Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 8, No. 1, 2026

Wind Turbine Fault Maintenance Decision System Based on Knowledge Graphs and Large Language Models

Zhengxi Li[#, *], Zihao Xu[#], Maomao Qi, Ruiyang Shou, Yuxuan Yang

School of Management Engineering, Nanjing Institute of Technology, Nanjing, China, 211167

[*]**Corresponding author**: Zhengxi Li (Email: x00209231222@njit.edu.cn)

[#]**These authors contributed equally.**

Abstract: With the continuous expansion of China's power grid and the deepening of the new energy strategy, wind power is playing an increasingly critical role in the global energy supply. However, the traditional wind power operation and maintenance model can no longer meet the dual challenges of low fault diagnosis efficiency and high maintenance costs, making an intelligent transformation imperative. To address this, this paper targets the issues of low intelligence in wind power fault operation and maintenance and the heavy reliance on expert experience for decision-making. It proposes an intelligent decision-making system that integrates knowledge graphs and Large Language Models (LLMs) and elaborates on the construction method for domain-specific knowledge graphs. By designing the system architecture and integration technology, the advantages of this method in improving the efficiency and accuracy of operation and maintenance decision-making are validated. The contribution of this paper lies in proposing an operation and maintenance decision-making framework that synergizes knowledge graphs and LLMs, clarifying the key technological pathways, and providing theoretical reference and practical direction for the intelligentization of wind power operation and maintenance.

Keywords: Knowledge Graph; Large Language Model; Wind Power System; Fault Maintenance; Decision System.

1. Introduction

According to statistics from the National Energy Administration, by the end of December 2024, the cumulative installed power generation capacity nationwide reached approximately 335 gigawatts, representing a year-on-year increaseof 14.6%. Within this total, wind power capacity amounted to roughly 52 gigawatts, marking an 18.0% year-on-year growth. Wind energy has emerged as a prominent renewable source due to its environmental and pollution-free advantages.[1] On one hand, the complex structure of wind turbine components and thevariable operating environment result in concealed fault characteristics with strong coupling effects, rendering traditional diagnostic methods inadequate for precise localisation and root cause identification. On the other hand, existing maintenance models rely heavily on expert experience and manual troubleshooting, leading to slow response times and high costs. This approach struggles to meet the demands of intelligent management for large-scale, highly concentrated wind farm clusters.[2] These challenges not only constrain the economic viability and reliability of wind power operations but also pose potential risks to the secure and stable operation of the power grid. Addressing these issues, this paper focuses on exploring how an intelligent decision-making system integrating knowledge graphs with large language models in the wind power domain can achieve structured representation and intelligent reasoning of fault knowledge. This approach aims to enhance the accuracy of fault diagnosis, improve responsespeed, and elevate the systematic level of O&M decision-making.

In recent years, knowledge graph technology has continued to advance in structured knowledge representation and reasoning. Large Language Models (LLMs), through their parameterised knowledge and generative capabilities, can effectively handle complex reasoning and multi-hop question-answering tasks [3], providing robust support for natural language interaction and intricate reasoning tasks. This has made knowledge-fusion-based intelligent systems feasible. Chen Hong et al. [4] proposed a knowledge graph-based fault diagnosis system for wind turbines. While improving entity recognition accuracy and introducing fault tree optimisation, it retains key limitations such as the absence of a dynamic knowledge base update mechanism and insufficient deep integration with large language models. Zhao Yan [5] proposed a planetary gearbox fault diagnosis method integrating knowledge graphs with GPT models. This approach employs multi-objective optimisation for feature selection and leverages few-shot learning to construct fault knowledge graphs, significantly enhancing diagnostic accuracy. However, the dynamic adaptability of its feature optimisation and graph construction processes, cross-model generalisation capabilities, and computational efficiency in real-time diagnostics require further validation and refinement. Overall, existing research predominantly focuses on general frameworks or individual technologies, failing to address the coordinated challenges of dynamic organisation and deep reasoning for fault knowledge within the wind power sector. Furthermore, there remains a lack of intelligent decision-making systems capable of covering the entire operation and maintenance lifecycle.

