Journal of Computer Science and Electrical Engineering
Print ISSN: 2663-1938
Online ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3073

A RETRIEVAL-AUGMENTED GENERATION (RAG)-BASED INTELLIGENT REVIEWER ASSIGNMENT SYSTEM FOR SCIENTIFIC PROJECT EVALUATION

JiTao Ma*, HongWei Huang, Jun Du
Network Management Center, Yunnan Science and Technology Information Research Institute, Kunming 650051, Yunnan, China.
Corresponding Author: JiTao Ma, Email: 34047697@qq.com

Abstract: With the rapid growth of scientific research projects and increasing complexity in interdisciplinary collaboration, traditional expert assignment methods—such as manual screening and keyword matching—are becoming inadequate. This study proposes an intelligent reviewer assignment system based on Retrieval-Augmented Generation (RAG), which enhances semantic understanding and improves the accuracy of matching scientific projects with suitable experts. The system constructs detailed knowledge profiles for both projects and experts across four dimensions: research questions, methods, results, and conclusions. Domain-specific prompts guide large language models (LLMs) to extract structured knowledge from unstructured textual inputs. These profiles are then transformed into semantic vectors using BERT-based embeddings and matched using cosine similarity. Experimental results show that the proposed method significantly outperforms baseline approaches in terms of precision, recall, and F1-score. Specifically, the model achieves 79% precision, 75% recall, and 77% F1-score at Top-5 recommendations. This work contributes to the development of more intelligent, accurate, and scalable systems for scientific peer review and expert assignment.
Keywords: Intelligent reviewer assignment; Retrieval-Augmented Generation (RAG); Semantic profiling; Expert matching

1 INTRODUCTION

1.1 Research Background

The rapid expansion of scientific research activities, particularly in cross-disciplinary domains, has significantly increased the complexity and volume of project evaluation tasks. According to recent statistics from the National Science Foundation, cross-disciplinary projects now constitute over 45% of total funding allocations, underscoring the urgent need for more intelligent and efficient mechanisms to assign expert reviewers. In this context, accurate reviewer assignment plays a dual role: (1) as a quality assurance mechanism that ensures methodological rigor and innovation in proposed research; and (2) as a knowledge alignment system that connects cutting-edge scientific inquiries with domain-specific expertise [1].

The importance of this process is further emphasized by its direct influence on critical decision-making outcomes such as funding allocation, institutional credibility, and the long-term trajectory of scientific discovery. Despite its significance, the traditional methods used for reviewer assignment face substantial limitations [2].

Manual screening remains the most widely adopted approach, with over 85% of funding agencies relying on administrative staff to match projects with experts based on keyword searches or personal networks. Although this method allows for some degree of contextual judgment, it suffers from low accuracy—particularly in interdisciplinary settings—where studies have shown low assignment success rates [3]. Additionally, manual processes are time-consuming, with large-scale grant calls often requiring more than three weeks for complete reviewer assignment [4].

Algorithmic approaches using TF-IDF, co-authorship graphs, or basic machine learning models offer partial automation but struggle with dynamic fields where terminology evolves rapidly. Commercial systems like Elsevier's Reviewer Recommender provide automated solutions but are limited to publication-based matching, neglecting key aspects such as methodological alignment or practical experience.

To address these systemic challenges, we propose a Retrieval-Augmented Generation (RAG)-based intelligent reviewer assignment system, which integrates semantic understanding, structured knowledge extraction, and explainable decision-making. The system employs a three-phase workflow:

- Knowledge Extraction: Utilizes domain-specific prompts to guide large language models (e.g., Qwen-72B) in extracting structured knowledge from unstructured project descriptions and expert publications. Knowledge is captured across four dimensions: research questions, methods, results, and conclusions.
- Semantic Profiling: Constructs semantic profiles for both projects and experts based on the extracted knowledge. These profiles are represented as vector embeddings, enabling deeper semantic understanding.
- Semantic Matching: Computes similarity scores between project and expert profiles using the cosine similarity. The system generates ranked recommendations based on the similarity scores.

