Advanced Engineering Informatics 66 (2025) 103468

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

A knowledge graph-enhanced large language model for question answering of hydraulic structure safety management

Dongliang Zhang [a,b], Gang Ma [a,b,c,*], Tongming Qu [a,b,d], Xudong Wang [b], Wei Zhou [a,b,c], Xiaomao Wang [a,b,e,*]

[a] Institute of Water Engineering Sciences, Wuhan University, Wuhan 430072, China
[b] State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan 430072, China
[c] Key Laboratory of Rock Mechanics in Hydraulic Structural Engineering of Ministry of Education, Wuhan University, Wuhan 430072, China
[d] Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong SAR, China
[e] CISPDR Corporation, Wuhan 430010, China

ARTICLE INFO

Keywords:
Hydraulic structure safety management
Domain knowledge QA
Intent parsing
Knowledge graph
Large language model

ABSTRACT

Early detection and mitigation of hazards in hydraulic structures are crucial for effectively reducing economic and life losses. However, traditional hydraulic structure safety management methods rely on error-prone individual experience and emergency manuals, which are insufficient for making timely, scientifically informed decisions during crises. To address this challenge, this study presents an AI-driven framework for hydraulic structure safety management based on knowledge-based question answering. First, an ontology model was developed through a detailed analysis of safety management texts. Next, a partition fusion Kolmogorov-arnold network (PFKAN) enhanced with attention mechanisms was designed to jointly extract entities and relational knowledge. A safety management knowledge graph (KG) was then constructed from this knowledge. Subsequently, a large language model (LLM) was employed with a voting strategy to interpret query intent and extract relevant domain-specific knowledge from the KG. Finally, domain knowledge was integrated into the LLM to generate professional responses. Experimental results show that the F1 scores for entity and relation extraction with PFKAN reached 0.91 and 0.90, respectively, and the F1 score for query intent parsing with the voting strategy was 0.95, demonstrating competitive performance. The KG-enhanced LLM significantly improves decision-making in hydraulic structure safety management, providing an accurate and scalable tool for engineering safety managers.

1. Introduction

Hydraulic structures are water infrastructures built on river channels to divert and retain water. By providing services such as flood control, water diversion, irrigation, and power generation, they have made significant contributions to the economic development of many countries. The safety management of hydraulic structures is a critical component in ensuring the safe and stable operation of water infrastructure. However, most hydraulic structures are predominantly composed of earth and rock materials. Long-serving structures may experience intermittent safety incidents that threaten lives and property [1]. A report by the Association of State Dam Safety Officials noted 173 large dam failures and 587 small-scale incidents in the United States between 2005 and 2013 [2]. In the event of hydraulic structure failure, floodwaters can spread rapidly and uncontrollably, posing severe threats to public safety and causing substantial economic damage. With the increasing frequency of extreme precipitation events driven by global climate change, effective hydraulic structure safety management has become

Abbreviations: **BERT**, Bidirectional Encoder Representations from Transformers; **BiLSTM**, Bidirectional Long Short-term Memory; **BIM**, Building Information Modeling; **CBAM**, Convolutional Block Attention Module; **CNN**, Convolutional Neural Network; **CRF**, Conditional Random Field; **GPT**, Generative Pre-trained Transformer; **KAN**, Kolmogorov-arnold Network; **KG**, Knowledge Graph; **LLM**, Large Language Model; **LSM-KG**, Levee Safety Management Knowledge Graph; **NER**, Named Entity Recognition; **NLP**, Natural Language Processing; **OCR**, Optical Character Recognition; **PFKAN**, Partition Fusion Kolmogorov-arnold Network; **QA**, Question Answering; **RAG**, Retrieval-augmented Generation; **RE**, Relation Extraction; **RNN**, Recurrent Neural Network; **SL**, Supervised Learning; **SVM**, Support Vector Machines.

* Corresponding authors at: Institute of Water Engineering Sciences, Wuhan University, Wuhan 430072, China.
E-mail addresses: magang630@whu.edu.cn (G. Ma), wangxiaomao@cjwsjy.com.cn (X. Wang).

https://doi.org/10.1016/j.aei.2025.103468
Received 12 January 2025; Received in revised form 12 March 2025; Accepted 12 May 2025
Available online 19 May 2025
1474-0346/

increasingly critical. Traditional hydraulic structure safety management has long relied on human experience. However, human error in decision-making is one of the three primary causes of structural failure [3]. For example, in 1964, the Gibson Dam in the United States failed to fully open the gate due to operational errors, resulting in nearly 1 m of floodwater overflowing the dam for 20 h before it collapsed. Similarly, in 1984, the Dibis Dam in Iraq failed because operators neglected to open the spillway in time, resulting in a breach that severely threatened downstream safety and property [4].

