StructRAG: Structure-Aware RAG Framework with Scholarly Knowledge Graph for Diverse Question Answering

Runsong Jia*
University of Technology Sydney
Sydney, NSW, Australia
runsong.jia@student.uts.edu.au

Bowen Zhang*
QDX Technologies
Canberra, ACT, Australia
bowen.zhang@qdx.co

Sergio José Rodríguez Méndez
The Australian National University
Canberra, ACT, Australia
Sergio.RodriguezMendez@anu.edu.au

Pouya G. Omran
The Australian National University
Canberra, ACT, Australia
P.G.Omran@anu.edu.au

Abstract

Recent advances in Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) have shown promise in academic question answering. However, existing approaches often fail to fully utilize document structural information and lack diversity in retrieved contexts. This paper presents StructRAG, a structure-aware RAG framework that leverages scholarly knowledge graphs for enhanced question answering. Our framework features three key innovations: (1) an automated knowledge graph construction pipeline based on Deep Document Model (DDM) that preserves document hierarchical structure, (2) a structure-aware retrieval mechanism that combines semantic relevance with source diversity, and (3) a context-enhanced generation approach that integrates structural metadata for improved answer synthesis. Experimental results on 329 computer science papers demonstrate that StructRAG significantly outperforms vanilla RAG baseline. While maintaining comparable semantic accuracy (91% vs 90%), our approach achieves substantially higher diversity in generated answers (Distinct-1: 62% vs 52%, Distinct-2: 89% vs 78%) and better answer quality across all metrics, with notable improvements in relevance (29%) and readability (36.5%). These results demonstrate that StructRAG effectively enhances both the diversity and quality of academic question answering.

CCS Concepts

• **Information systems → Information retrieval**; • **Information systems → Question answering**; • **Information systems → Knowledge representation and reasoning**; • **Information systems → Information extraction**;

Keywords

Knowledge Graph, Large Language Models, Retrieval-Augmented Generation, Knowledge Graph Construction, Deep Document Model

WWW Companion '25, April 28-May 2, 2025, Sydney, NSW, Australia

ACM ISBN 979-8-4007-1331-6/2025/04
https://doi.org/10.1145/3701716.3717819

ACM Reference Format:
Runsong Jia, Bowen Zhang, Sergio José Rodríguez Méndez, and Pouya G. Omran. 2025. StructRAG: Structure-Aware RAG Framework with Scholarly Knowledge Graph for Diverse Question Answering. In *Companion Proceedings of the ACM Web Conference 2025 (WWW Companion '25), April 28-May 2, 2025, Sydney, NSW, Australia.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3701716.3717819

1 INTRODUCTION

With the rapid development of natural language processing and information retrieval technologies, building efficient and intelligent question-answering systems has become an important goal for both academia and industry. Recent years have witnessed remarkable advances in LLMs. Models like GPT-4 [11] have demonstrated impressive capabilities in natural language understanding and generation tasks. To enhance the reliability and factuality of LLM outputs, RAG has emerged as a promising approach by incorporating relevant information from external knowledge sources into the generation process [7]. The RAG framework has seen significant evolution, from the initial knowledge retrieval-generation pipeline to more sophisticated variants like Self-RAG [16] that implement self-reflection mechanisms, and GPT-RAG [8] that optimizes retrieval strategies through reinforcement learning.

However, applying RAG to scientific literature question-answering presents unique challenges that existing approaches have not fully addressed [15]. First, academic papers contain rich structural information, including hierarchical relationships between sections and complex elements like figures and formulas, which traditional RAG systems primarily focused on textual similarity fail to utilize effectively. Second, while recent work has improved RAG's retrieval precision through techniques like query reformation and iterative refinement [21], the diversity of retrieved context remains underexplored, particularly in academic settings where questions often require multi-faceted analysis drawing from various sources and perspectives.

To address these challenges, this paper proposes StructRAG: a structure-aware RAG framework that leverages academic knowledge graphs to enhance both the accuracy and diversity of question answering. Our framework makes three key contributions:

(1) We develop an automated knowledge graph construction pipeline that preserves the hierarchical document structure

while capturing rich semantic relationships between academic entities.
(2) We introduce a structure-aware retrieval mechanism that utilizes both document hierarchy and semantic relevance to identify diverse yet coherent information sources.
(3) We design a path-enhanced generation approach that guides LLMs to produce comprehensive answers by incorporating structural context from knowledge graphs.