This study makes two primary contributions to the field of intelligent reviewer assignment:

- Architectural Innovation: We introduce the first RAG-based framework specifically designed for expert-project matching in academic peer review. This architecture combines prompt-driven knowledge profiling with neural retrieval and generation techniques.
- Empirical Validation: Through real world experiments involving, our model achieves 78% precision, greatly outperforming SVM-based baselines, demonstrating significant improvements in matching accuracy and robustness.

The paper is structured as follows: Section 2 reviews intelligent assignment systems and RAG applications. Section 3 presents the technical architecture. Section 4 validates performance against benchmarks, with conclusions in Section 5.

2 THEORETICAL FOUNDATION AND LITERATURE REVIEW

2.1 Intelligent Reviewer Assignment Systems

The development of automated reviewer assignment systems has progressed through distinct methodological phases, each addressing critical limitations in matching scholarly expertise to evaluation tasks. Early systems relied primarily on lexical matching algorithms that demonstrated limited effectiveness, with poor precision for interdisciplinary matching scenarios [5]. These initial approaches were constrained by their inability to recognize semantic relationships between conceptually similar but lexically distinct terms, with studies indicating they failed to identify most of equivalent term pairs [6]. The introduction of optimization algorithms, including the Hungarian method and linear programming techniques, brought mathematical rigor to the assignment process but introduced new challenges in computational complexity and dynamic constraint management. Subsequent machine learning approaches marked a significant advancement, with supervised learning models incorporating citation network analysis and publication timelines demonstrating improved performance. However, these systems still exhibited notable limitations, including substantial retraining latency and poor handling of early-career researchers' sparse publication records [7]. The current generation of neural systems has achieved transformative improvements through dynamic embedding techniques and cross-modal matching capabilities [8]. Despite these advances, challenges remain in maintaining real-time knowledge updates and ensuring transparent decision-making processes.

2.2 Retrieval-Augmented Generation in Academic Contexts

Retrieval-Augmented Generation (RAG) architectures have emerged as a powerful paradigm for knowledge-intensive academic tasks since their formalization by Lewis et al.[9]. These systems combine neural retrieval components with conditional generation capabilities, addressing fundamental limitations in traditional language models. The retrieval phase typically employs FAISS-optimized maximum inner product search, which has demonstrated good performance across corpora exceeding 2 million documents [10]. This is complemented by dynamic re-ranking mechanisms that perform better than cross-encoder BERT models. The generation component leverages advanced language models like Qwen-72B, which is better at factual checking compared to conventional methods while maintaining robust performance across specialized domains [11]. In practical applications, RAG systems have shown particular promise in grant review matching, where multi-modal analysis of proposal content has achieved measurable improvements in assignment quality. Journal reviewer suggestion systems incorporating live citation network data demonstrate 89% precision in matching, though they face challenges related to temporal lags in knowledge base updates. Conference paper assignment systems benefit from cross-institutional profile alignment, realizing 57% improvements in processing speed. However, significant research gaps persist, particularly in maintaining knowledge freshness and bridging interdisciplinary domains. These limitations highlight the need for continued innovation in developing more adaptive and transparent matching systems for scholarly applications.

3 METHODOLOGY

3.1 Overview of the Proposed Framework

This study proposes a Retrieval-Augmented Generation (RAG)-based intelligent reviewer assignment system to improve the accuracy and efficiency of matching scientific projects with appropriate experts. The proposed methodology consists of three main stages: (1) knowledge extraction using domain-specific prompts, (2) semantic modeling of both project and expert profiles, and (3) semantic matching based on multi-dimensional similarity. The system leverages large language models (LLMs) for structured knowledge extraction and integrates vector-based semantic representations to enable precise expert-project alignment.

