 2022.  [15] TOUVRON H, LAVRIL T, IZACARD G, et al. Llama: Open and efficient foundation language models[J]. arXiv preprint arXiv:2302.13971, 2023.  [16] LIU Y, OTT M, GOYAL N, et al. RoBERTa: A robustly optimized BERT pretraining approach. arXiv Preprint[J]. arXiv preprint arXiv:1907.11692, 1907.  [17] LONG C, LIU Y, OUYANG C, et al. Bailicai: A Domain-Optimized Retrieval-Augmented Generation Framework for Medical Applications[J]. arXiv preprint arXiv:2407.21055, 2024.  [18] OUYANG L, WU J, JIANG X, et al. Training language models to follow instructions with human feedback[J]. Advances in neural information processing systems, 2022, 35: 27730-27744.  [19] WANG H, ZHAO T, GAO J. BlendFilter: Advancing retrieval-augmented large language models via query generation blending and knowledge filtering[J]. arXiv preprint arXiv:2402.11129, 2024.  [20] WEI J, WANG X, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. Advances in neural information processing systems, 2022, 35: 24824-24837.  [21] VU T, IYYER M, WANG X, et al. Freshllms: Refreshing large language models with search engine augmentation[J]. arXiv preprint arXiv:2310.03214, 2023.  [22] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks[J]. Advances in Neural Information Processing Systems, 2020, 33: 9459-9474.  [23] SU W, TANG Y, AI Q, et al. Dragin: Dynamic retrieval augmented generation based on the real-time information needs of large language models[J]. arXiv preprint arXiv:2403.10081, 2024.  [24] GAO Y, XIONG Y, GAO X, et al. Retrieval-augmented generation for large language models: A survey[J]. arXiv preprint arXiv:2312.10997, 2023.  [25] ZHAO P, ZHANG H, YU Q, et al. Retrieval-augmented generation for ai-generated content: A survey[J]. arXiv preprint arXiv:2402.19473, 2024.  [26] GAO L, MA X, LIN J, et al. Precise zero-shot dense retrieval without relevance labels[J]. arXiv preprint arXiv:2212.10496, 2022.  [27] CHAN C M, XU C, YUAN R, et al. Rq-rag: Learning to refine queries for retrieval augmented generation[J]. arXiv preprint arXiv:2404.00610, 2024.  [28] SAWARKAR K, MANGAL A, SOLANKI S R. Blended rag: Improving rag (retriever-augmented generation) accuracy with semantic search and hybrid query-based retrievers[J]. arXiv preprint arXiv:2404.07220, 2024.  [29] ASAI A, WU Z, WANG Y, et al. Self-rag: Learning to retrieve, generate, and critique through self-reflection[J]. arXiv preprint arXiv:2310.11511, 2023.  [30] YAN S Q, GU J C, ZHU Y, et al. Corrective retrieval augmented generation[J]. arXiv preprint arXiv:2401.15884, 2024.  [31] WANG Y, REN R, LI J, et al. Rear: A relevance-aware retrieval-augmented framework for open-domain question answering[J]. arXiv preprint arXiv:2402.17497, 2024.  [32] SIRIWARDHANA S, WEERASEKERA R, WEN E, et al. Fine-tune the entire rag architecture (including dpr retriever) for question-answering[J]. arXiv preprint arXiv:2106.11517, 2021.  [33] ZHANG T, PATIL S G, JAIN N, et al. Raft: Adapting language model to domain specific rag[J]. arXiv preprint arXiv:2403.10131, 2024.  [34] LIU N F, LIN K, HEWITT J, et al. Lost in the middle: How language models use long contexts[J]. Transactions of the Association for Computational Linguistics, 2024, 12: 157-173.  [35] CHEN X, JIA S, XIANG Y. A review: Knowledge reasoning over knowledge graph[J]. Expert systems with applications, 2020, 141: 112948.  [36] 李贺，刘嘉宇，李世钰等 . 基于疾病知识图谱的自动问答系统优化研究［J］. 数据分析与知识发现, 2021,5(5)：115-126.  LI H, LIU J Y, LI S Y, et al. Optimizing automatic question answering system based on disease knowledge graph[J]. Data Analysis and Knowledge Discovery, 2021, 5(5): 115-126(in Chinese).  [37] 杜泽宇, 杨燕, 贺樑. 基于中文知识图谱的电商领域问答系统[J]. 计算机应用与软件, 2017, 34(5): 153-159.  DU Z Y, YANG Y, HE L. Question answering system of electric business field based on Chinese knowledge map[J]. Computer Applications and Software, 2017, 34(5): 153-159(in Chinese).  [38] TRAAG V A, WALTMAN L, VAN ECK N J. From Louvain to Leiden: guaranteeing well-connected communities[J]. Scientific reports, 2019, 9(1): 1-12. |
| 研究内容 | 1. 结合大模型的知识图谱构建方法  本研究首先致力于利用各种语料数据构建知识图谱，并进行有效的索引。这一过程包括从各种来源（如政策文件、行业报告、学术论文等）收集语料数据并进行预处理，然后通过文本切分、节点和关系提取等步骤，构建出包含实体、关系和属性的知识图谱。为了确保图谱的准确性和完整性，研究还采用了多次实体提取和子图合并策略，以整合不同数据块中的相关信息，并最终生成一个全局的知识图谱。此外，研究还为每个实体节点和关系边生成了文本键值对，以实现高效的检索。  2. 三层检索机制  在构建了知识图谱之后，本研究设计并实施了三层检索机制，以支持对用户查询的精准响应。这三层检索机制包括：细粒度检索，专注于检索具体的实体或关系；粗粒度检索，涉及更广泛的子图或全局图，以提供全面的相关信息；以及网络搜索，用于补充知识图谱中缺失的信息。通过这三种检索机制的组合使用，研究确保了检索到的内容与用户查询高度相关，同时充分利用了知识图谱的结构化优势，避免了上下文冗长和信息冗余的问题。  3.构建LAE-GPT系统  最后，本研究构建了一个面向低空经济领域的语料分析系统（LAE-GPT）。该系统能够实现低空经济语料自动化收集、智能问答和报告生成的功能。LAE-GPT通过网络爬虫的方式自动化收集低空经济语料并进行预处理，构建一个能够实现动态更新的低空经济知识图谱。之后，系统利用三层检索机制，在知识图谱和网络搜索中获取相关信息生成回答。通过在大模型中引入领域知识，LAE-GPT能够更准确专业地回答低空经济相关的问题。此外，LAE-GPT能够根据查询在知识图谱中检索相关信息，生成可视化低空经济分析报告，为政策分析、决策支持与研究工作提供了有力的技术支撑。 |
