

AIMI 시스템의 Agentic RAG 적용을 통한 소재 연구 효율성 증대 방안 연구

A Study on Enhancing Materials Research Efficiency by Applying Agentic RAG to the AIMI System

Yong-eun Cho^{*1, 2}, Won-Yong Shin^{† 1, 2}, Jun-Chae Na², Sung-Il Yang², Young-Jin Yu^{1, 2},
Ju-Hye Lee², Min-Hee Lee^{1, 2}, and Chan Jung^{1, 2}, Seung-Jun Han^{2, 3}

¹School of Mathematics and Computing (CSE), Yonsei University, Seoul 03722, Republic of Korea

²KAILOS LAB Co. Ltd., Seoul 06349, Republic of Korea

³Department of Mechanical Engineering (ME), Kyung Hee University, Yongin 17104, Republic of Korea

Email: kevin@kailoslab.com, wy.shin@yonsei.ac.kr, david@kailoslab.com, simon.yang@kailoslab.com, lucy.yu@kailoslab.com, may.lee@kailoslab.com, chloe@kailoslab.com, chann.jung@kailoslab.com, hansj2k@khu.ac.kr

Abstract

본 논문은 AIMI 시스템의 정보 검색 및 활용 능력 향상을 위해, 사용자 질의 기반 자율 계획, 동적 도구 (내부 벡터 DB, MongoDB, 웹 검색) 선택, 성찰 기반 검증 기능을 갖춘 Agentic Retrieval-Augmented Generation (Agentic RAG) 아키텍처를 제안한다. 로컬 LLM 과 LangChain/LangGraph 프레임워크를 활용하여 데이터 보안을 강화하고, 재료 연구의 효율성 및 정확성 증대를 목표로 한다.

I. Introduction

대규모 언어 모델 (Large Language Model, LLM)과 기존 검색 증강 생성 (Retrieval-Augmented Generation, RAG) [1]은 환각 현상 및 정적 파이프라인으로 인한 복잡한 질의 처리의 한계를 지닌다. 본 연구는 이러한 한계를 극복하고자 자율적인 계획 수립, 동적 도구 사용, 성찰 기능을 통합한 Agentic RAG 아키텍처를 재료 과학 및 공학 연구 시스템인 AIMI 에 적용하는 방안을 제안한다. AIMI 시스템은 MongoDB 기반 실험 데이터 분석/예측 시스템으로, 제안하는 Agentic RAG [2]는 논문/특허 (벡터 DB), 실험 데이터 (MongoDB), 웹 정보 등 다양한 내부 및 외부 정보 소스를 활용하며, 모든 LLM 구성 요소를 로컬 환경에서 운영하여 민감한 연구 데이터의 보안을 확보하고 연구 효율성 증대를 목표로 한다.

II. Methodology

제안하는 AIMI 시스템의 Agentic RAG 아키텍처는 사용자 질의를 분석하여 최적의 실행 계획을 수립하고, 내부 벡터 DB (논문/특허), MongoDB (구조화된 실험 데이터, TAG 원리 적용 [3]), 외부 웹 검색 등 필요한 도구를 동적으로 선택 및 조율하는 중앙 제어 에이전트 (AgentPlan)를 핵심으로 한다. 이후 검색 결과의 품질과 관련성을 평가 (EvalRetrieval)하고, 평가/필터링된 정보를 바탕으로 답변 초안을 생성 (Generate)한 뒤, 최종적으로 생성된 답변의 사실성 및 근거를 성찰 기반으로 검증 (Validate)하여 신뢰도를 높인다. 전체적인 AIMI 시스템은 LLM 기반 애플리케이션 개발을 위한 LangChain 과 에이전트 워크플로우 구축을 위한 LangGraph 프레임워크를 기반으로 로컬 LLM 을 활용하여 모듈식으로 구현된다. 이를 통해 자율적인 계획 수립, 질의 유형에 따른 동적인 다중 데이터 소스 (내부 DB, 외부 웹)의 유기적 활용 및 통합, 그리고 성찰 기반 검증을 통한 답변의 정확성 및 신뢰성 향상과 연구 데이터 보안 강화 효과를 기대한다.

III. Conclusion

본 논문에서는 재료 과학 및 공학 연구 시스템인 AIMI 를 위한 Agentic RAG 아키텍처를 제안하였다. 제안된 시스템은 자율적인 계획, 동적 다중 도구 활용, 성찰 기반 검증 기능을 통해 복잡한 연구 질의에 효과적으로 대응하고, 로컬 LLM 운영으로 데이터 보안을 확보하는 구조를 지닌다. 이는 기존 RAG 방식의 한계를 넘어 재료 연구자들이 방대한 정보 속에서 필요한 지식을 정확하고 효율적으로 탐색하고 활용하는데 기여할 것으로 기대된다. 향후 연구에서는 제안된 아키텍처를 실제 구현하고, 다양한 시나리오 기반 평가를 통해 성능을 검증할 계획이다.

Acknowledgement

This research was supported by SMEs Technology Innovation Development Program through the Technology Innovation and Promotion Agency (TIPA), funded by Ministry of SMEs and Startups (RS-2024-00511332) and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2021-NR059723, No. RS-2023-00220762).

References

- [1] Y. Gao *et al.*, "Retrieval-augmented generation for large language models: A survey," arXiv preprint arXiv:2312.10997.
- [2] A. Singh *et al.*, "Agentic retrieval-augmented generation: A survey on agentic RAG," arXiv preprint arXiv:2501.09136.
- [3] A. Biswal *et al.*, "Text2SQL is not enough: Unifying AI and databases with TAG," arXiv preprint arXiv:2408.14717.