Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 14, No. 1, 2025

Review on Construction and Application of Domain Knowledge Graphs: Cross-Domain Practices and Cutting-Edge Explorations Driven by Large Language Models

Zicheng Wang *, Guochen Yang

Department of information network security, People's Public Security University of China, Beijing, China

*** Corresponding author:** Zicheng Wang (Email: 18107078848@163.com)

Abstract: Knowledge graphs (KGs) are core infrastructure for artificial intelligence, and large language models (LLMs) have significantly advanced their construction and application. This paper systematically reviews the latest domestic and international research on domain KG construction. Domestically, studies focus on low-resource adaptation and prompt engineering in vertical fields (agriculture, healthcare, education, etc.), such as few-shot extraction via LoRA and zero-shot entity recognition with DeepSeek. Internationally, research emphasizes general frameworks in mineral resources, smart manufacturing, and other areas, including multimodal fusion with Cross-Modal Transformers and quintuple-based knowledge representation. It further analyzes technological evolution and existing challenges. Future directions include deepening LLM-KG synergy and enhancing interpretability. This study provides a systematic reference for intelligent knowledge management.

Keywords: Domain Knowledge Graphs; Large Language Models; Multimodal Fusion; Cross-Domain Application.

1. Introduction

As a crucial technology in the field of artificial intelligence, KGs describe concepts, entities, and their relationships in the real world in a structured manner, providing a fundamental framework for machines to understand semantics and achieve intelligent reasoning. In recent years, with breakthroughs in technologies such as natural language processing and large language models (LLMs), the construction methods and application scenarios of KGs have been significantly expanded—shifting from initial general domains to gradually penetrating vertical domains including agriculture, healthcare, education, and historical culture. Based on the latest domestic and international research literature, this paper systematically sorts out the technological evolution, cross-domain application practices, and future development trends of KG construction. It focuses on exploring the breakthroughs of KGs in construction efficiency, semantic understanding, and reasoning capabilities under the empowerment of LLMs, aiming to provide systematic references for research in related fields.

2. Research Progress on Knowledge Graph Construction at Home and Abroad

2.1. Research Progress on Domestic Knowledge Graph Construction

2.1.1. Agriculture and Animal Husbandry

In the agricultural field, KG construction is closely aligned with practical industrial needs, and specialized solutions are proposed based on the characteristics of domain-specific data. To address issues such as ambiguous semantic boundaries and overlapping entity roles in sheep disease-related texts, Zhang [1] put forward a KG construction method based on the CaRoMHPE model. This method takes the CasRel model as the foundation, adopts the RoBERTa-wwm-ext pre-trained model to enhance contextual semantic understanding, refines the semantic relationships between entities via a multi-scale cross-attention mechanism, and introduces hybrid position encoding to optimize entity boundary division in multi-head attention. Eventually, in the sheep disease knowledge extraction task, the method achieved an accuracy of 94.70%, a recall rate of 94.04%, and an F1-score of 94.37%—an improvement of approximately 9 percentage points compared to the baseline CasRel model—providing a high-quality knowledge base for intelligent diagnosis and treatment systems.

Focusing on the entire pig industry chain, Shi[2] proposed a low-cost domain KG construction framework empowered by the DeepSeek LLM. This study adopts a three-tier architecture consisting of ontology modeling, multi-source data fusion, and intelligent extraction to define 21 categories of core entities and their attribute relationships. It innovatively integrates DeepSeek with knowledge extraction tasks, enhances semantic understanding, and applies prompt engineering techniques, ultimately achieving an F1-score of 0.92 for entity recognition in pig disease prevention scenarios under zero-shot learning conditions.

2.1.2. Healthcare Field

Research on healthcare KGs focuses on addressing challenges such as strong data heterogeneity and complex professional terminology, forming a complete technical chain from knowledge extraction to application implementation. Tan [3] constructed a knowledge graph for spleen-stomach diseases using the BiLSTM-CRF model combined with the Neo4j database. Knowledge extraction was completed through a combination of the BIO sequence labeling method and manual rules. On the test set, the model achieved an accuracy of 96.19% and an F1-score of 87.71%, successfully supporting the query and relationship discovery of nodes such as TCM syndromes and treatment principles, and providing an effective solution for the structured expression of TCM knowledge.