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 details the overall framework and key technologies of StructRAG; Section 4 presents the experimental setup and results analysis; finally, Section 5 concludes the paper and discusses future work.

2 RELATED WORK

Recent advances in LLMs have significantly impacted academic knowledge processing and question answering. This section reviews key developments relevant to our research.

2.1 Knowledge Graph Construction for Academic Literature

Large-scale academic knowledge graphs have demonstrated substantial value in structuring scholarly information [10, 18]. Recent developments in knowledge graph construction have focused on capturing fine-grained document structure and automating the extraction process [6, 22]. Notable progress has been made in developing hierarchical document models that preserve structural relationships while enabling automated knowledge extraction. Open infrastructures for academic metadata have also emerged, providing comprehensive coverage of research outputs and their relationships [12].

2.2 LLM-Enhanced Knowledge Processing

The integration of LLMs with knowledge graphs has demonstrated significant improvements in academic information processing [4, 14]. Recent frameworks have focused on enhancing factual consistency and reducing hallucination in academic question answering [1]. Advanced techniques for scientific literature understanding have been developed [20], along with specialized methods for semantic query processing in scholarly knowledge graphs [5].

2.3 Diversity in Information Retrieval

While traditional information retrieval systems prioritize relevance metrics, recent research has emphasized the importance of result diversity. Several approaches have been proposed for optimizing both relevance and source variety in RAG systems [9]. New diversity metrics specifically designed for academic search have emerged [23], along with techniques for balancing semantic coverage with information redundancy [19]. These developments are particularly relevant for academic contexts, where comprehensive understanding often requires multiple perspectives.

Despite these advances, significant challenges remain in academic question answering systems. Current approaches face difficulties in balancing semantic relevance with source diversity, while the integration of structural information from knowledge graphs with LLM-based generation requires further development. This work addresses these gaps through enhanced diversity-aware retrieval mechanisms and improved context integration strategies, leveraging both document hierarchy and semantic relationships to enhance retrieval quality and diversity.

3 METHODOLOGY

In this section, we present our approach that combines automated knowledge graph construction with a diversity-aware retrieval mechanism for enhanced question answering. As shown in Figure 1, our methodology consists of three main components: automated knowledge extraction, diversity-based retrieval, and context-aware answer generation.

3.1 Automated Knowledge Graph Construction

We have developed an automated knowledge graph construction system that processes and transforms academic literature through a multi-stage pipeline. In the first stage, the system uses a marker component to convert academic papers from PDF format into structured Markdown documents. This conversion process not only extracts text content but also generates unique identifiers for each document through blake2b hash algorithms, ensuring traceability of the processing. The system implements an incremental processing mechanism by maintaining a mapping between original filenames and hash values, effectively avoiding redundant conversion of previously processed documents. Specifically, the system stores file mapping information in JSON format, ensuring cross-platform compatibility and processing transparency.

Based on the principles of DDM[1] [6], we implemented a granularity-based hierarchical document parsing solution that structures documents at three levels: sections, paragraphs, and sentences. This hierarchical document parsing approach is implemented through document structure ontology, ensuring the structural integrity of original documents during knowledge extraction. During the parsing process, the system identifies the hierarchical structure of documents and assigns unique identifiers to each hierarchical element, providing important contextual information for subsequent subgraph search, knowledge extraction, and semantic analysis.

In the knowledge graph construction phase, the system employs RDF and RDFS, organizing and storing extracted knowledge through OWL2 ontology-based schemas. The knowledge graph structure incorporates both domain-specific ontology concepts and instance data, while also maintaining document model descriptions to capture document component structures, content representations, and attribute relationships. Leveraging LangChain[3] for robust entity extraction and relationship identification, the system identifies important entities through a comprehensive taxonomy of academic concept types. Our entity type system is designed to capture the multifaceted nature of academic literature, particularly in computer science research. The taxonomy encompasses key research participants such as researchers and institutions, enabling the tracking of scholarly contributions and organizational collaborations. It also captures fundamental academic concepts including algorithms, methods, and frameworks, which represent the core technical contributions in computer science literature. To document

[1] https://w3id.org/kgcp/DDM

concrete research outputs, the system recognizes various research artifacts such as papers, systems, and software tools. Additionally, the taxonomy includes data-related entities like datasets and databases, which are crucial for experimental validation, as well as evaluation metrics that quantify research performance. This carefully structured entity type system enables systematic knowledge extraction while maintaining clear categorical boundaries.

