Translation and Linguistics
Transling

e-ISSN: 2807-3924 | p-ISSN: 2807-2766
https://doi.org/10.20961/transling.v5i3.109928
https://jurnal.uns.ac.id/transling
5(3), 190-203, 2025

A Knowledge-Graph-Based Intelligent Agent for Domain-Specific Question Answering

Mengxue Zhang[1*], Wangwu Huang[2]
[1,2] School of Foreign Languages, Hubei University of Technology, Wuhan, China
*Corresponding Author: 2959629127@qq.com

Article history:
Received
14-10-2025
Accepted
28-11-2025

Keywords:
linguistics; GraphRAG; intelligent agents; knowledge graph; Large Language Models

This is an open-access article under the CC BY SA license.

Abstract: This research focuses on the development of a question-answering intelligent agent for specialized domains based on a knowledge graph, using the field of English linguistics as an example. English linguistics encompasses a vast array of conceptual terminology and a complex knowledge structure that often lacks sufficient concrete language examples for illustration. Traditional information retrieval methods fall short in meeting the advanced demands of semantic comprehension and knowledge inference required in linguistics education. To address these challenges, this study employs Graph Retrieval-Augmented Generation (GraphRAG) technology to design and implement an intelligent agent on the Wenxin Agent Platform. This agent integrates large language models with structured knowledge graphs to enhance the performance of real-time question-answering. The knowledge graph functions as a structured repository that organizes domain-specific linguistic knowledge, thereby expanding the agent's knowledge reserve and enabling more precise responses that facilitate students' systematic understanding of linguistic concepts. Evaluation results indicate that the proposed agent surpasses general-purpose models in response accuracy and professionalism, depth and logical coherence, and explanatory transparency, while effectively addressing hallucination issues common in conventional large language models. The system demonstrates enhanced domain adaptability and improved reasoning performance within linguistic contexts.

Citation: Zhang, M. & Huang, W. (2025). A Knowledge-Graph-Based Intelligent Agent for Domain-Specific Question Answering. *Translation and Linguistics (Transling), 5* (3), 190-203. https://doi.org/10.20961/transling.v5i3.109928

1. INTRODUCTION

1.1 Background

The rapid advancement of artificial intelligence and natural language processing (NLP) has propelled large language models (LLMs) to the forefront of technology, demonstrating remarkable capabilities in semantic understanding, text generation, and intelligent question answering (Zhou & Xiao, 2024). Consequently, their application is expanding into knowledge services for specialized domains such as finance, healthcare, and law (Wang, C. B. et al., 2024). Deep learning-based NLP models, notably GPT-4, are seeing growing adoption in education. Trained on extensive educational corpora, these models can comprehend and generate natural language, enabling them to provide students with accurate explanations and tailored learning guidance (Gao et al., 2025). For instance, research by Brown et al. indicates that GPT-3 excels at generating educational content and addressing student inquiries, significantly improving learning efficiency and user experience (Brown et al., 2020).

However, question answering in specialized domains often involves intricate knowledge structures and demands rigorous logical reasoning. When lacking adequate domain-specific knowledge, LLMs are susceptible to hallucination, generating text that is fluent and grammatically correct but contains factual inaccuracies or logical inconsistencies that conflict with the input context or established knowledge (Dash et al., 2023). This tendency impedes their reliability in specialized academic applications. Further constraints include substantial hardware requirements and challenges in updating their knowledge, which hamper their efficacy in dynamically evolving fields (Dong et al., 2025). These limitations largely originate from their dependence on static pre-trained corpora for knowledge acquisition and an inability to interact effectively with structured knowledge sources (Jiang et al., 2025). Meanwhile, traditional information retrieval methods, which rely heavily on keyword matching and static databases, are often hampered by limited semantic understanding, weak reasoning abilities, and poor result interpretability (Ma, 2018). These limitations make it difficult for such methods to meet the increasing need for precise knowledge extraction and advanced intelligent reasoning (Wang, B. et al., 2022).