To tackle multi-source heterogeneous healthcare data, Meng [4] proposed a knowledge extraction framework based on the MC-BERT-CasRel model. This study takes the Unified Medical Language System (UMLS) medical terminology as the basis, uses the BTM (Biterm Topic Model) to assist ontology design through topic clustering of unstructured data, selects the optimal model for entity-relationship triple extraction by comparing the embedding effects of BERT variant models, and completes data fusion by combining the SapBERT model with the Levenshtein edit distance algorithm. Finally, a knowledge graph for digestive system diseases containing 10,010 entities and 29,044 relationships was constructed, which has been applied in clinical knowledge retrieval.

2.1.3. Education and Culture Field

KG research in the education field breaks through the traditional static knowledge representation model, focusing on the dynamic process of knowledge construction and in-depth integration with educational scenarios. Yang[5] designed a KG construction framework based on collaborative knowledge construction conversations. Through case analysis of real collaborative scenarios, they identified three group knowledge production paths as well as four group conversation development trends. Focusing on the promotion of traditional cultural reading in university libraries, Wang [6] proposed an LM-based KG construction process. This study extracts entity categories using the BERTopic model, completes entity-relationship extraction with the DeepSeek-R1 model, and ultimately realizes the visual display of knowledge elements in the Neo4j database.

2.1.4. Industrial and Engineering Fields

KG research in the industrial and engineering fields closely addresses industry pain points, realizing the value transformation from data to decision-making in scenarios such as industrial chain analysis and construction safety. To address the difficulty of extracting supply-demand relationships in the industrial chain, Zheng[7] proposed an LLM verification-enhanced industrial chain KG construction method. It designs a "industry-enterprise" dual-perspective hierarchical system and uses the VRTE-LLM framework to fine-tune the LLM with an industrial chain annotation corpus, correcting the completeness and bias errors of relational triples. In entity recognition and relationship extraction tasks, the method achieved accuracies of 80.9% and 83.9% respectively, and the constructed KG containing 39,627 triples has supported quantitative analysis and trend prediction of the industrial chain.

To solve the problem of multimodal data integration in construction safety management, Dong[8] proposed a multimodal LLM-based KG construction method. This study constructs a construction safety knowledge ontology, realizes the fusion extraction of text-based accident knowledge and on-site image information through multimodal LLMs, and achieves high accuracy in image content understanding and structuring tasks. The constructed multimodal KG significantly improves the comprehensiveness of safety hazard identification on construction sites, provides visualized knowledge support for management decisions, and verifies the application value of multimodal knowledge representation in the engineering field.

2.1.5. Historical Culture and Social Sciences Fields

KG research in the historical culture field is committed to breaking through the limitations of traditional narratives, realizing the systematic organization and in-depth excavation of historical resources. Zhang[9] proposed a method for constructing a knowledge graph of red figures of the Northeast Anti-Japanese United Army based on LLMs and LoRA. By designing information extraction prompt templates and constructing an instruction dataset, they performed few-shot LoRA fine-tuning on LLMs such as Qwen2.5-7B. When the number of training samples was 100, the precision of joint entity-relationship extraction reached 95.10%, successfully constructing a knowledge graph containing the life stories, deeds, and related events of red figures, which provides a technical paradigm for the digital preservation of red cultural resources under low-resource conditions. From an event-driven perspective, Yang[10] constructed an anti-Japanese war KG. They identified temporal, causal, and inclusive relationships among 1,022 anti-Japanese war events, completed entity extraction and relationship recognition with LLMs, and finally realized the visual storage and intelligent question-answering application of historical events in the Neo4j database.

2.2. Research Progress on International Knowledge Graph Construction

2.2.1. Mineral Resources Field

To address the difficulty in identifying theoretical components for porphyry copper deposit (PCD) exploration, Mo[11] proposed a mineralization system modeling method based on KGs and knowledge discovery. Using global PCD geological reports and exploration data as data sources, this study extracts entities such as "mineralized intrusions," "mineralization zones," and "ore-controlling structures" through the Universal Information Extraction (UIE) framework, and models the relationships of spatial entities in geological maps using ArcGIS Pro software. The constructed KG contains more than 100,000 association relationships of mineralization elements.