| 研究方案 | 1. 结合大模型的知识图谱构建方法  1.1 文本切分  将较长的文本分割成更小、更易于处理或分析的单元，以改善文本处理和信息检索的效率。考虑到大模型的上下文窗口限制，将文本划分为最大长度的块，并设置一定的重叠，以保留上下文信息。  1.2 提取节点和关系  使用大模型提示工程，从源文本块中提取包含详细属性的节点列表（涵盖实体、类型及属性）与关系列表（包括实体\_1、关系、实体\_2及属性）。通过精心设计提示词和优化提示词模板，提高知识抽取的准确性和效率。之后，用大模型为每个实体节点和关系边生成文本键值对，每个索引键是一个词或简短的短语，用于实现高效的检索；对应的值是一个段落文本，总结了从类型、关系、属性中提取的相关信息，帮助后续文本生成。为了同时平衡效率和质量的要求，让大模型进行多次实体的提取，以防止之前的过程可能遗漏任何额外实体。让大模型评估是否已提取所有实体，如果大模型回应遗漏了实体，则通过继续提示“上次提取中遗漏了实体”来使大模型继续搜集这些遗漏的实体。  1.3 生成全局图形  合并不同数据块中的子图，通过整合具有相同名称与类型但属性略有差异的实体的属性信息，实现去重与合并。在合并过程中平衡有效性和信息丢失的可能性，限制合并的迭代次数以确保图形的整体准确性和避免过多的信息冗余。逐步去除重复项，最终生成一个统一的全局图形。  2. 三层检索机制  为了响应用户查询，研究者们设计了三种不同的检索策略，提出了三级检索机制，包括局部检索、全局检索和网络搜索。  细粒度检索：侧重于检索具体的实体或关系。当大模型判断查询为一个细粒度查询时，会提取查询关键字，在知识图谱中检索特定关系和节点的详细信息。  粗粒度检索：涉及更广泛的子图或全局图。当大模型判断查询为一个粗粒度查询时，会先根据查询检索到相关的子图，然后进一步搜集与这些子图中的实体及其链接的Top-K相关实体的所有信息。这种方式能够全面覆盖知识图谱中的所有相关知识，并为后续的生成过程提供充足的上下文支持。  网络搜索：当知识图谱本身缺乏回答查询所需要的信息时，会采用大规模网络检索机制来补充缺失的信息。这种方法不仅解决了知识缺乏的问题，还保证了信息的实时性和最新性。  通过上面三种检索机制，研究者们确保了检索到的内容与查询高度相关，还能充分利用知识图谱的结构化优势，避免上下文冗长和信息冗余的问题。  3. 实验评估  3.1 数据集  为了对我们的方法进行全面分析，我们采用了一个航空领域数据集和三个多跳数据集进行评估。航空领域数据集我们采用了航旅纵横的航旅杯比赛数据集，数据主要来源于航旅纵横内部问答社区的样本，原始数据包含问题以及可回答该问题的文本，其中的评测答案均为人工标注结果。多跳数据集分别是HotPotQA、2WikiMultiHopQA和MuSiQue。这些数据集通常需要 2-4 个跃点的推理来回答问题，其中的每个问题分别与10、10和20个文档相关联，这些文档都取自维基百科。  3.2 基线模型  Naïve RAG：现有RAG系统中的标准基线。它将原始文本分割成块，并使用文本嵌入将它们存储在向量数据库中。对于查询，基础RAG生成向量化表示，直接根据表示中的最高相似度检索文本块，确保了高效且直接的匹配。  RQ-RAG：这种方法利用大模型将输入查询分解为多个子查询。这些子查询旨在通过使用重写、分解和消歧等显式技术来提高搜索准确性。  HyDE：这种方法利用大模型根据输入查询生成一个假设文档。然后使用这个生成的文档来检索相关的文本块，这些文本块随后被用来形成最终答案。4. 融合大模型与知识图谱技术构建LAE-GPT系统  3.3 评估指标  精确匹配（EM）是一种评估问答系统性能的指标，它用来衡量模型预测的答案与真实答案（ground truth answers）完全一致的比例。如果模型的输出与真实答案完全相同，则EM得分为1；如果不相同，则EM得分为0。这种评估方式非常严格，因为它要求预测答案与标准答案在字符级别上完全匹配。例如，在SQuAD数据集中，EM就是主要的评估指标之一。  F1分数是精确率（Precision）和召回率（Recall）的调和平均值，它综合考虑了这两个指标。精确率是指模型预测为正样本中真正为正样本的比例，而召回率是指模型能够正确预测为正样本的比例。  4. 系统实现  基于前面的方法，研究者们构建了融合大模型和知识图谱的低空经济语料分析系统（LAE-GPT）。该系统旨在收集低空经济语料，促进用户的深入理解和低空经济产业发展。LAE-GPT系统可以实现以下几个功能：  低空经济语料自动化收集：系统通过API的方式自动化采集低空经济语料数据并进行数据清洗、数据处理以及数据存储，以确保低空经济语料库的充足与准确性、实时性。新增的低空经济文本可以在原有的知识图谱上无缝集成。在知识库更新方面，为了能高效适应数据的动态变化，能够逐步更新低空经济知识库，而无需完全重新处理整个外部数据库。这种知识库更新方法有两个优势：一是实现新数据的无缝集成。二是减少计算开销，提升了整体效率。  低空经济智能问答：基于构建的低空经济知识图谱，利用三层检索机制，实现低空经济领域的精准、专业问答。  生成低空经济分析报告：生成详尽专业的低空经济分析报告，并附上了参考的真实信息来源，实现了知识溯源功能，确保了报告的权威性和可信度。 |
| 预期成果 | 1. 通过实验验证大模型融合知识图谱的检索增强生成方法在航空领域数据集和多跳数据集的准确性。  2. 构建低空经济语料分析系统（LAE-GPT），验证其在低空经济领域问答任务的专业性与准确性。  3.生成一份低空经济分析报告。 |
| 研  究  计  划 | 24.06.01——24.08.01 调研大模型、检索增强生成和知识图谱构建方法的文献24.08.01—24.08.01—24.11.01 设计结合大模型的知识图谱方法和三层检索机制  24.11.01-25.2.28 开展结合结合知识图谱的检索增强生成算法在公共数据集上的表现评测  25.2.28-25.09.30实现低空经济语料的自动化收集与管理；构建低空经济知识图谱。  25.09.30——26.3.01 构建低空经济语料分析系统LAE-GPT，完成智能问答和报告生成任务。 |
| 可能存在的问题 | 1. 算法还比较简单，需要进一步完善。  2. 需要加上消融实验。  3.评估指标可能需要增加，以更全面地进行分析评价。 |