In recent years, the field of English linguistics has experienced a rapid accumulation of research outputs, including academic papers, corpus data, experimental reports, and other multi-source heterogeneous materials. Much of this information exists in unstructured formats with fragmented semantic representations, rendering traditional keyword-based static retrieval methods insufficient for meeting the demands of problem-oriented semantic reasoning and in-depth information extraction. Simultaneously, linguistics instructors encounter numerous challenges in the classroom: the knowledge system is not only expanding but also increasingly interdisciplinary, making it difficult to cover essential content and incorporate recent advancements within constrained teaching hours. Instruction often remains superficial, lacking dynamic demonstrations of the underlying semantic relationships among complex theories or concrete illustrative examples. Furthermore, the dispersal of teaching resources compounds instructors' preparation efforts, hindering their ability to adapt instruction to diverse student needs. Students, on the other hand, face considerable obstacles in learning linguistics: confronted with abstract and poorly interconnected concepts, they often struggle to grasp the broader theoretical landscape, a classic case of "missing the forest for the trees." Without the capacity for self-directed exploration or semantic-level search, they find it difficult to synthesize knowledge from multiple learning resources effectively. Thus, there is a pressing need to develop a system equipped with capabilities

for knowledge organization, semantic comprehension, and logical reasoning to enable deep integration and intelligent linking of multi-source heterogeneous linguistic knowledge. Such a system would empower both teachers and students to perform semantic-aware retrieval and inference, thereby improving instructional efficiency and deepening learning outcomes. Its implementation would not only help overcome the inherent limitations of conventional classroom teaching but also offer students autonomous, structured, and problem-driven integrated learning support, effectively alleviating critical bottlenecks in both teaching and learning processes.

1.2 Retrieval-Augmented Generation (RAG) technology

To overcome the limitations of general-purpose models in semantic understanding and knowledge integration, the recently developed Retrieval-Augmented Generation (RAG) framework provides a promising alternative for constructing efficient question-answering systems (Chen et al., 2024; Siriwardhana et al., 2023). To mitigate these issues, MetaAI introduced Retrieval-Augmented Generation (RAG) in 2020 as a promising advancement in natural language processing (Fan et al., 2025). RAG improves the accuracy and reliability of model responses by incorporating information retrieval with text generation, drawing on external knowledge bases to produce well-grounded answers (Fan et al., 2025).

The RAG framework combines a retrieval module and a generation module within a unified model. By combining an external knowledge retrieval module with a neural language generator, RAG considerably improves a model's capacity to incorporate and leverage contextual information, exhibiting robust performance across general-purpose QA benchmarks (Wen et al., 2024). During content generation, the system retrieves pertinent text passages via vector similarity matching. This enables the model to supplement its parametric knowledge with dynamically retrieved external information, enhancing output veracity and reducing hallucinations in LLM-generated responses (Lewis et al., 2020). RAG establishes a tight coupling between external knowledge bases and LLMs: it uses LLMs to strengthen the multi-hop reasoning ability of knowledge bases, while the knowledge bases in turn provide curated domain knowledge to counteract hallucinations often found in general-purpose LLMs like DeepSeek (Siddharth & Luo, 2024).

1.3 Graph Retrieval-Augmented Generation (GraphRAG) technology

Nevertheless, conventional RAG methods depend on vector-based retrieval of unstructured text snippets, which can yield results lacking in semantic coherence and logical structure (Li et al., 2025). To address these issues, GraphRAG, an emerging extension of RAG augmented with knowledge graphs, effectively counteracts these weaknesses through the integration of graph-based representations and structured reasoning mechanisms (Buehler, 2024). GraphRAG, an enhanced variant of RAG, effectively tackles this limitation by incorporating graph-based representations (Ma, W. L., Sun, et al., 2025).

GraphRAG is an innovative framework that extends the original RAG architecture. By integrating knowledge graphs, it transforms textual knowledge into structured representations composed of entities and their interrelationships. The incorporation of knowledge graphs (KGs) offers a valuable technical enhancement. By providing structured and systematic modeling of domain knowledge, KGs supply large language models with reliable contextual support, thereby increasing the factual accuracy and logical coherence of generated outputs (Allemang & Sequeda, 2024). The framework facilitates multi-hop retrieval over the graph structure, enabling the identification of both directly pertinent entities and indirectly connected subgraph pathways. This ability

overcomes a major drawback of conventional RAG systems, which often fail to capture latent cross-document information, thus significantly improving the coherence and comprehensiveness of generated outputs (Ma, W. L., Sun, et al., 2025). As an advanced variant of RAG, GraphRAG enables the retrieval and semantic consolidation of relevant subgraphs from knowledge bases. It provides configurable options across three critical aspects: the retriever model, the retrieval strategy, and the level of retrieval granularity, allowing it to be tailored to specialized and intricate application contexts (Wang, H. et al., 2025).

Furthermore, knowledge graphs support dynamic updating, enabling the continuous assimilation of new information, effectively mitigating the knowledge update problem common in static LLMs (Dong et al., 2025). GraphRAG is a hybrid methodology that combines large language models with knowledge graphs within a retrieval-augmented generation paradigm. By accessing structured information from knowledge graphs, it supplies factual grounding during the LLM's generation phase, thereby increasing the accuracy and trustworthiness of system responses (Dong et al., 2025). The method employs graph-based representations to formalize domain knowledge, utilizing the inherent topology of knowledge graphs to make explicit the complex associations among entities. When coupled with the semantic comprehension and text generation abilities of large language models, GraphRAG enables effective knowledge access and fluent answer synthesis (Peng et al., 2024).