Tian[12] proposed a construction method for the Polymetallic Mineral Resources Knowledge Graph. Taking mineral geological survey reports, geological maps, and remote sensing images as data sources, this study extracts text-based entities through the UIE framework and uses remote sensing image interpretation technology to extract spatial entities such as "landform types" and "vegetation coverage." In the knowledge fusion stage, it realizes the association between text entities and spatial entities through geographic coordinate alignment, and designs KG applications from four dimensions: mineralization element analysis, semantic understanding, intelligent prediction, and decision support.

2.2.2. Smart Manufacturing Field

KG research in the smart manufacturing field focuses on process knowledge reuse, intelligent equipment operation and maintenance, and dynamic knowledge management, forming technical characteristics of multimodal fusion and incremental learning. To solve the problem of difficult multimodal data association in intelligent process planning, Xu[13] proposed a method for constructing a Multimodal Process Knowledge Graph based on Cross-Modal Transformers. Taking the mechanical machining process as the research object, this study developed the MF36 dataset containing 36 machining features, covering three modalities:

3D models, engineering drawings, and text descriptions. It designs a cross-modal framework: the LERT-CRF model is used to extract text-based entities such as "machining methods" and "tool types," while the PA-ViT model (with a pooling attention mechanism) extracts visual entities such as "geometric features" and "dimensional tolerances" from 3D models and engineering drawings. Through Cross-Modal Transformers, the alignment of text and visual features is realized, constructing cross-modal relationships such as "machining features-corresponding-machining methods" and "geometric features-constraining-process parameters." Experiments show that this KG improves the process planning efficiency of complex parts by 40% and achieves a multimodal knowledge fusion accuracy of 91.2%.

To address the continuous accumulation of manufacturing knowledge, Jing[14] proposed a manufacturing knowledge graph construction method based on meta-learning. Based on the manufacturing knowledge ontology (including entities such as "products," "processes," and "equipment"), a two-stage knowledge extraction model is designed: the first stage trains a "domain-general extractor" through meta-learning to acquire the commonalities of knowledge extraction across different manufacturing scenarios; the second stage fine-tunes the model with a small number of labeled samples for newly accumulated manufacturing knowledge, realizing low-resource entity recognition. Breaking through the limitations of traditional triple-based knowledge representation, Shao[15] proposed a quintuple-based (Entity 1, Relationship, Entity 2, Attribute, Context) assembly KG construction framework. To address the lack of contextual information in assembly processes, this study uses LLMs to process implicit relationships in assembly texts (e.g., the "tightening torque" attribute and "assembly sequence" context of "bolt-connecting-flange"). It decomposes assembly KG construction into five tasks: entity extraction, relationship recognition, attribute completion, context modeling, and knowledge fusion, which are implemented using LLM fine-tuning, rule matching, and context reasoning respectively. The constructed KG contains assembly knowledge of complex products such as aero-engines and automobile chassis, and its process reuse module matches similar assembly cases through the KNN algorithm, shortening the assembly process planning time by 45%.

2.2.3. Healthcare Field

International research on healthcare KGs focuses on large-scale clinical evidence integration and knowledge modeling of complex diseases such as mental illnesses, using LLMs to improve the scalability and accuracy of knowledge extraction. To solve the problem of fragmented evidence in mental illness research, Gao[16] constructed a Mental Disorders Knowledge Graph based on LLMs. This study integrates mental illness data from PubMed literature, clinical databases (e.g., UK Biobank), and electronic health records, extracts entities such as "disease types," "symptoms," and "risk factors" through LLMs, and constructs relationships such as "disease-accompanying-symptoms" and "risk factors-causing-diseases." MDKG contains more than 10 million relationships, among which nearly 1 million are novel associations not found in existing databases (e.g., DisGeNET). Experiments show that the prediction model enhanced by MDKG improves the AUC value of mental illness risk prediction by 18% and shortens the expert verification time by 70%.

