Application and Effectiveness Analysis of Graph RAG Technology in Intelligent Writing Content Optimization (Postprint)

Authors: Qin Jun

Date: 2025-09-19T15:52:10+00:00

Abstract

[Objective] Current large language models often treat intelligent writing as unstructured plain text data, neglecting the logical structure and connections within the text, which constrains content quality and effectiveness. [Method] Precisely and rapidly locating relevant information is key to improving content quality. Through an intelligent writing optimization technique that integrates RAG and Knowledge Graph (KG), we construct a knowledge graph to integrate information and preserve the internal logic and relationships of the text, thereby enhancing content writing effectiveness. [Results] This method can more precisely parse writing requirements, and by introducing the knowledge graph, it improves information retrieval accuracy and optimizes content generation quality. [Conclusion] This technology has been piloted for nearly half a year at the AIGC Application Research Center (Guangxi Laboratory) of the China Federation of Press Technology Workers, demonstrating that this method can significantly enhance the accuracy and efficiency of content generation.

Full Text

Application and Effectiveness Analysis of Graph RAG Technology in Intelligent Writing Content Optimization

Guangxi Daily Press, Nanning, Guangxi 530000

Abstract: Current large language models often treat intelligent writing as unstructured plain text data, neglecting the internal logical structure and connections within the text, which restricts content quality and effectiveness. Precisely and rapidly locating relevant information is key to improving content quality. Through an intelligent writing optimization technology that integrates RAG and knowledge graphs (KG), we construct knowledge graphs to integrate information, preserve the internal logic and relevance of text, thereby enhancing content writing effectiveness. This method can more accurately parse writing requirements, and by introducing knowledge graphs, improves information retrieval precision and optimizes content generation quality. This technology has been piloted for nearly half a year at the AIGC Application Research Center (Guangxi Laboratory) of the China News Technology Workers Federation, demonstrating that the method can significantly improve the precision and efficiency of content generation.

Keywords: Graph RAG; intelligent writing; content optimization; knowledge graph; retrieval enhancement
CLC Number: G202
Document Code: A
Article ID: 1671-0134(2025)08-132-05
DOI: 10.19483/j.cnki.11-4653/n.2025.08.029
Citation Format: Qin Jun. Application and Effectiveness Analysis of Graph RAG Technology in Intelligent Writing Content Optimization[J]. China Media Technology, 2025, 32(8): 132-136.

With the continuous advancement of artificial intelligence technology, intelligent writing assistance services have become increasingly critical for improving work efficiency and customer satisfaction. However, relying solely on large language models (LLMs) often yields limited results, as they cannot adequately encode and organize domain-specific knowledge. To overcome this bottleneck, the AIGC Application Research Center (Guangxi Laboratory) of the China News Technology Workers Federation was established to explore retrieval-augmented generation (RAG) technology, particularly Graph RAG that combines knowledge graphs and graph databases. GraphRAG effectively enhances the ability to capture logical structure and relevance in text by constructing a global entity-relationship network, compensating for the context loss caused by segmentation in traditional RAG when processing long texts [1]. In query-focused summarization tasks, it significantly improves the efficiency and precision of key information retrieval, thereby increasing the recall rate of critical information [2]. Graph RAG simultaneously constructs a multi-dimensional textual information structure network to optimize content generation for intelligent writing, enhancing retrieval performance and question-answering capabilities in customer service. It not only overcomes the precision loss and quality degradation issues caused by ignoring document structure and segmentation in traditional methods but also provides a novel methodology to unleash the potential of RAG and enhance the accuracy and relevance of LLMs when answering complex questions.

Guangxi Laboratory is jointly established by Guangxi Cloud Digital Media Group and Xinhua Zhiyun, accredited by the China News Technology Workers Federation.

1. Technical Overview

Early intelligent writing technologies primarily relied on simple rules and templates for text generation, later evolving into natural language processing techniques based on statistical models such as Hidden Markov Models. These approaches learned language structures through big data and statistical algorithms, improving text naturalness and finding widespread application in domain-specific intelligent content writing [3]. Notably, the MGC (Machine Generated Content) production model, a result of innovative collaboration between Guangxi Cloud Digital Media Group and Xinhua Zhiyun, serves as an excellent example. It utilizes intelligent production robots to achieve efficient content generation in areas such as weather, events, and exhibitions, significantly improving production efficiency during coverage of typhoons "Hongxia" and "Yuhong" in 2020. However, intelligent writing still faces challenges in understanding and creating complex content.