1.4 Objective and contributions

The objective of this study is to design and implement an intelligent agent leveraging a large language model and a knowledge graph to enhance the performance of intelligent question answering on specialized domains. This study takes the domain of English linguistics as its research context. This agent is intended to alleviate common student struggles such as knowledge fragmentation and comprehension challenges by offering real-time Q&A support, thereby stimulating interest and encouraging autonomous study.

The main contributions of this work are as follows:

(1) Construction of a specialized knowledge graph for English linguistics by transforming unstructured textual resources into structured knowledge representations. This graph offers extensive, systematic associations and navigational pathways across knowledge units, allowing learners to efficiently explore concepts, retrieve contextualized information, and support autonomous study (Gao et al., 2025).

(2) Development of an intelligent agent through the configuration of the knowledge base and large language model to facilitate the performance of intelligent question answering on English linguistic knowledge.

2. METHOD

2.1 Research questions

This study aims to address the following core research questions:

(1) How to construct a knowledge graph for the field of linguistics.

(2) How to build an intelligent agent based on large language model and knowledge graph.

(3) Whether the constructed intelligent agent can effectively utilize structured knowledge from the knowledge graph to generate responses with higher factual accuracy and reduced hallucinations.

2.2 Research methodology (Technical approach)

The methodology of this study centers on Graph Retrieval-Augmented Generation (GraphRAG), a framework that integrates a large language model with a knowledge

graph. The technical approach entails two key phases: the construction of a linguistic knowledge graph and systematically configuring the intelligent agent.

An "intelligent agent" is defined as an entity that perceives its environment and acts autonomously to achieve designated goals (Wei, 2024). Large language models play a pivotal role in building intelligent agents. Their advanced natural language processing capabilities support a variety of intelligent assistive functions in educational contexts (Wang, R., 2024; Xu et al., 2024). Knowledge graphs, conversely, are structured knowledge representation technologies that organize information into networks of entities, attributes, and relationships. They enable machines to comprehend knowledge and support its retrieval, reasoning, and practical application (Zheng et al., 2023). This research, being based on GraphRAG technology, combines large language models with knowledge graphs by first converting unstructured textual data into a knowledge graph, then extracting the structured information into the knowledge base of the Wenxin Intelligent Agent platform. Once configured, the deployed agent does not rely exclusively on the generative capacity of the Wenxin model when formulating responses, it also retrieves and integrates relevant information from the knowledge graph. This methodology effectively alleviates issues such as hallucinations in large language models, particularly within the domain of linguistics.

2.3 Working mechanism and advantages of the technical approach--GraphRAG

The deployed intelligent agent utilizes a Graph Retrieval-Augmented Generation (GraphRAG) architecture to process user queries. This workflow constitutes an advanced evolution of the classic RAG paradigm: when a user submits a linguistics-related query, the agent first conducts a semantic search within the vector knowledge base, constructed from the knowledge graph, to retrieve contextually relevant information. In contrast to traditional RAG, which operates over unstructured text, this approach queries a structured knowledge network enriched with entity and relationship semantics to identify the most pertinent information segments (Context). The system subsequently integrates these retrieved, logically interconnected knowledge segments with the original user query to construct a structured and enriched prompt, which is then processed by the "Wenxin Large Model 3.5." By leveraging both factual knowledge and underlying relational logic, the large language model engages in deep, comprehensive reasoning to produce responses that are not only accurate and reliable but also exhibit improved logical consistency and coherence.

The key advantages of this mechanism represent substantial advancements over traditional RAG:

Deeper reasoning and improved answer accuracy: Whereas traditional RAG retrieves isolated information snippets from plain text, GraphRAG supplies interconnected knowledge fragments extracted from the knowledge graph. This provides the large language model with enriched reasoning pathways and relational context, facilitating multi-hop reasoning, inconsistency resolution, and inductive summarization. Consequently, the system generates more logically robust and precise answers while effectively reducing hallucinations.

Enhanced traceability and explainability: Responses generated by the system can be traced back to specific entities and relational nodes within the knowledge base. This capability not only enables factual verification but also illuminates the underlying generative logic and reasoning trajectories (e.g., which entity attributes and relationships contributed to the answer), thereby substantially increasing the model's transparency and trustworthiness.