3.2 Prompt Design and Knowledge Extraction

Prompt engineering plays a central role in transforming unstructured scientific project descriptions into structured knowledge representations that can be effectively used for semantic matching with expert profiles. In this study, we employ carefully crafted prompts to guide large language models (LLMs) in extracting standardized information from project proposals across four key dimensions: research questions, methods, results, and conclusions. Each prompt is

designed to ensure consistency, completeness, and relevance of the extracted content, enabling accurate semantic modeling and subsequent expert matching.

- Research Questions: The identification of research questions forms the foundation of any scientific project, as it defines the core problem being addressed. To extract this critical information, we design prompts that encourage LLMs to not only identify the main question but also recognize its sub-questions, logical dependencies, and connections to existing literature. An example prompt for this dimension is: Identify the central research questions in the given proposal. List all sub-questions or hypotheses and explain how they relate to one another and to the broader research context.
- Research Methods: Accurately capturing the methodologies employed in a research project is crucial for identifying reviewers who possess the relevant technical expertise. Our approach involves using domain-specific prompts to extract detailed methodological information, including whether the study is experimental, theoretical, or data-driven. A representative prompt for this dimension is: Classify the research methodology used in the project as experimental, theoretical, or data-driven and describe it in detail.
- Research Results: Extracting results enables the system to assess the empirical impact of the research and match it with experts who have published comparable findings or worked on related phenomena. The goal is to capture both quantitative outcomes and qualitative interpretations. An example result-focused prompt is: Extract the main findings of the study. Describe the implications of the results for the field and any limitations in their generalizability.
- Research Conclusions: Finally, the conclusions provide insight into the broader significance of the research and its potential contributions to the field. We use prompts to extract not only the stated conclusions but also the inferred impacts on future research and applications. An illustrative conclusion prompt is: Summarize the major conclusions drawn from the research. Discuss how these findings advance the field or inform policy, practice, or future studies.

3.3 Semantic Profiling of Projects and Experts

Semantic profiling transforms unstructured textual inputs into structured, vector-based representations that enable accurate similarity comparisons between projects and experts.

3.3.1 Project profiling

For each research project, the four-dimensional knowledge extracted via prompts is encoded into a semantic vector using BERT-based embeddings. Each dimension—research questions, methods, results, and conclusions —is represented as a sub-vector. These are then combined into a composite profile that reflects the overall semantic structure of the project. This representation allows for precise semantic comparison with expert profiles.

3.3.2 Expert profiling

Expert profiles are constructed by applying the same set of prompts to each expert's representative publications from the past five years. The extracted knowledge from individual papers is aggregated to form an expert-level profile across the same four dimensions:

- Research Questions: Common themes and problems addressed in the expert's work.
- Methods: Frequently used techniques and methodologies.
- Results: Key findings and empirical contributions.
- Conclusions: Overall impact and theoretical or practical implications.

Each dimension is also vectorized and integrated into a comprehensive expert profile. This approach ensures that expert profiles reflect both historical expertise and current research focus.

3.4 Semantic Matching Between Projects and Experts

Once both project and expert profiles are constructed, semantic matching is performed to identify the most suitable reviewers.

We compute similarity scores using cosine similarity between the semantic vectors of projects and experts across each of the four dimensions. The overall match score is defined as a weighted sum:

$$\boldsymbol{SimScore(P,E)=\alpha\cdot cos(Q_P,Q_E)+\beta\cdot cos(M_P,M_E)+\gamma\cdot cos(R_P,R_E)+\delta\cdot cos(C_P,C_E)} \quad (1)$$

where $\alpha+\beta+\gamma+\delta=1$, and each term corresponds to similarity in research questions, methods, results, and conclusions respectively.

Experts are ranked based on their match scores, and the top-N candidates are selected for assignment, subject to conflict-of-interest filtering.

4 EXPERIMENT DESIGN

4.1 Data Description

To evaluate the performance of the proposed intelligent reviewer assignment system, we constructed an experimental dataset based on data collected from the ScholarMate platform (a Chinese academic social network). We randomly selected 100 scholars from the field of Information Systems. For each scholar, we collected their most recent 10 publications, forming a total dataset of 1,000 papers.

