A RAG Evaluation Detailed Results

This appendix provides detailed results from the comparative evaluation of Knowledge Graph-based Retrieval-Augmented Generation (KG-RAG) versus text-based RAG systems. The evaluation was conducted using 15 test questions designed to assess various aspects of question answering performance.

A.1 System Configuration and Performance Overview

KG-RAG System (KGRAGSystem) The knowledge graph-based RAG system demonstrated the following performance characteristics:

- Questions processed: 15
- Average latency: 6.715 seconds
- Answer Quality Metrics:
  - Overall Quality: 0.85
  - Relevance: 0.88
  - Completeness: 0.78
  - Factual Accuracy: 0.90

Text-RAG System (TextOnlyRAGSystem) The traditional text-based RAG system showed the following performance:

- Questions processed: 15
- Average latency: 6.161 seconds
- Answer Quality Metrics:
  - Overall Quality: 0.76
  - Relevance: 0.80
  - Completeness: 0.68
  - Factual Accuracy: 0.82

A.2 Comparative Performance Analysis

Response Time Comparison Table 1 shows the latency comparison between the two systems.

Table 1. Response Time Comparison

System	Average Latency (s)
KG-RAG	6.715
Text-RAG	6.161

Answer Quality Metrics Table 2 presents a comprehensive comparison of answer quality across multiple dimensions.

Table 2. Answer Quality Comparison

System	Overall Quality	Relevance	Completeness	Factual Accuracy
KG-RAG	0.85	0.88	0.78	0.90
Text-RAG	0.76	0.80	0.68	0.82
Improvement	**+11.8%**	**+10.0%**	**+14.7%**	**+9.8%**

Table 3. Answer Semantic Similarity

System Comparison	Average Similarity
KG-RAG vs Text-RAG	0.59

Answer Semantic Similarity The semantic similarity between answers from different systems was measured to assess consistency. Table 3 shows the results.

A.3 Representative Question-Answer Examples

Example 1: Knowledge Graph Construction Entities Question: What are the key entities involved in knowledge graph construction for scientific literature?

KG-RAG Answer: The key entities involved in knowledge graph construction for scientific literature include concepts such as "Knowledge Graph," "entities," and the concept that links between entities in a knowledge base can express specific relations. The graph highlights these as central concepts, indicating their importance in the construction process. The structured knowledge representation allows for precise identification of key concepts and their potential roles, though effectiveness depends on the completeness and connectivity of the graph.

Text-RAG Answer: Since no relevant documents were found, the answer was based on general knowledge. Key entities include: (1) Entities as concepts, objects, and ideas represented as nodes; (2) Relationships as interactions and dependencies represented as edges; (3) Literature as scientific papers providing data; (4) Metadata about literature for contextualization.

Example 2: Extraction Method Precision Question: Which extraction methods achieve the highest precision in scientific knowledge graph construction?

KG-RAG Answer: Based on the provided knowledge graph information, it was not possible to directly answer the question about extraction methods and precision metrics. The knowledge graph lacked specific entities related to extraction methods and precision metrics, highlighting a limitation in the current graph's coverage for this specific domain.

Text-RAG Answer: No document extracts were available to answer the question. General knowledge suggests common methods include rule-based extraction (high precision with well-defined rules), machine learning-based extraction (high precision with comprehensive training data), and hybrid approaches combining both methods.

Example 3: Common Entity Types Question: What entity types are most common in scientific knowledge graphs?

KG-RAG Answer: The knowledge graph included "Entity Type" as a concept and mentioned "uneven distribution of entity and relation frequencies in knowledge graphs," suggesting certain entity types are more prevalent. However, the graph lacked specific examples of common entity types in scientific domains, limiting the comprehensiveness of the answer.

Text-RAG Answer: No relevant information was available from document extracts. The answer could not be provided due to the absence of specific documentation about entity types in scientific knowledge graphs.

A.4 Key Findings and Observations

Advantages of KG-RAG

1. **Higher Answer Quality:** KG-RAG consistently outperformed Text-RAG across all quality dimensions, with improvements ranging from 9.8% to 14.7%.
2. **Better Factual Accuracy:** The structured nature of knowledge graphs enabled more accurate factual responses (0.90 vs 0.82).
3. **Enhanced Completeness:** KG-RAG provided more complete answers (0.78 vs 0.68), leveraging structured relationships.
4. **Improved Relevance:** The system demonstrated better relevance (0.88 vs 0.80) by utilizing graph structure for context.

Limitations and Challenges

1. **Slightly Higher Latency:** KG-RAG showed 9.0% higher response time (6.715s vs 6.161s) due to graph traversal overhead.
2. **Coverage Dependencies:** Performance heavily depends on knowledge graph completeness and connectivity.
3. **Domain-Specific Gaps:** Limited performance when specific domain knowledge is not well-represented in the graph.

Retrieval Performance Analysis The significant difference in retrieval metrics (KG-RAG: F1=0.41 vs Text-RAG: F1=0.00) indicates that the knowledge graph-based approach was more effective at retrieving relevant information, though both systems showed room for improvement in recall performance.

A.5 Implications for Knowledge Graph Design

The evaluation results highlight several important considerations for knowledge graph construction:

1. **Completeness is Critical:** The effectiveness of KG-RAG is directly related to the comprehensiveness of the underlying knowledge graph.

2. **Relationship Quality Matters:** Well-defined relationships between entities significantly improve answer quality and relevance.
3. **Domain Coverage:** Ensuring adequate coverage of domain-specific concepts and relationships is essential for optimal performance.
4. **Balance Trade-offs:** While KG-RAG provides higher quality answers, the slight increase in response time should be considered for real-time applications.

A.6 Conclusion

This detailed evaluation demonstrates that knowledge graph-based RAG systems offer significant advantages over traditional text-based approaches, particularly in terms of answer quality, factual accuracy, and completeness. However, the success of KG-RAG systems is contingent upon the quality and coverage of the underlying knowledge graph. The results support the value proposition of structured knowledge representation for enhanced question answering capabilities, while also highlighting areas for future improvement in graph construction and retrieval optimization.