In recent years, the emergence of large language models (LLMs) has further propelled the development of intelligent writing. Nevertheless, this technology still faces numerous shortcomings and challenges, including but not limited to accuracy issues caused by data bias, deficiencies in logical reasoning and common sense capabilities, difficulties in maintaining long-text contextual coherence, and hallucination problems, all of which hinder the widespread application of intelligent writing.

RAG (Retrieval-Augmented Generation) is a natural language query technology that enhances LLM performance by retrieving relevant information from external knowledge sources. It can significantly improve the relevance of question answering because it combines the generative capabilities of LLMs with the accuracy of external knowledge. RAG introduces reliable contextual information, effectively reducing the tendency of LLMs to hallucinate. Consequently, RAG has become the preferred method for enhancing generative model output and is widely applied in information extraction, text summarization, intelligent writing, and other processing tasks [4]. Although RAG has demonstrated good performance in various scenarios, it still has limitations when facing complex situations. First, it is constrained by difficulties in effectively associating key points when integrating different information to provide comprehensive insights. Second, low-quality user queries, such as those containing excessive abbreviations, affect model understanding. Furthermore, in complex text writing, RAG shows insufficient accuracy in entity representation, such as inadequate expression of logical relationships involving time, location, and persons, resulting in suboptimal performance in complex text writing tasks [5].

2. Methods and Effectiveness Analysis

The Graph RAG intelligent writing system workflow (shown in [Figure 1: see original paper]) consists of two main stages: First, in the knowledge graph construction stage, information is extracted from past customer service records to build detailed knowledge graphs, which are interconnected based on relational contexts. Additionally, the system generates embedding vectors for each node to facilitate subsequent semantic search. Second, in the knowledge retrieval and question-answering stage, the system parses user queries, identifies named entities and intents, and finally locates relevant subgraphs in the knowledge graph to generate answers, thereby providing precise and efficient information feedback for the intelligent writing system.

2.1 Knowledge Graph Construction

Aligned with Guangxi Cloud' s positioning as a media and technology service provider, we collected and organized knowledge related to Guangxi' s culture and media as our foundation. Utilizing extraction technologies with structured understanding capabilities and combining natural language processing (NLP), we transformed massive amounts of knowledge texts into structured entities, relationships, and attributes. The construction process includes entity extraction, attribute mining, and relationship identification. Subsequently, we optimized the graph using technologies such as graph deep learning (GNNs) to provide rich semantic features for retrieval tasks. Finally, we continuously evaluated and optimized through a user feedback mechanism.

[Figure 3: see original paper] Knowledge Graph Construction Process

(1) Corpus Preparation. To create a media culture knowledge graph with Guangxi characteristics, we constructed a unique corpus based on publicly available internet data related to Guangxi from sources such as Guangxi News Network, Nanguo Morning Post, and Baidu Baike, as well as characteristic private data including Guangxi Daily' s digital newspaper and image library. This covers multiple domains including politics, law, judiciary, economy, infrastructure, agriculture and rural areas, energy and water resources, transportation, environmental meteorology, and culture and leisure, providing a broad, abundant, and accurate data foundation for intelligent writing.

(2) Entity Recognition. For explicit information such as person names, place names, and organization names in the corpus, we utilized an entity recognition algorithm framework for label extraction. First, we extracted lexical, syntactic, and contextual semantic information from the text through a Roberta+Bi-LSTM model, where Roberta, as a large-scale text pretraining model, already contains some prior knowledge [7]. Finally, we used Conditional Random Fields (CRF) to constrain and extract labels.

We have completed approximately 300,000 general knowledge graph entities for persons, organizations, and geography, and approximately 50,000 basic entities for news region graphs, totaling 350,000 entity label entries.

[Figure 1: see original paper] RAG Processing Flow

Graph RAG technology effectively overcomes existing language model challenges in intelligent writing regarding entity accuracy, contextual logical relationships, and long-text expression [6]. Compared with traditional augmented generation techniques, Graph RAG enhances text logic and relevance by integrating knowledge graphs, substantially improving generation precision and efficiency. By constructing customized knowledge graphs for writing tasks, Graph RAG can rapidly locate entities and their relationships associated with key information, providing rich contextual information for writing. Graph RAG' s key technologies include entity recognition and relationship construction, intent detection, and embedding-based subgraph retrieval. Leveraging advanced natural language processing algorithms, Graph RAG can accurately identify entities and relationships in queries and deeply parse user query intent. Additionally, embedding-based subgraph retrieval technology enables efficient and precise subgraph retrieval, further enhancing text generation logic and coherence.