Over the years, hydraulic engineers and researchers have accumulated a vast body of safety management knowledge documented in books, manuals, papers, and online resources. However, this knowledge is dispersed across various platforms and formats, making it difficult to retrieve and apply promptly during emergencies, which can delay critical decision-making. Furthermore, decision-makers need to analyze text content and extract key information to formulate hazard prevention and control measures. This process is time-consuming and labor-intensive, which is problematic given that hydraulic structure failures can occur rapidly and lead to serious accidents. Additionally, the diverse types of hydraulic structure-related hazards (such as overtopping, breaching, and piping) involve large-scale management objects, multiple risk factors, and substantial resource allocation consumption. Relying solely on manual text review for decision-making increases the risk of errors, potentially resulting in unpredictable and severe consequences. Therefore, integrating intelligent technologies to create an efficient hydraulic structure safety management system is becoming increasingly urgent. Natural language processing (NLP) has been widely applied in the engineering field to rapidly extract knowledge from engineering texts. Initially, rule-based text processing methods were used for tasks such as word frequency statistics, topic extraction, and text classification [5,6]. These rules can accurately reflect the linguistic features of a specific domain, making them advantageous when processing texts with a well-defined structure. However, they require manual rule creation, and their coverage and generalization ability are limited. Subsequently, some supervised learning (SL) methods, such as support vector machines (SVM) [7] and n-gram models [8], overcome the limitations of manual feature construction by learning language patterns automatically from labeled texts. However, these methods still struggle to deeply understand text at the semantic level. With the advent of deep learning, which utilizes complex network architectures to capture contextual semantic relationships, NLP development has been greatly accelerated. Models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer networks have been widely applied in tasks such as construction text classification, risk factor extraction, and specification content parsing [9–11]. The emergence of transformer models has led to the development of large language models (LLM), represented by GPT [12–15], which have transformed the field of NLP. LLMs possess strong language comprehension abilities, allowing them to understand text semantics and generate coherent, logical text, thus providing technical support for the intelligent extraction and reuse of hydraulic structures hazard knowledge. However, LLMs also have notable limitations, such as the phenomenon of hallucination (generation of inaccurate or misleading information) and a lack of domain-specific knowledge [16,17]. Due to the strict requirements in engineering design and construction, most existing LLMs are designed for general purposes and insufficient to be directly applied in engineering practices. One feasible solution is to construct domain-specific knowledge bases to enhance LLM performance. Meanwhile, the scattered hydraulic structure's safety management knowledge forms disconnected knowledge islands, making it difficult to quickly and efficiently retrieve relevant information during a hazard. Therefore, to achieve effective hydraulic structure safety management, it is necessary to extract knowledge from extensive text sources and build an interconnected hydraulic structures hazard knowledge base. Furthermore, human expression habits often result in knowledge containing numerous function words that serve as connectors between concepts, as well as inverted sentence structures or nested knowledge. For example, the term "woven bags" can be both a component of the "Construction method" knowledge and an independent piece of knowledge within the "Hazard supplies" category (as illustrated in Fig. 1). Knowledge and the relation are often hidden within the semantic structure of the text, making knowledge extraction more difficult. Additionally, variations in language preferences among writers, including inconsistent terms and colloquial expressions in safety management texts, increase the risk of feature extraction errors. Therefore, it is essential to analyze knowledge features and their interrelationships from a semantic perspective while considering global semantic features to achieve accurate extraction of knowledge and relations.

To achieve accurate and efficient hydraulic structure safety management, with levee safety management as the application scenario, this study proposes a hydraulic structure safety management question-answering (QA) method based on a domain-specific knowledge graph enhanced large language model. First, a pre-trained model is used to comprehensively represent text semantics, integrating the Kolmogorov-arnold network (KAN) and attention mechanisms to combine the partitioned and global features of knowledge and relation, thus achieving high-precision knowledge extraction. Next, the extracted knowledge is used to construct a levee safety management knowledge graph (LSM-KG), with a voting strategy employed to interpret query intent and extract relevant domain-specific knowledge from the LSM-KG, thereby filling knowledge gaps in the LLM. Finally, leveraging the LLM's powerful semantic understanding and generation capabilities, combined with domain knowledge, the method provides accurate safety management knowledge to support informed decision-making for managers.

The main contributions of this study are as follows: (1) This is the first exploration of combining LLM with knowledge graph (KG) in the field of engineering safety management. This approach can effectively enhance the efficiency of management and the quality of decision-making. (2) A joint entity-relation extraction model tailored for hydraulic structure safety management was proposed. This model can efficiently construct domain-specific knowledge graphs. (3) A joint voting-based intent parsing method, integrating LLM, SL, and rule-based models, was introduced. This method significantly improves the accuracy of knowledge-based question answering. The remainder of this manuscript is organized as follows. Section 2 reviews and analyzes related studies; Section 3 presents the framework and detailed explanation of the proposed method; Section 4 compares and analyzes the performance of the method; Section 5 discusses the model's capabilities; Section 6 summarizes the research contributions and limitations.

