D. V. Pakutnik

RETRIEVAL-AUGMENTED GENERATION FOR AUTOMATED QUALITY CONTROL OF STUDENT ASSIGNMENTS

This paper formalizes the RAG-based automated grading pipeline, defines evaluation metrics for retrieval relevance and rubric alignment, and demonstrates that graph-structured RAG outperforms flat vector retrieval on complex assessment tasks.

Introduction

Large language models show strong potential for automated grading but remain prone to hallucinations and rubric misalignment. Retrieval-Augmented Generation addresses these limitations by grounding generation in externally retrieved evidence [1]. This work formalizes the RAG-based grading pipeline and examines the progression from flat vector retrieval to graph-structured knowledge representations.

I. Formal Framework

Let q denote the student response, r the reference answer, and $\mathcal{K} = \{d_1, \ldots, d_N\}$ a knowledge base of rubric criteria and exemplar responses. The RAG pipeline comprises three stages.

Retrieval. A retriever $\mathcal{R}$ selects the top-K relevant documents:

$$\mathcal{C}_K = \arg\max \sum_{d \in S} \mathrm{sim}(\mathbf{e}_q, \mathbf{e}_d) \tag{1}$$

where $\mathbf{e}_q, \mathbf{e}_d$ are dense vector embeddings and $\mathrm{sim}(\cdot,\cdot)$ is cosine similarity.

Generation. The LLM generator produces score s and feedback f:

$$y = (s, f) = \mathcal{G}(q, r, \mathcal{C}_K, \mathcal{P}) \tag{2}$$

where $\mathcal{P}$ encodes rubric instructions.

Evaluation. System quality is assessed via three metrics. Context relevance:

$$\mathrm{CR}(q, \mathcal{C}_K) = \frac{1}{K} \sum_{d \in \mathcal{C}_K} \mathbb{1}[\mathrm{rel}(d, q)] \tag{3}$$

Answer faithfulness:

$$\mathrm{AF}(y, \mathcal{C}_K) = \frac{\left|\{c \in \mathrm{claims}(y) : \exists\, d \in \mathcal{C}_K,\ d \models c\}\right|}{\left|\mathrm{claims}(y)\right|} \tag{4}$$

II. From Flat Retrieval to GraphRAG

Flat Vector RAG. Rubric-structured retrieval extracts information aligned with rubric knowledge points rather than ranking documents by similarity. Evaluation on a 701-essay dataset confirmed 94–99 % agreement with human evaluators, validating RAG-based grading for large-scale assessment.[2]

Graph-Structured RAG. Standard flat retrieval treats knowledge fragments in isolation, failing to capture multi-hop reasoning chains essential for complex assessment. To address this limitation, the GraphRAG framework was proposed, organizing rubrics and reference materials into a knowledge graph $G = (V, E)$. [3] Retrieval is performed via subgraph traversal:

$$\mathcal{C}_G = \bigcup_{v \in V_q} \mathcal{N}_h(v) \tag{5}$$

where V_q are query-matched nodes and $\mathcal{N}_h(v)$ is the h-hop neighborhood. This converts grading from open-ended generation into constrained graph verification. Experimental results confirm that GraphRAG significantly outperforms flat RAG.

III. Conclusion

The analysis reveals a clear progression from flat vector RAG, which grounds decisions in isolated rubric fragments, to graph-structured RAG capturing relational dependencies between concepts. Future directions include adaptive graph construction, integration with multi-agent architectures, and multimodal RAG for code and diagram evaluation.

References

1. RAG for Knowledge-Intensive NLP Tasks [website]. – URL: https://arxiv.org/abs/2005.11401 (date of access: 24.03.2026).
2. SteLLA: A Structured Grading System Using LLMs with RAG [website]. – URL: https://arxiv.org/abs/2501.09092 (date of access: 25.03.2026).
3. From Flat to Structural: Enhancing ASAG with GraphRAG [website]. – URL: https://arxiv.org/abs/2603.19276 (date of access: 25.03.2026).

Pakutnik Dmitry, Master's student of the Department of Intelligent Information Technologies of BSUIR, d.pakutnik@bsuir.by.

Scientific supervisor: Zakharov Vladimir, Associate Professor of the Department of Intelligent Information Technologies, PhD in Technology, zakharov@bsuir.by.