Effective processing of complex queries: For complex inquiries involving multiple entity relationships (e.g., "Compare the similarities and differences between Theory A and Theory B"), traditional RAG often retrieves disjointed information segments. In contrast, GraphRAG capitalizes on its inherent graph topology to retrieve coherent contextual information that aligns with all relevant entities and relationships, markedly improving performance on sophisticated queries.

2.4 The construction process of the intelligent agent based on large language model and knowledge graph

The intelligent agent for this study was developed and configured on the Wenxin Intelligent Agent platform. The detailed procedure for its construction is outlined below. The workflow of the construction process is depicted in Figure 1.

2.4.1 Building the knowledge graph in the domain of English linguistics

An automated GraphRAG pipeline driven by a large language model (LLM) was implemented to transform unstructured linguistic texts into a structured knowledge graph. This process operates without manual annotation and comprises the following core stages:

Entity and relation extraction: The pipeline initiates with text chunking, segmenting raw documents into manageable units (chunks) appropriate for computational processing. Each chunk is processed sequentially by the LLM for structured information extraction. This is accomplished through carefully engineered system prompts that guide the model to identify entity names (e.g., “General Linguistics”, “Saussure”), classify entity types, generate descriptive summaries (encapsulating core definitions, attributes, or functions), extract semantic relations between entities (e.g., “influenced”, “is a branch of”), and recover fully articulated claims, including subject, object, certainty measure, and source attributions. The outputs are structured as tuples, establishing the foundational nodes and edges of the graph.

Knowledge cleaning, enrichment, and vectorization: The initially extracted data undergoes refinement in subsequent stages. Entity description consolidation employs the LLM to integrate, deduplicate, and refine fragmented entity references, yielding consistent and high-quality definitions. Thereafter, entity embeddings are produced via a text embedding model, encoding each entity into a high-dimensional vector representation. These embeddings are stored in a LanceDB vector database to enable efficient semantic similarity retrieval.

Graph structure mining and knowledge abstraction: This phase constitutes a principal advantage of GraphRAG over conventional RAG. Community detection utilizes graph clustering algorithms (e.g., Louvain) to analyze the entity-relationship graph, automatically aggregating semantically related entities into thematic communities. The most novel component is community report generation, wherein the LLM is tasked, acting as a domain specialist, to conduct thorough analyses of each community and automatically generate structured JSON reports comprising titles, summaries, evaluative ratings, and detailed insights. These reports provide a synthesized interpretation of the underlying graph structure, elevating raw data into abstracted knowledge.

The pipeline concludes by serializing all refined components, cleaned entities, relations, community assignments, community reports, and embeddings, into a collection of Parquet files. These files serve as a comprehensive and structured repository for the knowledge graph, which is subsequently visualized within Neo4j.

2.4.2 Extracting the knowledge graph data

To facilitate the transfer and utilization of linguistic knowledge across heterogeneous platforms, structured data extracted from a Neo4j graph database serves as the

foundational knowledge source for agent construction. Specifically, Cypher, Neo4j's declarative graph query language, was used to design targeted queries for retrieving comprehensive sets of entities (nodes) and their interrelationships from the knowledge graph. The retrieved data is systematically extracted into multiple comma-separated values (CSV) files. This procedure constitutes more than a simple data extract, it represents a essential stage of structured data serialization. By effectively translating nodes, attributes, and relations from the graph into a universal tabular format independent of any specific platform, the process establishes a standardized, machine-readable knowledge base essential for enhancing the large language model with structured domain knowledge.

2.4.3 Building the knowledge base on the Wenxin Intelligent Agent Platform

The extracted CSV files from knowledge graph are subsequently uploaded to the knowledge base module of the Wenxin Intelligent Agent Platform. This module functions not as simple file storage, but as a vector database system with advanced natural language processing capabilities. During ingestion, the platform automatically parses, segments, and converts textual content from the CSV files, including entity names, attributes, and relationship types, into high-dimensional vector embeddings. These embeddings are indexed within an optimized vector space, ensuring that semantically related queries and knowledge elements are proximally located. Through this procedure, linguistic knowledge originally represented as a graph structure in Neo4j is effectively transformed and enriched into a vector knowledge base capable of high-speed semantic similarity search. This infrastructure enables the intelligent agent to move beyond keyword matching and achieve nuanced semantic comprehension and accurate knowledge retrieval.

2.4.4 Configuring the intelligent agent for English linguistics

Agent configuration plays a vital role in determining its efficacy as an educational assistant. The following core configurations were applied within the Wenxin Intelligent Agent Platform:

Configuring the intelligent agent with a large language model: The "Wenxin Large Model 3.5" was selected as the agent's core reasoning and generation engine. This model provides strong natural language understanding and generation abilities, offering a well-balanced performance profile that underpins fluid and context-aware dialogue.