To this end, this study proposes a collaborative decision-making system for wind turbine faults that deeply integrates knowledge graphs with large language models. By constructing a dynamically evolving fault knowledge graph and developing an interactive diagnostic module with multi-step reasoning and semantic comprehension capabilities, it establishes an intelligent operation and maintenance support framework covering the entire process. This system not only achieves precise fault localisation and decision generation for resolution but also delivers interpretable, actionable closed-loop services tailored to diverse roles. It propels wind power operations and maintenance from an experience-driven to a knowledge-driven paradigm, enhancing operational efficiency and reliability while providing an extensible technical pathway and practical foundation for the intelligent development of new power systems.

2. Current State of Research

2.1. Fundamental Concepts of Knowledge Graphs

Knowledge graphs, as a pivotal achievement in contemporary AI, precisely describe concepts, entities, and their relationships through structured means. They systematically integrate and correlate vast amounts of knowledge, providing users with intelligent knowledge retrieval and reasoning services [6]. As a data modelling technique, it represents data links as a graph where nodes denote data entities and edges signify relationships between them. This transforms textual descriptions of objective facts into structured language, enabling conceptual knowledge representation that visually illustrates knowledge structures and interconnections. A specific example is illustrated in Figure 1.

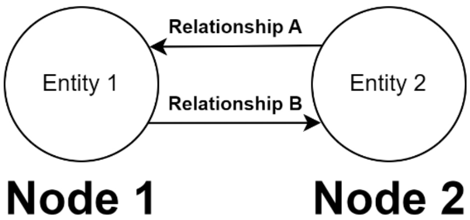

Figure 1: Example of Entity Relationships in a Knowledge Graph

Knowledge graphs are stored in graph form, with their fundamental building blocks being triples: ‘head entity, relation, tail entity’. Entities within these triples are represented by individual nodes, while the relations are depicted by graph edges and arrows. Converting data into this triplet format not only provides an effective means of describing the data but also enables its efficient storage and utilisation [7]. Consequently, as a form of graph database, knowledge graphs exhibit strong directionality and logical coherence. Constructing a knowledge graph necessitates information organisation and logical reasoning, involving the efficient consolidation and structuring of acquired data. This process entails extracting structured information—such as entities, relationships, and entity attributes—from semi-structured and unstructured data sources [8].

Knowledge graphs overcome the limitations of traditional data analysis methods in revealing complex relationships through deep association and semantic reasoning, enabling the unified management and deep utilisation of vast domain knowledge. As a key technology in the AI era, its immense potential in fields such as intelligent search and recommendation systems is garnering widespread attention from both academia and industry.

2.2. The Application of Knowledge Graphs in Wind Power Operation and Maintenance

Knowledge graphs demonstrate significant value in wind power operation and maintenance through their robust semantic expression and relational modelling capabilities. By integrating multi-source data such as SCADA, vibration monitoring, and temperature monitoring, they construct a knowledge system spanning the entire turbine lifecycle. This system provides an intuitive representation of complex relationships between equipment, faults, and maintenance strategies, thereby supporting fault diagnosis and operational decision-making.

In recent years, research in this domain has evolved from static knowledge storage towards dynamic intelligent reasoning. By incorporating advanced technologies such as graph neural networks, knowledge graphs can uncover deep patterns within operational data, enabling early fault warning and root cause identification. Compared to traditional data-driven approaches, this knowledge-based maintenance paradigm demonstrates significant advantages in terms of interpretability and adaptability to diverse scenarios.

2.3. Large Language Model Technology and Its Applications

Large Language Models (LLMs) are natural language processing models based on deep learning, pre-trained using vast amounts of textual data. Their core capability lies in understanding and generating human language. Unlike traditional rule-based natural language processing techniques, LLMs learn linguistic patterns and knowledge to automatically produce accurate and fluent language outputs across multiple tasks.