Finally, the system converts the extracted structured knowledge into standard TTL (Turtle) format. In this process, the system generates standardized URIs for each document element, establishes semantic associations between entities, while preserving necessary document metadata. Through the RDF graph structure, the system establishes a complete knowledge network, including hierarchical and semantic relationships between entities. This approach not only achieves structured knowledge representation but also maintains bidirectional associations with original documents, providing a reliable foundation for subsequent knowledge retrieval and verification[2].

During the knowledge graph construction process, the system employs a series of optimization measures to improve processing efficiency. First, parallel processing technology enhances the performance of batch document conversion. Second, the system implements memory management mechanisms, avoiding memory overflow issues through batch processing of large documents. Finally, the system adopts a batch writing strategy when generating knowledge graphs, reducing the number of I/O operations and improving overall performance. These optimization measures enable the system to efficiently process large-scale academic literature collections, providing effective technical support for academic knowledge management and utilization.

Compared to previous work, we have integrated and enhanced the PARSE [22] and DDM [6] models, enabling automated construction of academic knowledge graphs from research papers in PDF and other formats with a single-step process. The implementation is publicly available at https://w3id.org/kgcp/DPFusion.

3.2 Diversity-based Retrieval Process

The retrieval process comprises three sequential stages designed to maximize both relevance and source diversity:

Initial Entity Recognition and Matching. The process begins with GPT-4 identifying key entities from the user query. These entities are then matched against the knowledge graph to locate relevant nodes and their associated paragraphs. This step ensures that retrieved content is semantically aligned with the query's focus.

Semantic-based Retrieval. We compute semantic similarity scores between the query entities and paragraphs in our knowledge graph by using cosine similarity of their embeddings, leveraging Sentence-BERT [13]. This step can be formalized as:

$$P = \{p_1, \ldots, p_n\} \tag{1}$$

is the set of all paragraphs in the knowledge graph,

$$sim(q, p) = \cos(\text{embed}(q), \text{embed}(p)), \tag{2}$$

where q is the query, embed() represents the embedding function, and $sim(q, p)$ is the semantic similarity score.

Source Diversity Reranking. From the top-15 paragraphs with the highest similarity scores, we implement a source-aware reranking strategy that maximizes information source diversity while maintaining semantic relevance. This can be formalized as:

$$P = \{p_1, \ldots, p_{15}\} \tag{3}$$

the top-15 retrieved paragraphs,

$$P^* \subseteq P \text{ such that } |P^*| = k \text{ (where } k = 5) \tag{4}$$

and $|\{S(p)|p \in P^*\}|$ is maximized,

where $S(p)$ denotes the source document of paragraph p. This approach ensures that the selected paragraphs represent the broadest possible range of source documents while maintaining high relevance to the query.

3.3 Context-Aware Answer Generation

The final component employs the LLaMA3 70B model [17] for answer generation, with specific attention to context utilization and structural information. Our approach implements strict context boundaries through carefully designed prompts that constrain the model to generate answers solely based on the provided paragraphs.

To enhance answer accuracy, we augment the context with structural metadata derived from our DDM representation. This includes each paragraph's location within its source document's hierarchy (e.g., section position, subsection relationships) and its source document information. This structural context helps the model better understand the relative importance and relationships between different pieces of information.

The model generates a single coherent paragraph that synthesizes information from the diverse set of source paragraphs. This approach ensures that the final answer maintains high factual accuracy while incorporating multiple perspectives from different academic sources. Also, the generation process incorporates DDM structural metadata to enhance context understanding:

```
Paper-[PaperID]-Section-[SectionID]-
Paragraph-[ParagraphID]
```

This structural context enables LLM to better comprehend the relationships and relative importance of different information pieces within the academic document hierarchy.

4 RESULTS

4.1 Experimental Setup

We conducted our experiments on a dataset comprising 329 academic papers from the computer science domain. To establish a comprehensive evaluation framework, we first randomly selected 50 paragraphs from this dataset. For each selected paragraph, we employed LLM (GPT-4) to generate a scientific question based on its content, ensuring these questions were representative of complex academic queries. These paragraph-question pairs served as our evaluation benchmark.

For each generated question, we employed GPT-4 [11] for entity recognition, obtaining an entity list that captured the key concepts and terms within the query. These identified entities were then used

[2]Example implementations and TTL files demonstrating this structure are available at the project's repo https://w3id.org/kgcp/DPFusion in the folder `code/PARSE/Papers/Pipeline/output`.