Configuring the intelligent agent with a knowledge base: The agent was integrated with the previously constructed (in 2.4.3) knowledge base. This setup implements the Graph Retrieval-Augmented Generation (GraphRAG) paradigm, wherein the agent first retrieves relevant information from the knowledge graph-based repository (repository means knowledge base) before supplying this structurally rich context, populated with entities and relationships, to the large language model. This process significantly improves the accuracy and logical consistency of generated responses.

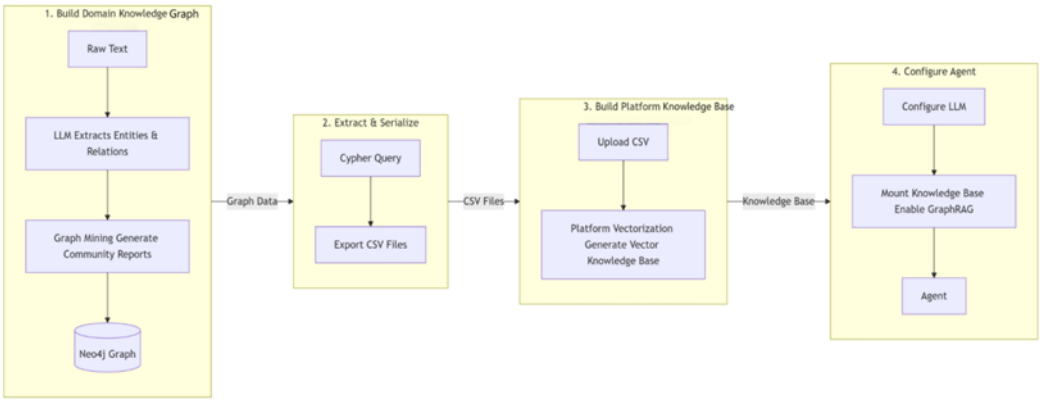

Figure 1

2.5 Comparative experiments and evaluation

This study aims to evaluate the performance of a GraphRAG-based intelligent agent (experimental group) on linguistic knowledge question-answering tasks. The agent's performance is compared against that of the general-purpose large model ERNIE Bot (control group), which operates without any knowledge base retrieval augmentation, using the same set of evaluation questions. The experiment is designed to test the following hypotheses:

(1) This agent can effectively utilize structured knowledge from the knowledge graph to produce more accurate responses with reduced factual inaccuracies (i.e., hallucinations).

(2) Compared to general-purpose large models, this agent can generate answers with greater depth and more logical coherence.

A representative question, “Please explain and compare the differences and connections between 'Synchronic Linguistics' and 'Diachronic Linguistics'.”, was selected as the test case.

3. FINDING AND DISCUSSION

3.1 Analysis of the comparative experiment results

Here are the responses from both systems to this question.

Figure 2. Intelligent Agent

Figure 3. ERNIE Bot

From the results of the comparative experiment, we have verified the two hypotheses stated above. First, we can know that the constructed intelligent agent can effectively utilize structured knowledge from the knowledge graph to generate responses with higher factual accuracy and reduced hallucinations, which also serves as an answer to the third research question. Second, this agent can generate answers with greater depth and more logical coherence. A more detailed analysis is as follows.

In terms of factual accuracy and professionalism, ERNIE Bot's responses (Figure 3) captured the core definitions correctly but suffered from a lack of rigor and precision. For instance, it described synchronic linguistics as focusing on the "current structure and function" of language, a formulation that is unduly restrictive, since synchronic analysis may target any fixed time point, not exclusively the present. In contrast, the intelligent agent (Figure 2) delivered a precise definition, clearly stating that synchronic linguistics examines "the non-historical aspects of language at a specific point or given period," while diachronic linguistics investigates "change and development over time." The agent's phrasing was rigorously professional and aligned closely with established textbook definitions. The entire response was free from factual inaccuracies or misleading examples, demonstrating superior accuracy and a significant reduction in hallucinations. This improvement can be attributed to the agent's use of structured knowledge retrieved from the knowledge base, where core information is sourced directly from authoritative structured data, substantially mitigating hallucination.