Beyond demonstrating significant effectiveness in general-purpose tasks, large language models are also expanding deeply into vertical domains[9]. For instance, domestically developed open-source LLMs such as DeepSeek and Doubao have undergone application validation in healthcare[10] (e.g., auxiliary diagnosis of ocular diseases) and agricultural science[11] (e.g., intelligent identification and management of pests and diseases), demonstrating strong domain adaptability and potential for decision support. Particularly in agricultural pest diagnosis, integrating knowledge graphs with LLM technology has further enhanced diagnostic accuracy and interpretability.

However, within the specialised context of wind power operation and maintenance, LLM applications still face critical challenges such as insufficient domain knowledge and misinterpretation of technical terminology. To overcome these limitations, an effective technical approach involves embedding wind power domain knowledge graphs as the semantic foundation for LLMs. By deeply integrating structured knowledge—such as equipment specifications, fault case studies, and maintenance protocols—into the model's reasoning process, this significantly enhances its comprehension of professional contexts and the quality of its generated outputs. This, in turn, underpins highly reliable operational decision-making and interactive analysis.

3. System Architecture and Technical Implementation

3.1. Overall Architecture Design

The system proposed in this study adopts a layered architecture design, comprising the data acquisition layer, knowledge construction layer, intelligent reasoning layer, and application service layer. The data acquisition layer integrates multi-source heterogeneous data, including real-time monitoring data, historical maintenance records, and technical documentation. The knowledge construction layer employs automated knowledge extraction techniques to derive entity relationships from unstructured text, thereby constructing a wind turbine fault knowledge graph. The intelligent reasoning layer combines the semantic comprehension capabilities of large language models (LLMs) with the structured reasoning capabilities of the knowledge graph to achieve intelligent fault diagnosis. The application service layer provides diverse services such as fault-related question answering, maintenance guidance, and decision support.

3.2. Dynamic Update Mechanism for Knowledge Graphs

To address the challenge of knowledge updating posed by the rapid iteration of wind power technology, this study designed a dynamic knowledge update mechanism incorporating large language models (LLMs). This mechanism comprises three core components: incremental information extraction based on LLMs, multi-version knowledge management, and semantic consistency verification.

When new fault reports, maintenance records or technical specifications emerge, the LLM can automatically parse unstructured text, identify entities and their relationships, and generate candidate knowledge triples. The system subsequently filters this content through manual review or confidence assessment; once confirmed, it synchronises updates to the knowledge graph. Throughout this process, the LLM continuously performs semantic-level quality checks, assisting in the identification of logical conflicts and redundant content. This effectively safeguards the accuracy, timeliness, and consistency of the knowledge system during its dynamic expansion.

3.3. Multimodal data fusion

Within the intelligent wind turbine operation and maintenance platform developed in this study, multimodal information fusion technology enables the correlation of visual inspection results with relevant textual descriptions, thereby constructing a more comprehensive representation of fault characteristics. For instance, images of damage on blade surfaces can be linked to corresponding maintenance records, material properties, and other pertinent information, providing more comprehensive data support for fault analysis.

4. LLM-based Knowledge Graph Construction

The construction of wind power system knowledge graphs encompasses processes such as ontology development, knowledge extraction, knowledge fusion, and knowledge storage [12]. The establishment of wind power system knowledge graphs commences with ontology design, forming a structured blueprint of knowledge by defining core concepts—including wind turbines, components, faults, and operations and maintenance—along with their interrelationships. Subsequently, a layered strategy combining rule matching with large language models (LLMs) extracts entities and relationships from multi-source data such as standard documentation and fault reports. Knowledge fusion is achieved through entity linking and conflict resolution, ensuring information consistency. Finally, the cleaned triples are stored in a graph database, providing precise knowledge support for fault diagnosis and intelligent operations and maintenance. The workflow diagram for constructing the wind power system knowledge graph is shown in Figure 2.