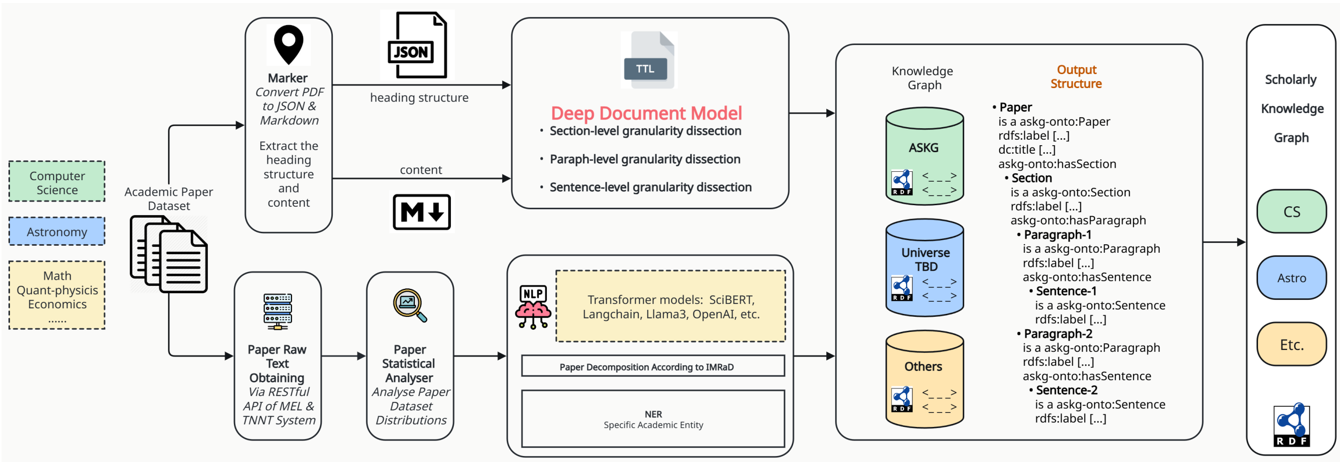

Figure 1: System architecture for academic paper processing and knowledge graph construction.

to match against all paragraphs in our dataset. Our system initially retrieved the top-15 most semantically similar paragraphs based on these entities, from which we selected the top-5 paragraphs that maximized source diversity while maintaining relevance to serve as the context for answer generation. Then we utilize LLaMA3 70B to generate answers based solely on the selected paragraphs.

For the baseline, we implemented a VanillaRAG approach that splits documents into chunks of 1000 tokens, selecting the top-5 most semantically similar chunks as context. Our StructRAG method first retrieves the top-15 most relevant paragraphs based on semantic similarity, then selects the top-5 paragraphs that maximize source diversity while maintaining relevance.

We evaluated our system using two sets of metrics. First, we assessed answer quality using four key dimensions: Relevance (how well the answer addresses the query), Accuracy (factual correctness), Completeness (coverage of necessary information), and Readability (clarity and coherence). These metrics were scored on a 5-point scale by Claude 3 [2], an advanced language model.

While our implementation uses GPT-4 for entity recognition and LLaMA3 70B for answer generation due to their strong performance in academic text understanding and structured generation tasks, our framework maintains a plug-and-play architecture that can accommodate different language models. The choice of these specific models demonstrates the framework's capability, but the structure-aware approach and diversity-based retrieval mechanism are model-agnostic and can be integrated with any suitable LLMs based on specific requirements and resource constraints.

Furthermore, we measured the diversity and semantic quality of the generated answers using three metrics: Distinct-1 (unigram diversity), Distinct-2 (bigram diversity), and Semantic Similarity (comparing generated answers with standard answers).

4.2 Performance Analysis

Table 1 presents the comparison between StructRAG and the baseline method across the four quality metrics. StructRAG demonstrates substantial improvements across all dimensions, with particularly notable gains in Relevance (29% improvement) and Readability (36.5% improvement). The significant enhancement in Completeness (from 3.22 to 4.12) suggests that our diversity-aware retrieval strategy effectively captures a broader range of relevant information.

