Leveraging Knowledge Graphs and Generative AI for Augmented Research Paper Retrieval

Rima Dessí, Erick Mendez Guzman, Alya Alshaami, Amna Alowais, Hamda Alhammadi, Nada Alzarooni, Weam N.A Jarbou and Zarak Khan

Higher Colleges of Technology (HCT) - University City, Sharjah, UAE

Abstract

Recent advancement of Large Language Models (LLMs) has significantly enhanced Natural Language Processing and led to improved human-machine interaction. Despite their excellent benefits, LLMs still face challenges, particularly in generating accurate and reliable answers. To overcome these limitations Retrieval-Augmented Generation (RAG) systems have been proposed, which incorporate external knowledge to improve the responses generated by LLMs. In this paper, we argue that integrating a scientific Knowledge Graph (KG) into RAG systems can further enhance the effectiveness of LLMs in academic research paper retrieval. We further discuss the benefits and limitations of such approaches.

1. Introduction

The advancement of Large Language Models (LLMs) has transformed the field of Natural Language Processing and fostered significant improvements in capabilities such as human-machine interaction [1, 2]. These models such as the GPT series are based on transformer architecture and trained on extensive and diverse datasets. Further, such models have illustrated advanced capabilities of understanding human language and providing coherent answers. However, they face critical limitations, including (1) the tendency to produce inaccurate or *hallucination* answers, and (2) lack of domain-specific knowledge. These challenges arise primarily because their knowledge capabilities rely on static training data, meaning, they have limited access to up-to-date information after training. Further, the training data often lack very domain-specific information such as scientific research, technology, finance, and humanities. , etc. which limits the model's effectiveness in such areas.

To overcome the aforementioned challenges, several Retrieval-Augmented Generation (RAG) approaches have been proposed to enhance the performance of Large Language Models (LLMs) by supplying relevant external knowledge. The RAG methods combine a retrieval component with a generative model to allow LLMs to access more relevant and up-to-date data and provide more knowledge to LLMs beyond their training data. Upon a query submission, the retrieval model identifies and provides the most relevant document which then LLMs exploit this external knowledge to generate a higher degree of accuracy in responses. In the case of research papers, this process is particularly beneficial by enabling models to identify studies, methodologies, and experimental results to the given query and provide more reliable and precise scientific results.

Although these approaches achieved impressive performance by integrating external relevant knowledge, they fail to capture the semantic and structural relations among the documents provided as external knowledge to LLMs. To address these limitations Knowledge Graphs (KGs) are integrated in RAG approaches. KGs such as Wikidata[1] provide an excellent source of machine-understandable structured knowledge, enabling RAG systems to retrieve documents that are more semantically relevant and interconnected.

Overall, this paper argues that Scientific KGs such as CS-KG [3] can significantly improve the quality and factual accuracy of content generated by Large Language Models (LLMs) through their effective

---

✉ rdessi@hct.ac.ae (R. Dessí); eguzman@hct.ac.ae (E. M. Guzman)

[1]https://www.wikidata.org/wiki/Wikidata:Main_Page

integration into Retrieval-Augmented Generation (RAG) systems, particularly in the context of retrieving research papers and specialized information.

2. Towards Integration of KG in RAG Systems for Research Paper Retrieval

Integration of KGs into RAG systems provides significant benefits to LLMs generative capabilities. Since KGs serve factual and structured knowledge in the form of entities and their relations enable RAG systems to retrieve much relevant data effectively. To illustrate this integration, we propose the development of a Knowledge Graph-Enhanced Academic Research Publication Retrieval System, which later can easily be utilized by LLMs.

The primary goal is designing and implementing a prototype tool that integrates a Scientific Knowledge Graph with RAG to retrieve relevant research publications. This system will utilize structured knowledge to generate search results that are more contextually accurate and relevant to researcher queries, specifically targeting abstracts and titles.

Additionally, the system will improve search relevance through the integration of the KG and LLMs. By leveraging the knowledge graph, the system will enhance search relevance by connecting research papers through semantic relationships, such as shared topics, data, and methodologies. This approach will enable the generation of more meaningful recommendations that extend beyond traditional keyword-based searches.

Finally, we aim to lay the foundation for future knowledge graph-based research tools. This includes documenting the technical challenges and solutions encountered during the project, thereby creating a roadmap for future research and development of knowledge graph-based scholarly tools. The documentation will identify additional functionalities that could further support researchers, such as personalized recommendations and trend analysis.

3. Challenges and Limitations

The integration of Scientific Knowledge Graphs (KGs) into Retrieval-Augmented Generation (RAG) systems can provide numerous benefits across different domains. KG-enhanced RAG can significantly improve especially in the context of academic research by enabling researchers to find relevant information more efficiently and with greater accuracy. However, there are challenges and limitations need to be considered while integrating KG into RAG system for the more enhanced research academic retrieval. The most notable ones include:

Data Quality and Completeness: The effectiveness of RAG system relies heavily on the KG completeness. The number of published research papers and findings are increasing everyday. Therefore, it is curicial to the KG should be regularly updated.

Complexity of Integration: Integration of KGs into RAG systems proposes significant technical challenges such as different data formats which needs to be addressed.

Scalability Issues: As mentioned before, the number of available research papers and findings are increasing everyday. To provide up-to-date results the KG should be continuously augmented with them. This poses several scalability issues due to the volume of the data.

4. Conclusion and Future Directions

In this paper, we argue that KGs provide substantial benefits when integrated into RAG systems, particularly in the field of academic research. Scientific KGs such as can generate search results that are more contextually accurate and relevant to researcher queries, specifically targeting abstracts and titles. While the challenges remain such as scalability and complexity, scientific KGs are excellent resources to be integrated into the RAG system to enable LLMs to generate a higher degree of accuracy in responses.

5. Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-4 and Grammarly for Grammar and spelling checks. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] B. Peng, Y. Zhu, Y. Liu, X. Bo, H. Shi, C. Hong, Y. Zhang, S. Tang, Graph retrieval-augmented generation: A survey, CoRR abs/2408.08921 (2024). URL: https://doi.org/10.48550/arXiv.2408.08921. doi:10.48550/ARXIV.2408.08921. arXiv:2408.08921.

[2] L. Bahr, C. Wehner, J. Wewerka, J. Bittencourt, U. Schmid, R. Daub, Knowledge graph enhanced retrieval-augmented generation for failure mode and effects analysis, CoRR abs/2406.18114 (2024). URL: https://doi.org/10.48550/arXiv.2406.18114. doi:10.48550/ARXIV.2406.18114. arXiv:2406.18114.

[3] D. Dessì, F. Osborne, D. R. Recupero, D. Buscaldi, E. Motta, CS-KG: A large-scale knowledge graph of research entities and claims in computer science, in: The Semantic Web - ISWC 2022 - 21st International Semantic Web Conference, Virtual Event, October 23-27, 2022, Proceedings, 2022.