[Figure 2: see original paper] Graph RAG Processing Flow

News Knowledge Graph Entity Annotation Statistics Table

2.2 Retrieval and Question Answering

During retrieval, Graph RAG relies not only on traditional keyword matching but also introduces logical reasoning based on entity relationships. According to query requirements, we first use natural language processing technology to extract key information such as names, attributes, and descriptions. We then map the query to corresponding knowledge graph nodes and, based on relationships between nodes, quickly locate knowledge subgraphs that may contain answers [9].

During the reasoning process, noise issues are inevitable. The system employs techniques such as entity linking and subgraph exploration to extract coarse-grained knowledge from dynamic knowledge graphs. Simultaneously, it introduces a self-aware knowledge retrieval method that utilizes the ranking capabilities of LLMs to filter noise and redundant information, improving retrieval result relevance and accuracy. To optimize question-answering performance, we introduce an embedding-based subgraph retrieval model for complex queries that uses deep learning technology to learn the mapping relationship between semantic information and knowledge graphs, thereby enabling effective parsing and response to semi-structured data. In this process, Graph RAG demonstrates significant advantages in multiple aspects. First, it improves knowledge retrieval accuracy: through graph structure modeling, Graph RAG can more precisely retrieve relevant knowledge, reducing LLM "hallucination" issues (Peng et al., 2024) [10]. Second, it enhances contextual understanding: graph structures can capture complex relationships between entities, helping LLMs better understand context and generate more coherent content (Dong et al., 2024) [11]. Additionally, it supports multi-hop reasoning: Graph RAG can perform multi-hop reasoning to handle complex problems requiring integration of multiple knowledge sources (Min et al., 2025) [12]. These characteristics enable

Graph RAG to more efficiently and accurately meet user needs in retrieval and question-answering, providing higher-quality support for intelligent writing. Experiments show that Graph RAG can achieve more precise retrieval in complex queries than traditional methods, significantly improving question-answering performance.

(1) Query Entity Recognition and Intent Detection. Query entity recognition and intent detection are indispensable components of intelligent writing. To efficiently utilize knowledge graphs, we must first identify key entities contained in the query and capture the user' s specific intent. Using entity recognition and intent detection models combined with BiLSTM, Attention mechanisms, and CRF algorithms makes the entity recognition process more accurate.

Based on entity recognition results, we introduce a pre-trained BERT model to extract potential user needs by incorporating contextual information. Through a Softmax classifier, we achieve precise discrimination of user intent. For example, user queries may be classified into different categories such as "seeking solutions," "inquiring about person information," or "expressing emotions." Simultaneously, the system employs a multi-turn dialogue mechanism that continuously learns contextual information from queries through neural networks, gradually narrowing the detection scope and improving the model' s understanding of complex query intent.

For specific writing needs such as biographies or event news, we optimized the entity and intent detection models, ensuring high efficiency and accuracy across different writing scenarios, thereby effectively improving the overall performance of intelligent writing optimization technology.

(2) Embedding-based Subgraph Retrieval. The embedding-based subgraph retrieval method is an approach that can precisely extract information from large knowledge graphs. This method can convert specific information and user intent (such as writing themes and content points) into vector representations, quickly find the most relevant subgraphs through vector retrieval, and then use LLMs to integrate this information into useful, coherent text. Practice has shown that adopting embedding-based subgraph retrieval significantly enhances the system' s ability to meet complex creative needs, making the entire writing process smoother and more intelligent, providing richer creative inspiration and background information to help creators produce higher-quality works [13].

In embedding-based knowledge base retrieval (EBR-based Knowledge Base Retrieval), we analyze user queries to extract a set of key named entities P and use these entities to locate the top k historical questions or entries most relevant to user needs in the knowledge graph. We calculate the similarity between entities in the query and nodes in the knowledge graph, computing the cosine similarity between entity value v and all graph nodes n corresponding to part k for each entity pair (k, v) P. We then aggregate node scores belonging to the same query question or other information records to rank these records, thereby improving retrieval precision. Consequently, this method can quickly identify the top k historical query records most relevant to the user' s current query and extract relevant subgraphs accordingly, improving retrieval accuracy [14].

(3) Answer Generation. Answer generation is a crucial component of Graph RAG technology, involving the synthesis of content that aligns with user query intent using retrieved relevant information. To achieve high-quality answer generation, textual information must perfectly align with query intent to ensure answer relevance and accuracy. The system employs a Transformer architecture-based sequence-to-sequence (Seq2Seq) model that first encodes the input query intent and relevant subgraphs, then generates answers during the decoding phase based on the encoded information.

During the encoding stage, the model integrates entities with attributes, relationships between entities, attribute values, etc., to form a comprehensive representation vector. It adopts a multi-dimensional attention mechanism to more accurately capture subgraph features highly relevant to query intent. Introducing entity-level and relationship-level attention enables the model to better understand the semantic structure in knowledge graphs, thereby more accurately selecting relevant information during answer generation.