The diversity metrics, shown in Table 2, further validate the effectiveness of our approach. StructRAG achieves higher scores in both Distinct-1 (62% vs 52%) and Distinct-2 (89% vs 78%), indicating more diverse vocabulary usage in the generated answers. Notably, this increased diversity does not come at the cost of semantic accuracy, as evidenced by the slightly higher Semantic Similarity score (91% vs 90%).

The experimental results demonstrate that StructRAG's source-diverse retrieval strategy effectively enhances both the quality and diversity of generated answers. The improved performance can be attributed to two key factors: (1) the fine-grained knowledge representation provided by our enhanced DDM framework, and (2) the effective balance between semantic relevance and source diversity in our retrieval mechanism.

The consistent improvement across all metrics suggests that our approach successfully addresses the challenge of maintaining answer quality while increasing information diversity. This is particularly evident in the simultaneous improvement of both diversity metrics and semantic similarity, indicating that StructRAG effectively combines information from multiple sources while maintaining coherence and accuracy.

4.3 Knowledge Graph Analysis

To comprehensively evaluate the quality and coverage of the constructed knowledge graph, we conducted multi-dimensional quantitative analysis of entity distribution. Through detailed analysis of entity distribution patterns, we discovered several important characteristics and patterns. From the overall hierarchical structure

Table 1: Comparison of Methods Based on Relevance, Accuracy, Completeness, and Readability

Method	Relevance	Accuracy	Completeness	Readability
VanillaRAG	3.58	4.12	3.22	3.40
StructRAG	**4.62**	**4.42**	**4.12**	**4.64**

Table 2: Comparison of Methods Based on Distinctness and Semantic Similarity

Method	Distinct-1(%)	Distinct-2(%)	SemanticSimilarity (%)
VanillaRAG	52	78	90
StructRAG	**62**	**89**	**91**

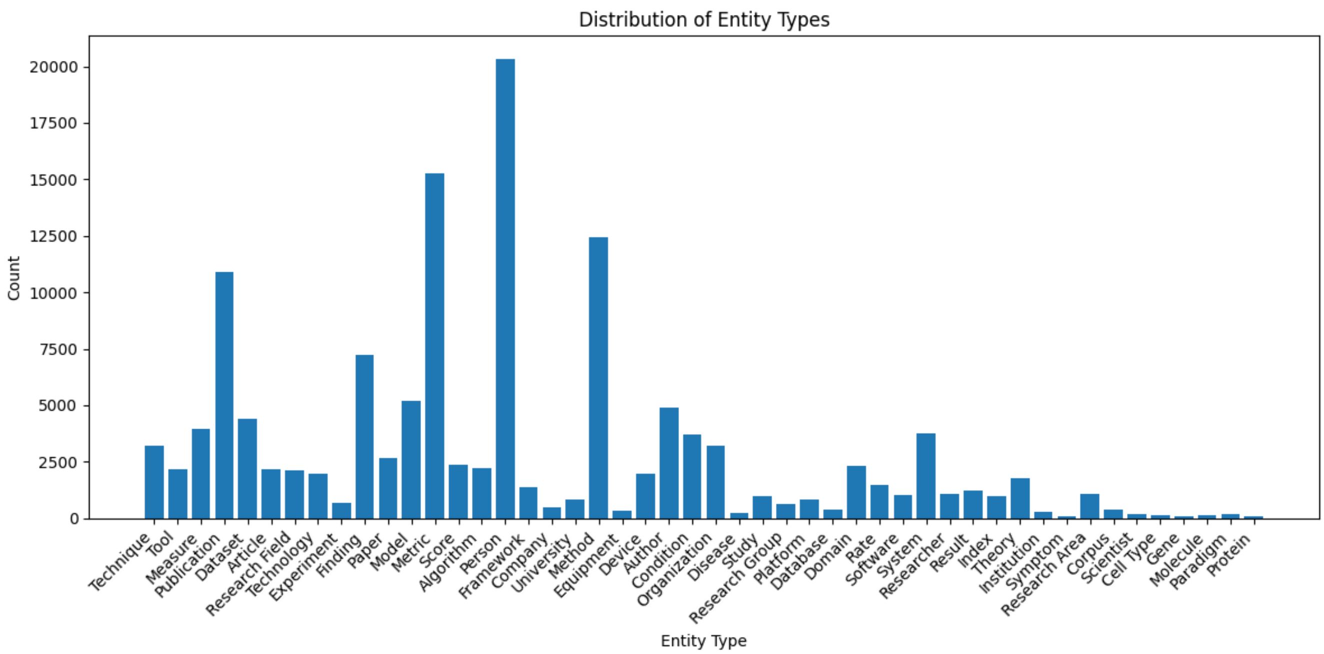

Figure 2: Distribution of entity types in the knowledge graph.