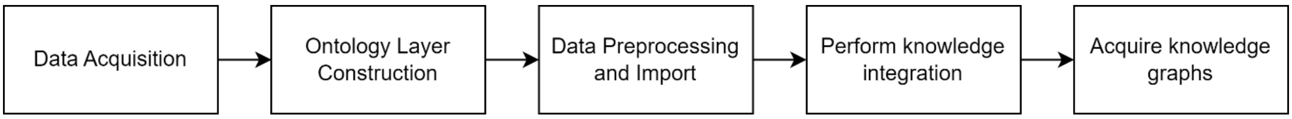

Figure 2: Process Flowchart for Constructing a Wind Power System Knowledge Graph

4.1. Acquisition of the triplet

In constructing knowledge graphs, extensive data processing of accessible textual materials is required. Traditional data processing methods fall into two primary categories: the first involves manual triplet selection, whereby researchers undertake extensive literature review to annotate unstructured text, yielding results as shown in Table 1. However, this approach yields limited data and operates with low efficiency, proving inadequate for processing vast volumes of textual information.

Table 1: Triplets following manual screening

Subject	Relationship	Object
Wind turbine	Comprises	Blades
Wind turbine	Comprises	Gearbox
Wind turbine	Comprises	Generator
Wind turbine	Comprises	Tower
Wind turbine	Comprises	Yaw system
Wind turbine	Comprises	Pitch system
Wind turbine	Comprises	Control system
Pitch Control System	Belongs to	Control system
Yaw System	Belongs to	Control system
Gearbox	Connects	Generator
Routine Maintenance	Includes	Grease lubrication
Routine Maintenance	Includes	Bolt tightening
Routine Maintenance	Includes	Filter element replacement
Fault Repair	Includes	Blade repair
Fault Repair	Includes	Gearbox replacement
Condition Monitoring	Utilises	Vibration sensors
Condition Monitoring	Utilises	Temperature sensors
Performance Analysis	Based Upon	SCADA data
Gearbox	Possible	Bearing wear
Blades	Possible	Cracks
Generator	Possible	Insulation ageing
Yaw System	Possible	Yaw deviation
Bearing Wear	Causes	Increased gearbox vibration
Cracks	Causes	Reduced aerodynamic performance of blades
Insulation Ageing	Causes	Generator short circuit
Operations and Maintenance Engineer	Responsible for	Routine maintenance
Operations and Maintenance Engineer	Operates	Wind turbine
Technical Specialist	Analyse	Fault report
Wind Farm	Possess	Wind turbine
Manufacturer	Produce	Wind turbine
Maintenance Manual	Describe	Routine maintenance procedure
Fault Code Table	Correspond	Fault types
IEC 61400 Standard	Specification	Turbine design
GB/T19072 Standard	Applicable	Wind power equipment
SCADA System	Collect	Operational data
Vibration Sensor	Upload	Vibration data
Temperature Sensor	Upload	Temperature data
Historical Data	For use in	Fault prediction

The second category comprises machine learning-based approaches that learn patterns or regularities from training data and employ machine learning algorithms to predict or infer alignment relationships between entities [13]. While such methods offer significant efficiency gains, the inherent limitations of machine learning often result in processed information exhibiting low accuracy and interpretability, failing to meet the high-precision requirements of wind power operation and maintenance.

LLMs can comprehend and execute information structuring tasks through natural language instructions, substantially reducing reliance on manual rules and custom models inherent in traditional approaches. This enhances the flexibility and efficiency of knowledge automation construction.

Consequently, this study employs LLMs to extract triplets. This involves cleaning and segmenting the original corpus to ensure semantic integrity and syntactic correctness of input text; Subsequently, based on domain knowledge or task requirements, preliminary entity recognition or keyword extraction may be necessary to help the LLM focus on core information. This ultimately constructs clear, unambiguous text segments serving as contextual foundations for prompts. Prompt engineering aims to precisely ‘program’ the triplet extraction task to the LLM via natural language instructions. Following entity recognition and linking, relation classification, and triplet assembly, outputs are strictly formatted according to prompt specifications. Owing to the inherent ‘hallucination’ tendencies and randomness of text generation in LLMs, raw outputs require rigorous post-processing. This includes parsing the model's string outputs, removing duplicates, correcting formatting errors, and ensuring each triplet is correctly resolved into three distinct elements. Subsequently, cross-validation (such as comparison with source text or generating multiple outputs under different prompts to take the intersection) or matching against existing knowledge bases eliminates false triples generated out of thin air or inconsistent with the original text. Finally, entities and relations with differing expressions but identical meanings are unified. After cleaning and normalisation, the accuracy of the new triples is significantly enhanced.