During the answer decoding stage, to improve generated text conciseness and fluency, we employ a multi-step decoding strategy. We control the number of generation steps through preset thresholds to avoid excessive length or information redundancy. Additionally, by introducing an adverb generation mechanism, we can enhance article expressiveness and impact while maintaining text conciseness.

To validate module effectiveness, we compared results with and without knowledge graph assistance. Results show that answer generation incorporating knowledge graph information is not only more accurate in lexical selection and grammatical construction but also better aligned with user query intent at the semantic level. Simultaneously, for content coherence and logical consistency metrics, the knowledge graph-integrated answer generation module demonstrates superior performance [15].

2.3 Effectiveness Evaluation

To further objectively evaluate the quality of Graph RAG-generated content, we conducted double-blind testing at the AIGC Application Research Center (Guangxi Laboratory) of the China News Technology Workers Federation. Evaluators, unaware of text sources, scored Graph RAG-generated texts and human-written texts on the same topics across multiple dimensions including coherence, completeness, fluency, and creative expression. Results show that in simulated customer service environments, content using the Graph RAG system achieved significant improvements in accuracy and richness. Graph RAG reduced the information error rate by approximately 18% compared to traditional RAG methods in intelligent writing tasks, substantially reducing correction workload caused by information errors.

Regarding generated content quality, we evaluated the generated texts using metrics such as BLEU and ROUGE. Using natural language generation evaluation metrics like BLEU and ROUGE, we objectively assessed the generated texts through machine evaluation. Graph RAG generally improved BLEU-4 scores and increased ROUGE-L scores. This indicates that Graph RAG-generated texts not only retain key information but also better demonstrate text diversity and richness.

Question Answering Performance

Through testing, we found that the key to success lies in effective knowledge graph construction and embedding strategies. This strategy strengthens the close connection between entity relationships and content generation. Additionally, the carefully designed subgraph retrieval algorithm efficiently narrows the retrieval scope, enabling rapid location of required knowledge points. Combined with improved entity recognition and intent detection technologies, Graph RAG demonstrates excellent performance in understanding and responding to user writing needs [16].

3. Application in Intelligent Writing Content Optimization

The core of AIGC intelligent writing content optimization lies in how to enhance the relevance, accuracy, and originality of generated text through technical means. Graph RAG technology injects new vitality into intelligent writing by combining large language models with knowledge graphs, which not only enhances structured text processing capabilities but also substantially improves information retrieval quality.

In intelligent writing, the algorithm further identifies and optimizes information-dense nodes in subgraphs, improving information specificity and utilization. Embedding-based subgraph retrieval enables knowledge factors highly relevant to writing tasks to emerge quickly, reducing retrieval latency and improving information precision and processing efficiency. During the answer generation stage, the optimized subgraph serves as environmental input for reinforcement learning, interacting with intelligent writing algorithms to generate more precise and coherent content. Simultaneously, continuous iteration and optimization of knowledge graphs and generation models based on user feedback make writing results more aligned with user needs and content more suitable for practical application scenarios.

The introduction of Graph RAG technology significantly improves system writing quality. Based on half-year experimental data, content generation error rates decreased by approximately 30%, while information consistency and accuracy improved by about 40%. Such data results demonstrate the tremendous potential and practical value of Graph RAG technology in the intelligent writing field.

Over the past year, the practical application of Graph RAG technology at the AIGC Application Research Center (Guangxi Laboratory) of the China News Technology Workers Federation has fully proven its significant effectiveness and feasibility in optimizing intelligent writing content. By integrating traditional RAG with modern knowledge graphs, this technology not only maintains the intrinsic logic and inter-textual relevance but also significantly improves information retrieval precision, thereby substantially enhancing text quality and effectively solving information fragmentation problems. Additionally, Graph RAG technology demonstrates excellent scalability, applicable to various intelligent writing scenarios including news writing, academic generation, report compilation, and official document writing. Looking ahead, Graph RAG has broad application prospects in intelligent writing, with potential to shine in fields such as short video script creation, novel writing, official document drafting, and technical reporting, providing authoritative and efficient intelligent writing assistance for various industries.

However, Graph RAG technology also faces some challenges, such as high graph construction costs, relatively fixed retrieval methods, and dependence on long-context reasoning (Luo et al., 2025) [17]. Additionally, some studies have shown that in certain practical tasks, Graph RAG performance may be inferior to traditional RAG (Xiang et al., 2025) [18]. In the future, with continuous technological optimization and expanding application scenarios, Graph RAG technology is expected to become an important infrastructure in the intelligent writing field, bringing more efficient and precise intelligent support to content creation.