In terms of depth and logical coherence, ERNIE Bot's answers exhibited limited depth and logical integration, presenting concepts in isolation without connecting them to broader linguistic developments or related dimensions. The intelligent agent, on the other hand, did not emphasize the "complementarity" of the two subfields but also introduced the more nuanced concept of "interdependence." It explicitly articulated that "synchronic research requires diachronic historical context for support, while diachronic research relies on synchronic states as reference," reflecting a sophisticated grasp of their

dialectical relationship. Moreover, the agent elevated the discussion in its conclusion by situating the concepts within the wider context of "linguistic research," alluding to the complexity and diversity of language, and linking the ideas to the evolution of the discipline, rather than treating them in isolation. This capability stems from the knowledge graph's capacity to store not only factual nodes but also relational structures. By providing access to interconnected concepts and their positions within a broader semantic network, the knowledge base equips the agent's reasoning core, the large language model, with rich contextual and inferential pathways, resulting in responses of greater depth and logical coherence.

In terms of traceability and explainability, the provenance of ERNIE Bot's responses cannot be ascertained, as they are generated based on its training data, which may commingle accurate and erroneous information, with no mechanism for verification. In contrast, the intelligent agent's answers allow retrieval logs to be examined, making it possible to trace which specific entities and relationships in the knowledge base contributed to the response. This inherent verifiability and explainability represent a distinct advantage over general-purpose large models.

3.2 Reasons behind the agent's performance advantages

The superior performance of the intelligent agent developed in this study, in terms of response accuracy, domain expertise, and explanatory transparency, is not coincidental. The underlying reason lies in the GraphRAG architecture's effective integration of the semantic comprehension capabilities of large language models (LLMs) with the structured reasoning strengths of knowledge graphs (KGs), thereby overcoming inherent limitations of both conventional LLMs and standard RAG systems.

First, incorporating a knowledge graph supplies the LLM with high-quality, structured domain-specific context. The linguistic knowledge graph used by this agent systematically organizes otherwise scattered and isolated linguistic concepts, such as "synchronic linguistics" and "diachronic linguistics", into entities and relations, forming a semantically enriched associative network. When processing user queries, rather than depending exclusively on internal parametric knowledge, which can be incomplete or outdated, as in general-purpose models, the agent first retrieves relevant entities and their surrounding relational subgraphs from the knowledge graph. This procedure supplies logically interconnected knowledge that is subsequently used during LLM generation. By doing so, it substantially reduces the likelihood of hallucinations at the source, thereby strengthening response accuracy. For instance, when defining "synchronic linguistics," the agent delivers a precise explanation because the knowledge nodes it retrieves are derived from authoritative structured sources, rather than being probabilistically generated by the LLM.

Second, the GraphRAG architecture facilitates a paradigm shift from "fragment retrieval" to "relational reasoning," empowering the system with advanced multi-hop inference capabilities. Standard RAG retrieves information snippets (or chunks) from large volumes of unstructured text, which are often semantically isolated and lack interconnections. In contrast, the knowledge graph's structure inherently represents semantic relationships between concepts via edges (relations). During retrieval, the agent acquires not isolated pieces of text, but a compact relational network. When generating answers, the LLM can leverage these explicit logical connections for inference. This capability is the key reason why the agent can articulate complex relationships, such as the complementarity and interdependence between synchronic and diachronic linguistics, in depth. It not only recognizes individual concept definitions but also

comprehends their contextual positions and interactions within the knowledge network, enabling comparative, inductive, and deductive reasoning (i.e., multi-hop reasoning).

3.3 Challenge

Although the results of this study are promising, the implemented agent still encounters a significant challenge, which indicates valuable avenues for future research.

The retrieval and reasoning efficiency for complex queries requires further optimization. Queries involving numerous entities and deep relational paths may result in computational overhead due to the retrieval of large subgraphs and complex reasoning processes, ultimately increasing response time. Enhancing retrieval algorithms and developing more effective reasoning-path pruning strategies, without compromising answer quality, are crucial for improving user experience.

3.4 Application of the linguistic intelligent agent in teaching and learning

The *English Curriculum Standards for Compulsory Education (2022 Edition)* explicitly advocate for "fully leveraging modern information technology to support and serve English language teaching and learning." It encourages educators to use digital technologies in rational and innovative ways to effectively address students' individualized learning needs (Ministry of Education of the People's Republic of China [MOE], 2022). In the age of artificial intelligence, greater emphasis is placed on students' autonomous learning capabilities. Students are expected to choose learning materials and practice content according to their progress and needs, and to independently manage their study schedules. This makes learning more targeted and flexible, catering to personalized educational demands and improving outcomes. Consequently, the intelligent agent becomes a supportive tool rather than a substitute for learning (Zhong, 2025).

Regarding teaching content generation, the agent can automatically produce course handouts, summaries of key concepts, and analyses of representative examples by drawing on structured linguistic concepts, theoretical frameworks, and relational data within the knowledge graph. For example, when teaching the "difference between phonetics and phonology," the system can retrieve definitions, research priorities, and viewpoints of prominent scholars from both subfields, generate comparative analyses, and supply relevant teaching examples and discussion materials, significantly reducing preparation workload.