perspective, the entity distribution exhibits clear pyramid characteristics. As shown in Figure 2, among all entities, Person (20,331, 18.9%), Metric (15,242, 14.2%), Method (12,431, 11.6%), and Publication (10,904, 10.1%) constitute the highest level, accounting for 54.8% of the total. The dominant position of Person entities indicates that our knowledge graph effectively captures the personnel networks in academic research, while the high proportion of Metric and Method entities reflects the in-depth characterization of research methodology. Finding (7,226, 6.7%), Model (5,197, 4.8%), Author (4,885, 4.5%), and Dataset (4,389, 4.1%) entities form the middle layer, accounting for approximately 25%, these entities mainly represent research outcomes and methodological components. From a methodological perspective, we found that methodology-related entities (Method, Algorithm, Technique) account for 15.7% of the total, while evaluation-related entities (Metric, Measure, Score) account for 20.1%, this 4:5 ratio reflects the academic community's emphasis on rigorous evaluation. Publication-related entities (Publication, Paper, Article) account for 14.7% of the total, comparable to the proportion of methodology-related entities, indicating a good balance between research methods and their formal publication. This distribution characteristic highly aligns with the natural language processing field's emphasis on methodological innovation and experimental validation. Regarding research organizational structure, the ratio between individual entities (Person, Author, Researcher) and institutional entities (Organization, University, Research Group) is approximately 6:1, demonstrating the knowledge graph's ability to maintain institutional context while providing fine-grained expression of individual academic contributions. This hierarchical entity distribution structure not only reflects our knowledge graph's ability to effectively express both the breadth and depth of academic research but also indicates its structural maturity. This characteristic enables the knowledge graph to effectively support diverse academic application scenarios, from macro-domain analysis to specific research queries.

To better understand the structural characteristics of entity distribution, we employed document clustering analysis based on entity type distribution. Using t-SNE dimensionality reduction and K-means clustering algorithms, we divided the document collection into three main clusters, as shown in figure 3. Cluster 0 (blue dots, approximately 15%) primarily aggregates papers focused on algorithms and theoretical methods, typically containing higher proportions of Algorithm and Method entities. Cluster 1 (orange dots, approximately 45%) is characterized by datasets and experimental evaluation, showing higher proportions of Dataset, Evaluation, and Metric entities. Cluster 2 (green dots, approximately 40%) displays a more balanced and diverse entity distribution, suggesting these documents may cover multiple research aspects or interdisciplinary work.

Through heatmap analysis, we further revealed the importance distribution of entity types across different clusters. The results indicate that while each cluster shows preferences for certain entity types, the boundaries between them are not absolute. This "soft boundary" characteristic reflects the interdisciplinary and comprehensive nature of modern computer science research. Notably, we observed that Method and Algorithm entities maintain relatively high importance across all clusters, confirming their fundamental status in computer science research.

During knowledge retrieval, the DDM [6] enables a more nuanced understanding of document structure and entity relationships. The system can leverage these distribution characteristics, enhanced by DDM's hierarchical document representation, to balance different types of entity information, thereby improving the diversity and completeness of retrieval results. Furthermore, these distribution characteristics, combined with DDM's ability to capture document structure from section to sentence level, help us better understand the organizational structure of disciplinary knowledge.

5 CONCLUSION AND FUTURE WORK

In this paper, we presented an enhanced approach to academic question answering that integrates source diversity into the retrieval process. By improving the PARSE pipeline[3] [22] and implementing a source-aware retrieval strategy, we demonstrated that considering source diversity in context selection can enhance both answer quality and information comprehensiveness. Our experimental results showed improvements across multiple evaluation metrics, including relevance, accuracy, completeness, and readability, while maintaining high semantic similarity with standard answers.