4.2. Visualisation of Triplets

The structured storage of knowledge forms the foundation for efficient querying and deep reasoning, while intuitive visualisation serves as the pivotal link in empowering systems with knowledge value. Within this system, wind turbine fault maintenance knowledge extracted, cleansed, and fused by large language models is ultimately stored as ‘entity-relationship-entity’ triples within the graph database Neo4j, accessible via the configured environment variable (PATH). This data can be output as a knowledge graph, as illustrated in Figure 3. Neo4j employs graph theory search algorithms and provides a descriptive Cypher query language [14].

The primary step in the construction process involves theoretical mapping of the data model, elevating the discrete triples output by the LLM into a semantically rich, computable property graph model. This process transcends simple format conversion, constituting deep semantic modelling. The specific procedure comprises the following three aspects:

1) Abstracting and Classifying Nodes: Entity Typing Each entity is abstracted as a graph node. To enable effective knowledge organisation and efficient retrieval, the system assigns one or more type labels to each node. This process involves semantic classification of entities—for instance, categorising gearboxes as core components, classifying excessive vibration as a fault symptom, and labelling dynamic balancing calibration as a remedial measure. The design of the tagging system directly determines the ontological structure of the knowledge graph, forming the syntactic foundation for subsequent querying and reasoning.

2) Semantic Definition of Relationships: Refining Relationships: Binary relationships between entities are concretised as directed edges in the graph, each assigned a precise relationship type. This step aims to eliminate ambiguities inherent in natural language, constructing machine-understandable semantic links. For instance, a gearbox exhibiting vibration anomalies is mapped as an edge from the gearbox node to the vibration anomaly node, with the edge type designated as EXHIBITS. Standardised relationship type design is pivotal to ensuring knowledge graph quality.

3) Property Attachment and Enrichment Knowledge Deepening: Both nodes and relations can bear property key-value pairs to characterise their intrinsic features and contextual information. This evolves the knowledge model from a simple association network into an information-rich property graph. For instance, a fault cause node may attach properties such as confidence level and historical occurrence frequency; a maintenance relation may attach properties such as required tools, standard operation time, and safety level. This process significantly enhances the depth and practicality of the knowledge.

Given that Neo4j's graph database query language exhibits greater efficiency in querying nodes than edges [15], it proves better suited to subsequent model training processes in this research. Neo4j Output of Wind Turbine Failure System Knowledge Graph is shown in figure 3.