In terms of learning process support, the intelligent agent can identify knowledge weaknesses based on the cognitive state demonstrated by students during interactions, and subsequently deliver targeted exercises and explanatory content. For example, if a student performs poorly in a test related to "syntactic tree diagram analysis," the agent can not only retrieve and explain key rules but also generate step-by-step training exercises based on related concepts within the knowledge graph, thereby strengthening the student's understanding of the knowledge architecture.

In the domain of academic assessment, the agent utilizes the knowledge graph to automatically construct multi-dimensional and hierarchically structured test items. It can evaluate not only grasp of discrete concepts but also devise integrated, application-based questions that encourage a transition from memorization to analytical and creative thinking. Simultaneously, automated scoring and real-time feedback mechanisms allow instructors to promptly assess class-wide understanding and refine teaching strategies accordingly.

4. CONCLUSION

This study designed and implemented an intelligent agent for English linguistics using a GraphRAG-based framework, which combines the natural language processing capabilities of the Wenxin Large Model 3.5 with the semantic support of a structured knowledge graph to improve both the efficiency and depth of linguistics education.

Through the construction of a linguistic knowledge graph, unstructured text was converted into a semantic network composed of entities, attributes, and relationships. Integrated with the retrieval augmentation of the Wenxin Agent Platform, the system provides high-precision and interpretable answers to complex linguistic questions. Experimental findings indicate that, in comparison to general-purpose large language models, the proposed agent delivers superior performance in factual accuracy, logical coherence, and multi-hop reasoning, notably reducing hallucinations and improving answer traceability. The system functions not only as an intelligent aide for self-directed learning, but also as a support tool for instructors in creating teaching materials and evaluating student progress. It strongly supports the principle endorsed in the *English Curriculum Standards for Compulsory Education (2022 Edition)* calling for "deep integration of information technology with English teaching."

Future work can focus on the following key directions. Firstly, optimizing reasoning efficiency for complex queries. To tackle potential latency issues in processing queries involving deep relational chains, we will explore more efficient subgraph retrieval algorithms and reasoning path pruning strategies. The goal is to significantly reduce computational overhead without compromising answer quality, thereby enhancing user experience and enabling the agent to handle multi-hop reasoning tasks more fluently. Secondly, future research will extend beyond merely converting unstructured text into a knowledge graph. It aims to construct a “holographic” knowledge graph capable of comprehending and integrating multimodal information (text, images, audio, and video), thereby enabling a more three-dimensional and comprehensive representation of linguistic knowledge. This approach can significantly enhance the intuitiveness and depth of learning.

REFERENCES

Allemang, D., & Sequeda, J. (2024). Increasing the LLM accuracy for question answering: Ontologies to the rescue! *arXiv*. https://doi.org/10.48550/arXiv.2405.11706

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv*. https://doi.org/10.48550/arXiv.2005.14165

Buehler, M. J. (2024). Generative retrieval-augmented ontologic graph and multiagent strategies for interpretive large language model-based materialsdesign. *ACS Engineering Au*, *4*(2), 241–277. https://doi.org/10.1021/acsengineeringau.3c00063

Chen, J., Lin, H., Han, X., Liang, H., Wu, W., & Shi, H. (2024). Benchmarking large language models in retrieval-augmented generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(16), 17754–17762. AAAI Press.

Dash, D., Thapa, R., Banda, J. M., Swaminathan, A., Cheatham, M., & Kaggal,V. (2023). Evaluation of GPT-3.5 and GPT-4 for supporting real-world information needs in healthcare delivery. *arXiv*. https://doi.org/10.48550/arXiv.2304.13714

Dong, Z. A., Qin, K. H., Zhou, Z. L., Wang, S. F., Xiong, J., & Liu, Z. B. (2025). GraphRAG-based question answering system for traditional Chinese medicine knowledge. *Journal of*

Qufu Normal University (Natural Science Edition), 1–11.

Fan, X. S., Li, Z. Q., Zhang, X. H., & Wang, S. L. (2025). A Q&A system forsteel industry standards based on large models and knowledge graph-enhanced RAG. *Metallurgical Automation*, *49*(4), 134–145.

Gao, L. P., Cai, W. J., Zheng, W., Zheng, J. B., & Ma, C. Y. (2025). Application and practice of an intelligent tutoring system based on large models and knowledge graphs in blended teaching of software testing courses. Software Guide, 1–8.