2.2.4. Urban and Crisis Management Field

KG research in the urban and crisis management field focuses on using LLMs to improve knowledge extraction from unstructured texts. To address the difficulty in mining association patterns of compound urban crises (e.g., natural disasters overlapping with public health events), Hao[17] proposed an LLM-based KG construction method. Using 1,941 urban crisis news reports from Shenzhen as data sources, this study enhances the domain adaptability of LLMs through Retrieval-Augmented Generation (RAG) technology, extracts entities such as "crisis types," "impact scopes," and "response measures," and identifies compound relationships such as "crisis-triggering-crisis" (e.g., "heavy rain-leading to-waterlogging-causing-traffic disruption"). The constructed KG contains an association network of 13 types of urban crises (natural disasters, public health, technical failures, etc.), and typical compound chains such as "extreme weather-infrastructure failure-public service disruption" are identified through community detection algorithms. Experiments show that this KG achieves an 82% accuracy in compound crisis prediction, providing decision support for "early warning-response" of urban crises.To address the problem of "fragmented cybersecurity threat reports," Wang[18] proposed a cybersecurity KG construction method (SC-LKM) based on semantic chunking and LLMs. Breaking through the limitation of "fixed-length chunking" in traditional GraphRAG, this study adopts hierarchical semantic chunking technology to split threat reports into "report chapters-topic paragraphs-key sentences," preserving the semantic integrity of the text. Through LLMs, entities such as "attack types," "vulnerability numbers," and "affected devices" are extracted, and relationships such as "attack-exploiting-vulnerability" and "vulnerability-impacting-device" are constructed. A predefined cybersecurity rule base (e.g., the MITRE ATT&CK framework) is introduced to verify the rationality of triples, eliminating issues such as "attack type misjudgment" and "vulnerability association errors." The constructed KG shortens the traceability time of cybersecurity incidents by 50% and improves the utilization rate of threat intelligence by 45%.

2.2.5. Transportation and Geographic Information Field

KG research in the transportation and geographic information field focuses on "spatial relationship modeling" and "traffic entity prediction," improving the spatiotemporal awareness of KGs through high-definition (HD) maps and co-occurrence relationships. To address the insufficient coverage of long-tail scenarios in traffic entity prediction for autonomous driving, Yin[19] proposed a KG construction method based on HD maps and co-occurrence relationships. This study adopts a "top-down" construction strategy, taking concepts such as "traffic signs," "lane lines," and "intersections" defined in HD maps as the ontological foundation. It models entity associations through a co-occurrence semantic network (e.g., "parking signs-often appearing at-residential area intersections" and "bus lanes-adjacent to-bus stops") and uses prior knowledge to make up for the deficiencies of labeled datasets. In the traffic entity prediction module, by mining association rules such as "seeing traffic lights-predicting-intersections ahead," the prediction accuracy of long-tail scenarios (e.g., traffic entity recognition under special weather conditions) is improved by 27%, providing a theoretical basis for the "safety calibration" of autonomous driving.

To meet the needs of "knowledge visualization and automated construction" in virtual geographic scenarios, Guo[20] proposed a 3D KG construction method based on multi-agent systems. A multi-agent system is designed: the "entity extraction agent" identifies entities such as "buildings," "roads," and "vegetation" in virtual geographic scenarios (e.g., 3D urban models); the "relationship modeling agent" constructs spatiotemporal relationships such as "spatial adjacency," "inclusion," and "orientation"; the "visualization agent" realizes the 3D display of the KG. Entity relationship modeling is optimized through spatiotemporal constraint rules (e.g., "buildings-located on-both sides of roads"). Experiments conducted on a virtual scenario containing 2,183 objects show that the constructed 3D KG improves the cognitive efficiency of geographic knowledge by 40%, providing support for "intelligent interaction" in virtual geographic environments.

3. Technological Evolution of Knowledge Graph Construction Driven by Large Language Models

3.1. Multimodal Knowledge Fusion

Multimodal knowledge fusion (integrating text, images, 3D models, logs, etc.) has become a key technology in domain KG construction. Domestic research focuses on scenario-specific adaptation, while international research emphasizes the design of general frameworks. Domestically, Wang[21] proposed a bimodal knowledge extraction framework that integrates device manual information (text + images) and user behavior logs, realizing the association between images and text through relative position layout matching (RPLM) to construct a spatiotemporal KG for smart homes, with an accuracy of 96.39%; Dong[8] integrated text-based accident knowledge and on-site images in their construction safety KG, improving the comprehensiveness of hazard identification. Internationally, Xu[13] realized the joint feature extraction of 3D models, engineering drawings, and text with Cross-Modal Transformers; Yang[22] improved the accuracy of equipment image recognition in their LNG carrier PSC (Port State Control) inspection KG through an optimized visual prompt tuning (EVPT) model.