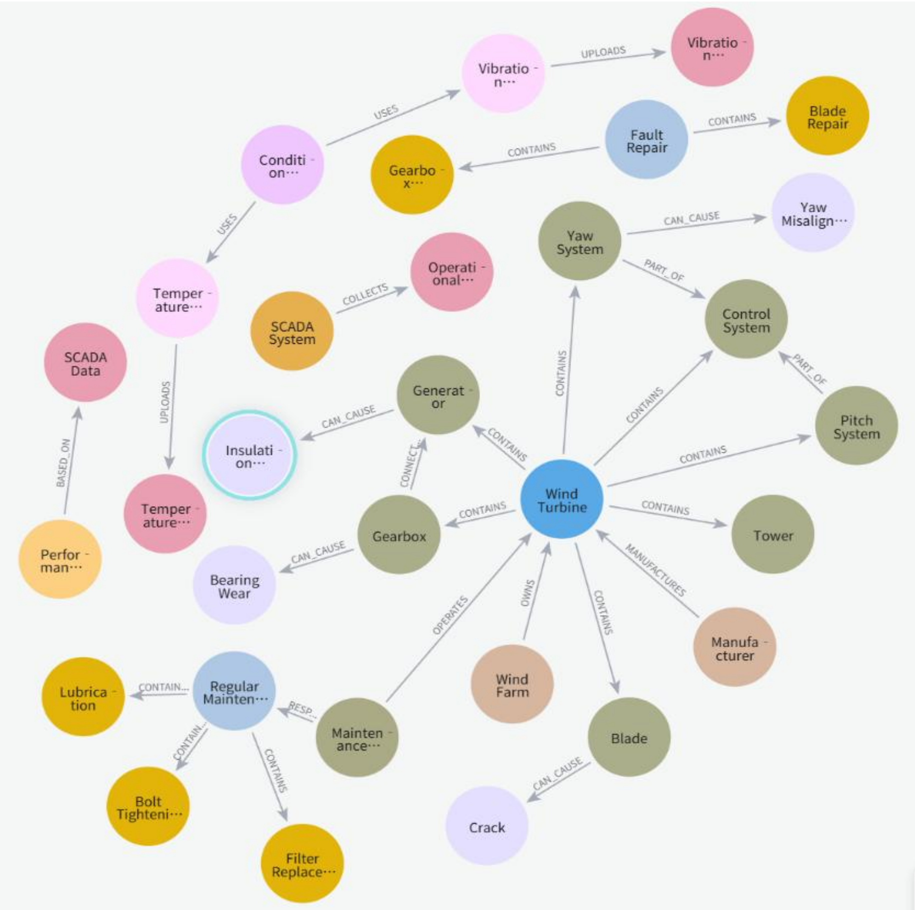

Figure 3: Neo4j Output of Wind Turbine Failure System Knowledge Graph

5. Research Findings and Discussion

In earlier research, scholars including Xu Yan et al. [16] employed a layered collaborative knowledge extraction strategy for the original corpus, applying distinct and optimally suited technical methods to knowledge sources of varying granularity (document-level, sentence-level, semantic-level). The core of this strategy lies in the principle of 'divide and conquer'. By stratifying knowledge according to granularity and selecting the most suitable and efficient technical tools for each layer (ranging from rule-based regular expression matching, through structure-based parsing for clause segmentation, to deep learning-based semantic understanding), it achieves collaborative, precise, and efficient knowledge extraction from macro-level document structures down to micro-level semantic entities.

In processing the raw corpus, this research initially relied primarily on manual data annotation. This effort annotated nearly 500 valid entity relationships, with a subset used to construct the knowledge graph sample. As the corpus size expanded, this approach proved inadequate for annotating the vast dataset. To address this challenge, the research shifted towards exploring large language models to assist in automated and semi-automated entity and relationship extraction.

Specifically, we first employed the high-quality manually annotated data as a seed set. This enabled targeted prompt engineering optimisation and exemplar learning for the large language model, equipping it to comprehend domain-specific entity types and relationship definitions. Building upon this foundation, an iterative human-machine collaborative optimisation strategy was adopted: the model generated preliminary predictions on unannotated corpora, while researchers conducted sampling verification and error correction on these predictions. The corrected data was then fed back to the model for further refinement. This process effectively establishes a self-improving closed-loop system.

By incorporating large language models, this research significantly enhances data processing scale and overall research efficiency while maintaining annotation quality. It lays a reliable foundation for constructing large-scale, high-quality knowledge graphs and provides a reference pathway for automated information extraction from massive text volumes in similar studies.

6. Conclusion

This study systematically explores the conceptual framework for constructing a wind turbine fault maintenance decision support system based on knowledge graphs and large language models, alongside the implementation pathways for knowledge graphs. By integrating the structured knowledge representation capabilities of knowledge graphs with the robust natural language understanding capabilities of LLMs, a maintenance system endowed with intelligent reasoning and decision support functions is proposed.

However, the system demonstrates limitations in handling complex fault reasoning, particularly in diagnosing and analysing multi-system coupled faults, where its reasoning mechanisms and decision-making logic necessitate further refinement and optimisation.