References: [1] Hu, Y., Lei, Z., Zhang, Z., Pan, B., Ling, C., & Zhao, L. (2024). GRAG: Graph Retrieval-Augmented Generation. arXiv preprint arXiv:2405.16506. [2] Edge D, Trinh H, Cheng N, et al. From Local to Global: A Graph RAG Approach to Query-Focused Summarization[J]. arXiv preprint arXiv:2404.16130, 2024. [3] Liang Y. Integration and Innovation of Artificial Intelligence Writing and News Editing[J]. News Tide, 2023(7): 37-39, 52. [4] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, Haofen Wang. Retrieval-Augmented Generation for Large Language Models: A Survey[J]. arXiv preprint arXiv:2312.10997v5 [cs.CL], 2024. [5] D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody, S. Truitt, J. Larson. From Local to Global: A Graph RAG Approach to Query-Focused Summarization [cs.CL]. arXiv preprint arXiv:2404.16130v1, 24 Apr 2024. [6] Dong, Y., Wang, S., Zheng, H., Chen, J., Zhang, Z., & Wang, C. (2024). Graph-Based Retrieval-Augmented Generation for Complex Knowledge Reasoning. arXiv preprint arXiv:2411.03572. [7] Huang W, Zhou W, Li H, Zhang T. Construction of Chinese Intangible Cultural Heritage Knowledge Graph Based on BERT+BILSTM+CRF[J]. Modern Computer, 2023, 29(24): 74-78. [8] Derong Xu, Wei Chen, Wenjun Peng, Chao Zhang, Tong Xu, Xiangyu Zhao, Xian Wu, Yefeng Zheng, Yang Wang, Enhong Chen.

Large Language Models for Generative Information Extraction: A Survey[J]. arXiv preprint arXiv:2312.17617v2 [cs.CL], 2024. [9] Shen S, Yan D, Bian J, He H. Enhanced Semantic Representation and Retrieval Based on Academic Knowledge Graph[J]. Journal of Hunan University (Natural Sciences), 2024, 51(6): 108-118. [10] Peng, B., et al. (2024). Graph Retrieval-Augmented Generation: A Survey. arXiv preprint arXiv:2408.08921v2. [11] Dong, Y., et al. (2024). Graph-Based Retrieval-Augmented Generation for Complex Knowledge Reasoning. arXiv preprint arXiv:2411.03572v1. [12] Min, C., et al. (2025). Scalable GraphRAG Deployment Framework for Enterprise Environments. arXiv preprint arXiv:2507.03226v1. [13] Ma H, Qian Y, Leng H, et al. Research Progress Review on Knowledge Graph Embedding[J]. Computer Engineering, 2025, 51(2): 18-34. [14] D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody, S. Truitt, J. Larson. From Local to Global: A Graph RAG Approach to Query-Focused Summarization [cs.CL]. arXiv preprint arXiv:2404.16130v1, 24 Apr 2024. [15] Z. Xu, M.J. Cruz, M. Guevara, T. Wang, M. Deshpande, X. Wang, Z. Li. Retrieval-Augmented Generation with Knowledge Graphs for Customer Service Question Answering[D]. arXiv preprint arXiv:2404.17723v2, 2024. [16] D. Li, S. Yang, Z. Tan, J.Y. Baik, S. Yun, J. Lee, A. Chacko, B. Hou, D. Duong-Tran, Y. Ding, H. Liu, L. Shen, T. Chen. DALK: Dynamic Co-Augmentation of LLMs and KG to answer Alzheimer' s Disease Questions with Scientific Literature [cs.CL]. arXiv preprint arXiv:2405.04819v2, 12 May 2024. [17] Luo, H., E, H., Chen, G., Lin, Q., Guo, Y., Xu, F., Kuang, Z., Song, M., Wu, X., Zhu, Y., & Tuan, L.A. (2025). Graph-R1: An Agentic GraphRAG Framework via End-to-End Reinforcement Learning. arXiv preprint arXiv:2507.21892v1. [18] Xiang, Z., et al. (2025). When to use Graphs in RAG: A Comprehensive Analysis for Graph Retrieval-Augmented Generation. arXiv preprint arXiv:2506.05690v1.

Author Bio: Qin Jun (1980–), male, from Binyang, Guangxi, master' s degree, senior professional title, Director of Information Technology Center at Guangxi Daily Press, research direction: media convergence technology.

Editor: Li Yansong

Note: Figure translations are in progress. See original paper for figures.