3.2. LLM Empowerment Drives Paradigm Innovation in Knowledge Extraction

Equipped with semantic understanding and contextual modeling capabilities, Large Language Models (LLMs) break through the reliance of traditional methods on manual annotation. Domestic research focuses on low-resource adaptation and prompt engineering, while international research emphasizes general capability and knowledge discovery. Domestically, examples include: achieving few-shot knowledge extraction via LoRA (Low-Rank Adaptation) fine-tuning technology, and realizing zero-shot entity recognition using the DeepSeek model. Internationally, efforts include integrating knowledge via LLMs to discover nearly new associations[23], and leveraging LLMs to assist in quintuple-based knowledge extraction, thereby enhancing the integrity of complex knowledge representation.

4. Challenges and Prospects of Knowledge Graph Construction and Application.

4.1. Analysis of Existing Challenges

Despite significant progress in KGs empowered by LLMs, numerous challenges remain. Technically, the accuracy of knowledge extraction still needs to be improved in complex scenarios; in terms of knowledge reasoning, the "black-box problem" of LLMs remains unresolved. Application-wise, domain adaptation costs are high—although few-shot learning has alleviated this issue to some extent, a certain amount of labeled data is still required. Additionally, data privacy and security are particularly prominent in fields such as healthcare and industry, and the protection of sensitive data requires attention.

4.2. Future Research Directions

Future research can be carried out in the following directions: First, deepen the synergy mechanism between LLMs and KGs, explore technologies such as dynamic knowledge injection and continuous learning, and realize bidirectional enhancement between KGs and LLMs. Second, strengthen multimodal knowledge fusion and reasoning, develop technologies such as cross-modal feature alignment and visual semantic understanding, and expand more application scenarios by drawing on existing multimodal modeling experience. Third, improve the interpretability of KGs, construct transparent reasoning paths, and address the demand for interpretability in relational reasoning. Fourth, explore few-shot and zero-shot KG construction technologies to reduce domain adaptation costs.

4.3. Prospects for Interdisciplinary Applications

The future development of KGs will place greater emphasis on interdisciplinary integration. In the agricultural field, combining IoT sensor data with KGs can realize precision planting decisions; in the healthcare field, integrating multimodal KGs with precision medicine is expected to promote the generation of personalized treatment plans and expand application scenarios; in the education field, combining KGs with learning analytics can achieve intelligent learning situation diagnosis. Interdisciplinary applications will transform KGs from knowledge representation tools to the core of intelligent decision-making.

5. Conclusion

This paper systematically sorts out the latest research progress in KG construction and application at home and abroad, focusing on analyzing the technological innovations and cross-domain practices driven by LLMs. Domestic research has demonstrated remarkable application innovations in fields such as agriculture, healthcare, and education, while international research shows greater foresight in terms of technical generality and cross-domain integration. The introduction of LLMs has promoted the comprehensive upgrading of technologies such as knowledge extraction, fusion, and reasoning, but challenges such as accuracy, adaptation costs, and privacy security still exist. Future research needs to further deepen the synergy between LLMs and KGs, strengthen multimodal processing and interdisciplinary applications, and provide more powerful technical support for knowledge management and decision-making in the intelligent era.

References

[1] Zhang Z J, Sun X H, Wang C, et al. A method for constructing sheep disease knowledge graph based on CaRoMHPE [J]. Transactions of the Chinese Society of Agricultural Engineering, 2023.

[2] Shi Z Y, Lei J, Sun T, et al. Research on low-cost construction of domain knowledge graphs empowered by DeepSeek [J]. Journal of Library and Information Sciences in Agriculture, 2025, 37 (03): 4-17.