Jiang, X., Wang, H. Y., Yang, S. T., Chen, X. Y., Zhou, M. Y., & Dong, L. G. (2025). Research on multi-hop question answering technology based on the integration of large language models and knowledge graphs. Telecommunications Science, 1–25.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474.

Li, B. W., Wang, Y. Z., Ding, Z. J., Wang, B., Wen, S. B., Dong, Y. H., & Ji, Z.(2025). Intelligent search technology for Jiaodong gold deposits based on large models and GraphRAG. *Earth Science Frontiers, 32*(4), 155–164. https://doi.org/10.13745/j.esf.sf.2025.4.77

Ma, K. (2018). Research on key technologies of geological big data representation and association [Doctoral dissertation, China University of Geosciences]. CNKI. https://kns.cnki.net/kcms2/article/abstract?v=9oehDy4zW5YO2ZN3AfUGiOSiMDg3yk7NulMyYZDKHndWnqU7ftlZzWDmDNf2Mt166oEd9pIHnZlMNcI2bxif6oVNFK_-qhKKgOH9sA3MLJzLcdg67t1EfNtw6dQY05clX08s0A2CB0RhH3tZWkSQGtVHIPLadrakvyjWNkc9lBDq7U-ZuYMYGNQMgiHS-_kG

Ma, W. L., Sun, T., Zhao, R. X., & Xian, G. J. (2025). Generating graphical summaries of scientific literature based on LLM and GraphRAG. Data Analysis and Knowledge Discovery, 1–21.

Ministry of Education of the People's Republic of China. (2022). *Compulsory education English curriculum standards (2022 edition)*. Beijing Normal University Publishing Group.

Peng, B., Zhu, Y., Liu, Y., Li, H., & Zeng, M. (2024). Graph retrieval-augmented generation: A survey. *arXiv.* https://doi.org/10.48550/arXiv.2408.08921

Siddharth, L., & Luo, J. (2024). Retrieval augmented generation using engineering design knowledge. *Knowledge-Based Systems*, *303*, 112410. https://doi.org/10.1016/j.knosys.2024.112410

Siriwardhana, S., Weerasekera, R., Wen, E., & Kaluarachchi, T. (2023). Improving the domain adaptation of retrieval augmented generation (RAG) models for open domain question answering. *Transactions of the Association for Computational Linguistics*, *11*, 1–17. https://doi.org/10.1162/tacl_a_00524

Wang, B., Ma, K., Wu, L., & Wang, C. (2022). Visual analytics and information extraction of geological content for text-based mineral exploration reports. *Ore Geology Reviews*, *144*, 104818. https://doi.org/10.1016/j.oregeorev.2022.104818

Wang, C. B., Wang, M. G., Wang, B., Chen, J. G., Ma, X. G., & Jiang, S. (2024). Quantitative prediction of mineral resources integrating knowledge graphs. *Earth Science Frontiers, 31*(4), 26–36. https://doi.org/10.13745/j.esf.sf.2024.5.3

Wang, H., Yan, W., Zhang, X. M., Zhu, S., Jiang, Z. G., & Zhu, Z. R. (2025). Disassembly sequence planning for end-of-life power batteries based on GraphRAG. China Mechanical Engineering, 1–15.

Wang, R. (2024). The new paradigm of AI-empowered continuing education in universities. *Continuing Education Research, 3*, 12–16.

Wei, Y. D. (2024). The "cognitive turn" of artificial intelligence: The necessary path to artificial general intelligence. *Journal of Shanxi Normal University (Social Science Edition), 51*(5), 9–17. https://doi.org/10.16207/j.cnki.1001-5957.2024.05.010

Wen, S., Qian, L., Hu, M. D., & Chang, Z. J. (2024). A review of question answering technology based on large language models. *Data Analysis and Knowledge Discovery, 8*(6), 16–29.

Xu, S., Tong, J. R., & Hu, X. E. (2024). Next-generation personalized learning: Generative artificial intelligence-enhanced intelligent tutoring systems. *Open Education Research, 30*(2), 13–22. https://doi.org/10.13966/j.cnki.kfjyyj.2024.02.002

Zheng, W., Gu, R., Wu, X., & Ma, C. (2023). Personalized learning path recommendations for software testing courses based on knowledge graphs. *Computer Education, 12*, 63–70.

Zhong, M. (2025). Building an AI agent spoken language practice model based on characters from English textbooks. *Teaching and Administration, 11*,55–58.

Zhou, Y. Z., & Xiao, F. (2024). A glimpse into new advances in artificial intelligence and big data earth science research. *Earth Science Frontiers, 31*(4), 1–6. https://doi.org/10.13745/j.esf.sf.2024.6.99