Among these, for each scholar, 9 of the 10 papers were used as the training set to build the expert profile, resulting in a total of 900 papers for training. The remaining one paper per scholar was used as the test set, comprising 100 test papers. The goal of the experiment was to use the proposed method to find the top-N most relevant experts for each test paper. The original author of the test paper was considered the ground truth for correct expert assignment.
This setup allowed us to simulate a realistic expert assignment scenario, where the system must identify the appropriate reviewers based on the content of the paper, without prior knowledge of the author's identity.

4.2 Baseline Methods

We compared our proposed method with three baseline approaches:
Random Assignment: Experts are randomly selected without considering any semantic or topical information.
Keyword Matching: A traditional approach that matches papers and experts based on shared keywords extracted using TF-IDF.
SVM-based Assignment: A machine learning approach using Support Vector Machines trained on manually labeled project-expert pairs to predict relevance.
These baselines represent different levels of sophistication in the expert assignment process, ranging from purely random selection to supervised learning models.

4.3 Metrics

To quantitatively assess the performance of the proposed method and the baselines, we employed three widely used evaluation metrics: precision, recall, and F1-score. These metrics are defined as follows:
Precision: measures the proportion of assigned experts who are correct, which is calculated as:

$$\text{Precision}=\frac{\text{TP}}{TP+FP}$$

Recall: measures the proportion of correct experts that were successfully identified, which is calculated as:

$$\text{Recall}=\frac{\text{TP}}{TP+FN}$$

F1 – score: provides a balanced measure of precision and recall, which is calculated as:

$$\text{F1-Score}=\frac{2*\text{Precision}*\text{Recall}}{Precision+Recall}$$

4.4 Results Analysis

The experimental results are summarized in Table 1, which compares the average precision, recall, and F1-score of the four methods at Top-5 recommendations.

Table 1 Performance of Top-5 recommendations

Method	Precision	Recall	F1 - score
Random Assignment	0.18	0.15	0.16
Keyword Matching	0.37	0.32	0.34
SVM-based Assignment	0.46	0.41	0.43
Proposed Model	0.79	0.75	0.77

As shown in the table, the proposed model significantly outperforms all baseline methods across all three metrics. Specifically, the proposed model achieves a precision of 0.79, indicating that nearly 80% of the recommended experts are correct. It also obtains a recall of 0.75, meaning that 75% of the correct experts are successfully identified among the Top-5 recommendations. The F1-score of 0.77 further confirms its superior balance between precision and recall.
The poor performance of the random assignment method highlights the necessity of using semantic or topic-based matching strategies. While keyword matching improves upon random assignment, its limited ability to capture semantic relationships restricts its effectiveness. The SVM-based method performs better due to its use of supervised learning; however, it still falls short of the proposed model, which benefits from structured knowledge profiling and semantic similarity computation. Additionally, the use of prompt-driven knowledge extraction ensures that both projects and experts are represented in a consistent and comprehensive manner, leading to more accurate matches.

5 CONCLUSION

This study presents a novel intelligent reviewer assignment system based on Retrieval-Augmented Generation (RAG), aiming to enhance the accuracy and efficiency of matching scientific research projects with appropriate expert reviewers. Traditional methods, such as keyword-based or optimization-based approaches, often fail to capture the semantic complexity of both project content and expert expertise. To address these limitations, we propose a structured knowledge profiling framework that leverages domain-specific prompts and large language models (LLMs) to extract multi-dimensional knowledge from project proposals and expert publications.

Specifically, we design four types of prompts to guide the extraction of structured knowledge across four key dimensions: research questions, methods, results, and conclusions. These prompts ensure consistent and comprehensive knowledge representation for both projects and experts. Based on this structured knowledge, we construct semantic profiles using BERT-based embeddings, enabling fine-grained similarity comparisons. Finally, we implement a semantic matching algorithm that computes relevance scores between projects and experts across all four dimensions, resulting in more accurate and context-aware reviewer assignments.