[3] Tan P, Liu H N, Wei C F. Construction of knowledge graph for spleen-stomach diseases based on BiLSTM-CRF and Neo4j [J]. Science Technology and Engineering, 2025, 25 (22): 9436-9444.

[4] Meng Q Q, Zheng M R, Tian Y L, et al. Research on knowledge graph construction for multi-source heterogeneous healthcare data [J]. Information Science, 2025.

[5] Yang Y, Lai S N. Research on knowledge graph based on collaborative knowledge construction conversations [J]. Modern Educational Technology, 2025, 35 (08): 97-106.

[6] Wang L L. Construction of knowledge graph for traditional cultural reading promotion in university libraries based on large language models and its optimization path [J]. Information Science, 2024, 42 (12): 130-140.

[7] Zheng A Z, Zhang K L, Li Y L, et al. Research on industrial chain knowledge graph construction based on large model verification enhancement [J]. Journal of Frontiers of Computer Science and Technology, 2023.

[8] Dong L, Wu F J, Shi J Y, et al. Construction and application of multimodal knowledge graph for construction safety based on large language models [J]. Computer Engineering and Applications, 2025, 61 (09): 325-333.

[9] Zhang W D, Chen X P, Weng H Q. Research on construction method of knowledge graph for red figures of the Northeast Anti-Japanese United Army based on large models and LoRA [J]. Information Studies: Theory & Application, 2025, 48 (08): 187-195.

[10] Yang R C, Wang D. Research on construction and application of anti-Japanese war knowledge graph based on event-driven approach [J]. Information Science, 2025, 43 (04): 117-128.

[11] Mo L Z, Zhang Z J, Yang J, et al. Knowledge graph construction and knowledge discovery for porphyry copper deposits [J]. Ore Geology Reviews, 2025, 186.

[12] Tian M, Xie Z, Qiu Q, et al. PMinrKG: polymetallic mineral resources knowledge graph construction and its applications integrated with multimodal data [J]. International Journal of Digital Earth, 2025, 18 (1).

[13] Xu Q, Zhang C, Ma D, et al. Automated multimodal process knowledge graph construction for intelligent process planning with Cross-Modal Transformers [J]. Robotics and Computer-Integrated Manufacturing, 2026, 98.

[14] Jing Y, Zhou G, Zhang C, et al. Knowledge graph construction with meta-learning for continuously accumulated manufacturing knowledge [J]. Computers in Industry, 2025, 172.

[15] Shao P, Huang Z, Qiao L, et al. A novel assembly knowledge graph construction framework enhanced by large language model [J]. The International Journal of Advanced Manufacturing Technology, 2025, 140 (3-4).

[16] Gao S, Yu K, Yang Y, et al. Large language model powered knowledge graph construction for mental health exploration [J]. Nature Communications, 2025, 16 (1).

[17] Hao H, Chen X, Chen Y, et al. Uncovering compound urban crises with large language model-assisted knowledge graph construction [J]. International Journal of Disaster Risk Reduction, 2025, 127.

[18] Wang P, Zhang Y, Zhou Z, et al. SC-LKM: A Semantic Chunking and Large Language Model-Based Cybersecurity Knowledge Graph Construction Method [J]. Electronics, 2025, 14 (14).

[19] Yin Z, Chen Y, Gu J, et al. A knowledge graph construction method based on co-occurrence for traffic entity prediction [J]. International Journal of Applied Earth Observation and Geoinformation, 2025, 142.

[20] Guo Y, Zhu J, Wu J, et al. An automated construction method of 3D knowledge graph based on multi-agent systems in virtual geographic scene [J]. International Journal of Digital Earth, 2025, 18 (1).

[21] Wang F, Tao Y, Liu J W, et al. A bimodal fusion construction method for spatiotemporal knowledge graph in smart home space [J]. Journal of Computer Applications, 2025.

[22] Yang Q, Gan L, Mao Q, et al. Construction of a multimodal knowledge graph for LNG carrier port state control inspections based on improved visual prompt tuning [J]. Ocean Engineering, 2025, 339.

[23] Zhang E, Zhang S W, Jia J L. A dynamically updatable knowledge graph construction method for computer-aided process planning and design [J]. Journal of Engineering Design, 2025, 36 (7-9).