Several directions for future research emerge from this work. First, we plan to enhance our evaluation framework through:

(1) Incorporating a larger dataset of academic papers and more diverse scientific questions to validate our approach across different domains
(2) Introducing human expert evaluation to complement the LLM-based metrics for answer quality assessment
(3) Conducting comprehensive ablation studies to analyze the contribution of each pipeline component
(4) Establishing systematic comparisons between LLM-based and human evaluations to better understand potential biases

Second, we aim to explore additional dimensions of diversity in our retrieval strategy:

- **Author Diversity**: Incorporating author information could provide varied perspectives and methodologies. This would involve analyzing author networks and ensuring representation of different research groups and academic generations. The integration of author diversity could help capture different schools of thought and methodological approaches within the field.
- **Institutional Diversity**: Considering the affiliations of authors could enrich the retrieved context by including perspectives from different:
  - Geographic locations and research traditions
  - Types of institutions (universities, research institutes, industry labs)
  - Research groups with varying methodological approaches
- **Temporal Diversity**: Balancing between recent developments and foundational works by:
  - Tracking the evolution of research ideas over time
  - Ensuring representation of both contemporary and seminal works
  - Capturing the progression of methodologies and findings

Finally, we plan to develop more sophisticated evaluation metrics specifically designed for measuring these additional dimensions of diversity. This would involve creating new quantitative measures for assessing the balance between different diversity aspects while maintaining answer quality. Such metrics would help in understanding the impact of various diversity factors on the overall performance of academic question answering systems.

[3] https://w3id.org/kgcp/PARSE

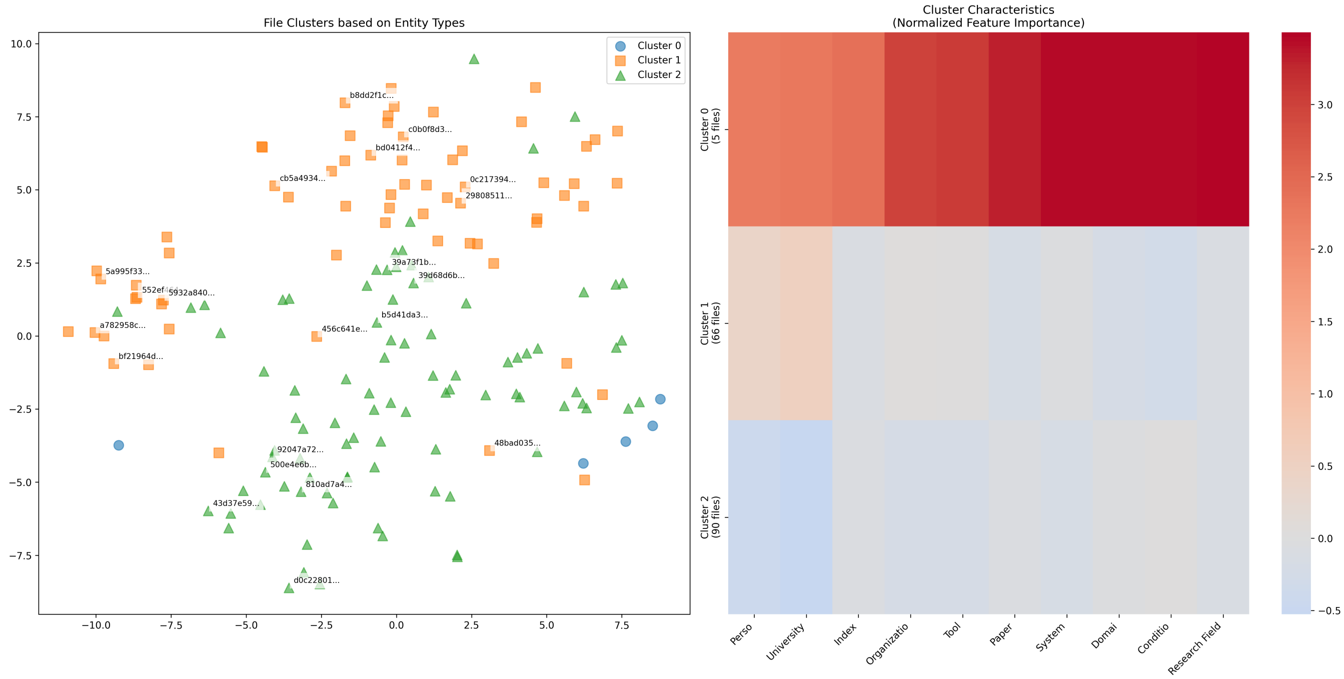

Figure 3: Document clustering analysis based on entity types.

References

[1] Geetika Agrawal and Huan Liu. 2024. Can Knowledge Graphs Reduce Hallucinations in LLMs? A Systematic Study. *Findings of ACL 2024* (2024).