The experimental results demonstrate the effectiveness of our approach. Using a dataset collected from the ScholarMate platform consisting of 100 scholars and their recent 10 papers each, we constructed a test scenario where one paper per scholar was used for evaluation. Our model achieved 0.79 precision, 0.75 recall, and 0.77 F1-score at Top-5 recommendations, significantly outperforming baseline methods including random assignment, keyword matching, and SVM-based assignment. This indicates that the proposed method can better understand the conceptual structure of research and align it with relevant expert domains.

One of the key innovations of this work lies in the integration of prompt-driven structured knowledge extraction with semantic vectorization and multi-dimensional matching. Unlike previous systems that rely heavily on surface-level features or manual feature engineering, our approach enables automated, fine-grained, and semantically rich profiling of both projects and experts. Furthermore, by applying the same set of prompts consistently across both data sources, we ensure comparability and coherence in knowledge representation, which enhances the overall performance of the assignment process.

Despite its promising results, the proposed system has several limitations. First, the current dataset is limited to the Information Systems field, which may affect the generalizability of the model to other disciplines, especially those with different writing styles or publication practices. Second, while our method captures expert knowledge based on recent publications, it does not account for real-time changes in an expert's interests or availability. Third, the system assumes that the original author of a paper is the most suitable reviewer, which may not always be the case in practice due to potential conflicts of interest or workload constraints.

Future research will focus on addressing these limitations and further improving the system's applicability and robustness. First, we plan to expand the dataset to include multiple academic fields and investigate cross-domain transfer capabilities. Second, we aim to incorporate additional information such as expert preferences, availability, and past review quality into the assignment model to make it more practical for real-world applications. Third, we will explore dynamic updating mechanisms to ensure that expert profiles remain current as their research evolves. Finally, integrating explainability features into the model will help users understand the rationale behind reviewer assignments, thereby increasing trust and transparency in the peer review process.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Sedaghat A R, Bernal-Sprekelsen M, Fokkens W J, et al. How to be a good reviewer: A step-by-step guide for approaching peer review of a scientific manuscript. Laryngoscope Investigative Otolaryngology, 2024, 9(3): e1266.

[2] Aksoy M, Yanik S, Amasyali M F. Reviewer assignment problem: A systematic review of the literature. Journal of Artificial Intelligence Research, 2023, 76: 761-827.

[3] Bornhorst J, Rokke D, Day P, et al. B-122 Estimated improvement of sigma error metrics associated with manual secondary result review, and subsequent artificial intelligence driven quality assurance. Clinical Chemistry, 2023, 69(Supplement_1): hvad097-456.

[4] Horbach S S, Halffman W W. The changing forms and expectations of peer review. Research Integrity and Peer Review, 2018, 3: 1–15.

[5] Liang D, Xu P, Shakeri S, et al. Embedding-based zero-shot retrieval through query generation. arXiv preprint arXiv:2009.10270, 2020.

[6] Al Hashimy A S H, Kulathuramaiyer N. An automated learner for extracting new ontology relations. In: Proceedings of the 2012 International Conference on Advanced Computer Science Applications and Technologies (ACSAT), IEEE, 2012: 19-24.

[7] Jiménez-Ruiz E, Cuenca Grau B. LogMap: Logic-based and scalable ontology matching. In: Proceedings of the International Semantic Web Conference. Berlin, Heidelberg: Springer, 2011: 273–288.

[8] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019: 4171–4186.

[9] Lewis P, Perez E, Piktus A, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv preprint arXiv:2005.11401, 2020.

[10] Johnson J, Douze M, Jégou H. Billion-scale similarity search with GPUs. IEEE Transactions on Big Data, 2021, 7(3): 535 – 547.

[11] Wang H, Zhang D, Li J, et al. Entropy-optimized dynamic text segmentation and RAG-enhanced LLMs for construction engineering knowledge base. Applied Sciences, 2025, 15(6): 3134.