Future research will focus on the following key areas: Firstly, innovation in deep reasoning mechanisms, combining the strengths of symbolic reasoning and neural network reasoning to enhance the system's reasoning capabilities in complex scenarios. Secondly, cross-wind farm knowledge transfer, investigating how operational experience from a single wind farm can be effectively generalised to other sites, addressing application challenges in data-scarce environments. Thirdly, real-time optimisation, improving knowledge updating and reasoning efficiency to meet the real-time demands of online fault diagnosis. Finally, the introduction of trustworthy artificial intelligence technologies will enhance system reliability and transparency through uncertainty quantification and explainability analysis.

References

[1] Sukanta R, Shawli B, Yogesh K, et al. Recent technology and challenges of wind energy generation: A review[J]. Sustainable Energy Technologies and Assessments, 2022, 52(PC):

[2] Chen, Yueping. Research on Intelligent Operation and Maintenance of Wind Power Based on Knowledge Graph Technology [Doctoral Dissertation]. Xiangtan University, 2021. DOI: 10.27426/d.

[3] Liu Xinliang, Xu Yusi, Li Dubai, et al. A Review of Knowledge Graph-Based Question-Answering Methods [J/OL]. Computer Applications, 1-21 [2025-11-19].

[4] Chen Hong, Chen Xincai, Gong Xiaoyun, et al. Construction and Application of a Wind Turbine Diagnostic System Based on Knowledge Graphs [J]. Journal of Zhengzhou University (Engineering Science), 2023, 44(06): 54-60+98.

[5] Zhao Yan. Fault Diagnosis of Wind Turbine Planetary Gearboxes Based on Knowledge Graphs and GPT Models [J]. Small Electric Machines, 2025, 53(04): 78-82.

[6] Qu K, Li C K, Wong M T B, et al. A Survey of Knowledge Graph Approaches and Applications in Education[J].Electronics,2024,13(13):2537-2537.

[7] Liu Xinliang, Xu Yusi, Li Dubai, et al. A Review of Knowledge Graph-Based Question-Answering Methods [J/OL]. Computer Applications, 1-21 [2025-11-20].

[8] Miao Lin, Wu Yu, Liu Xuhong, et al. A Review of Knowledge Graph Information Extraction Techniques [J]. Computer Simulation, 2025, 42(09): 279–285+338.

[9] Zheng Yining, Yu Zhen, Li Bu-fan, et al. A Review of Tool Usage in Large Language Models [J/OL]. Acta Automatica Sinica, 1–16 [20 November 2025].

[10] Guo Yuyu, Zhang Yue, Xu Yanhui, et al. Application of Large Language Models in Health Education for Paediatric Cataract Patients [J/OL]. Journal of Central South University (Medical Edition), 1–11 [20 November 2025].

[11] Wu, H. Y., Chen, M., & Guo, Q. H. (2025). An expert system for rice pests and diseases based on large language models and knowledge graphs. Transactions of the Chinese Society for Agricultural Engineering, 1–12. [2025-11-20].

[12] Shi Zhiyuan, Zhang Jialei, Kong Zhiwei, et al. Intelligent Question-Answering System for Wind Power Equipment Integrating Knowledge Graphs and Large Language Models [J]. Dongfang Electric Review, 2024, 38(03): 77-84.

[13] Cong Shuo, Su Guibin, Liu Lin, et al. A Review of Research on Entity Alignment in Knowledge Graphs: From Traditional Methods to Cutting-Edge Techniques [J/OL]. Computer Engineering and Applications, 1–26 [2025-11-20].

[14] Jiang Wei, Wang Minghua, Chen Jinming, et al. Calculation of Power Supply Reliability in Distribution Networks Based on Neo4j Graph Database. Automation of Electric Power Systems, 2022, 46(15): 104-111.

[15] HU S, ZOU L, XU Y J, et al. Answering natural language questions by subgraph matching over knowledge graphs [J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(5): 824-837.

[16] Xu Yan, Wang Tao, Yuan Yang, et al. Research on Multi-granularity Knowledge Graph Construction and Intelligent Question-Answering Applications in Power Design[J/OL]. Journal of Civil Engineering and Management, 1-9[2025-11-23].