[2] Anthropic. 2023. Claude: Introducing Claude. https://www.anthropic.com/index/introducing-claude.

[3] Baskaryan and othres. 2024. LangChain: Building context-aware reasoning applications. https://github.com/langchain-ai/langchain. https://github.com/langchain-ai/langchain

[4] Ming Chen, Jing Liu, and Yue Wang. 2024. Knowledge-RAG: A Knowledge-Aware Framework for RAG in Scientific Domain. *arXiv preprint arXiv:2402.15568* (2024).

[5] Runsong Jia, Bowen Zhang, Sergio José Rodríguez Méndez, and Pouya G. Omran. 2024. Leveraging Large Language Models for Semantic Query Processing in a Scholarly Knowledge Graph. *arXiv preprint arXiv:2405.15374* (2024).

[6] Runsong Jia, Bowen Zhang, Sergio José Rodríguez Méndez, and Pouya G. Omran. 2024. Towards Building an RDF-based Deep Document Model and Retrieval Augmented Generation System for Enhanced Question Answering with Large Language Models. In *Proceedings of the Posters, Demos, and Industry Tracks at the International Semantic Web Conference 2024*. Baltimore, USA.

[7] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* 33 (2020), 9459–9474.

[8] Zhihong Liu, Shuai Wang, Chao Zhang, and Yang Yang. 2024. GPT-RAG: Generative Pre-trained Transformer for Retrieval-Augmented Generation. *arXiv preprint arXiv:2401.08845* (2024).

[9] Zhengyang Liu, Xuezhi Wang, and Yiran Chen. 2024. DiverseRAG: Enhancing Retrieval-Augmented Generation through Multi-Source Integration and Diversity Optimization. *arXiv preprint arXiv:2401.08046* (2024).

[10] Kyle Lo, Lucy Lu Wang, Mark Plank, and Sebastian Uribe. 2023. Semantic Scholar at Scale: Building Infrastructure for AI-Powered Scientific Discovery. *Nature Communications* 14, 1 (2023), 1842.

[11] OpenAI. 2023. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774* (2023).

[12] Jason Priem, Heather Piwowar, and Richard Orr. 2024. OpenAlex: The Fully Open Catalog of Global Scholarship. *Quantitative Science Studies* 3, 1 (2024), 1–20.

[13] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. 3982–3992.

[14] Daniel Sanmartin, Yifan Zhang, and Wei Chen. 2024. KG-RAG: Knowledge Graph Enhanced Retrieval-Augmented Generation. *arXiv preprint arXiv:2405.12035* (2024).

[15] Yunfan Shi, Yixin Pan, Kai Sun, and Dong Yu. 2023. A Survey on Retrieval-Augmented Generation for Large Language Models. *arXiv preprint arXiv:2312.10997* (2023).

[16] Akari Sun, Wangchunshu Yu, Tom Griffiths, and Jacob Steinhardt. 2023. Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection. *arXiv preprint arXiv:2310.11511* (2023).

[17] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and Efficient Foundation Language Models. arXiv:2302.13971 [cs.CL]

[18] Kuansan Wang, Anshul Subramanian, and Omri Ben-Eliezer. 2024. Microsoft Academic Graph 2024: Advancing Discovery at Web Scale. *Scientometrics* 129, 1 (2024), 1–23.

[19] Xiaojun Wang, Chen Li, and Mei Zhang. 2024. Balanced Retrieval for Scientific Question Answering. *arXiv preprint arXiv:2402.18229* (2024).

[20] Zhirong Wu, Liu Yang, and Hao Chen. 2024. Scientific Literature Understanding with Structure-Aware LLMs. *arXiv preprint arXiv:2403.09132* (2024).

[21] Zhangyang Yu, Yue Wang, Haotian Wang, Wenpeng Wang, and Xu Yin. 2023. Progressive Retrieval-Augmented Generation. *arXiv preprint arXiv:2310.03275* (2023).

[22] Bowen Zhang, Sergio José Rodríguez Méndez, and Pouya Ghiasnezhad Omran. 2023. ASKG: An Approach to Enrich Scholarly Knowledge Graphs through Paper Decomposition with Deep Learning.. In *ISWC (Posters/Demos/Industry)*.

[23] Wei Zhang, Yan Liu, and John Wang. 2024. Diversity-Aware Academic Search: Metrics and Methods. *Information Retrieval Journal* 27 (2024), 121–145.