2. Related studies

2.1. Research on engineering knowledge mining and reuse

A substantial amount of hydraulic structure's safety management knowledge is stored in unstructured text, posing challenges in knowledge management and effective utilization [18]. Text mining, by deeply modeling text features and extracting key information, has played a crucial role in areas such as construction document classification [19,20], contract information extraction [21], and automatic compliance checking [22–24]. Early text mining methods primarily relied on domain-specific dictionaries and statistical models [6,25]. Which, while feasible to some extent, had limited capacity for processing long-distance semantic information. The advent of deep learning has advanced long-distance text representation. For instance, bidirectional long short-term memory networks (BiLSTM) [26] alleviate the issue of long-distance semantic loss, and combining BiLSTM with conditional random fields (CRF) [27] further improves information extraction accuracy. Pretrained models, such as bidirectional encoder representations from transformers (BERT), have overcome the challenges of text semantic distortion, finding widespread application in areas such as underground engineering water inrush risk mining [28] and potential

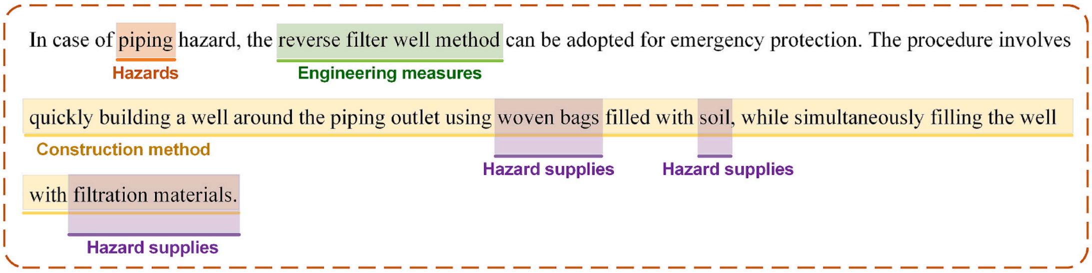


Fig. 1. Knowledge nesting phenomenon in hydraulic structure safety management texts.

hazard association extraction from safety incident texts [29]. Although these efforts have effectively mined textual information, they overlooked the organization and storage of this information within knowledge systems. Hydraulic structures risk events occur rapidly and require a quick response. Extracting information after risk events is difficult to meet the needs of safety management.

KG is a graph-based semantic knowledge repository, that represents knowledge and its relationships through entities and relations. KGs can effectively store and manage information from diverse data sources [30]. Unlike traditional relational databases, KGs model concepts and relations explicitly, offering stronger semantic representation and more flexible knowledge expression capabilities [31], thus supporting efficient information retrieval [32]. In the process of KG construction, some studies have divided entity and relation extraction into two sequential tasks [33–35]; however, this approach suffers from error propagation, where errors in earlier steps cannot be corrected in subsequent steps [36]. To address this, some scholars have proposed methods to mitigate error impacts by sharing an encoding layer [37,38], but such methods may generate isolated entities with no relational connections during the decoding process. To overcome this limitation, Cao et al. [11] proposed a table-filling global pointer network, which successfully improved the accuracy of risk factor relation extraction, offering a valuable reference for high-precision knowledge extraction. While KGs serve as excellent repositories for domain knowledge, existing methods primarily focus on the construction of KGs. How to effectively feedback KG knowledge into hydraulic structure safety management remains an area for further exploration.

QA is an effective method for reusing domain-specific knowledge. Tian et al. [39] used a twin neural network to deeply match questions and answers, enhancing construction safety management efficiency. Lin et al. [40] developed a BIM knowledge question-answering system based on BERT and cosine similarity, achieving an accuracy rate of 65 %. However, such systems require large amounts of labeled data and struggle to integrate effectively with KGs. With advances in artificial intelligence, LLMs have shown impressive performance in semantic understanding. One approach to applying LLMs in vertical domains is to introduce domain knowledge through domain preference alignment or incremental learning [41,42], which has shown initial success in e-commerce [43] and the medical field [44]. However, such methods require vast amounts of data and substantial computational power. To reduce computational costs, Álvaro et al. [45] mounted academic data onto an LLM via text chunking and evaluated the effectiveness of vector-based retrieval-augmented generation (RAG) techniques in manufacturing QA. Han et al. [46] built a knowledge QA system for the carbon neutrality domain based on LLMs, providing expert-level accuracy. Which demonstrates the potential of LLMs in hydraulic structure safety management QA. However, naive RAG is still limited by the similarity between questions and answers, and valuable answers are often randomly distributed across the knowledge base, with a lack of effective relationships between pieces of knowledge, making it difficult to obtain all relevant answers in one pass. Knowledge graphs can provide effective knowledge associations. However, challenges remain in the development of QA systems based on KGs, particularly with respect to relationship mapping, i.e., accurately parsing user intent and retrieving matching answers from the knowledge base. Existing methods still face issues such as insufficient semantic understanding and imperfect matching [47]. Furthermore, while LLMs demonstrate logical coherence, their grasp of domain-specific knowledge often remains superficial and may even include erroneous information, making them unsuitable for direct application in hydraulic structure safety management. Enhancing LLMs with domain-specific knowledge could be an effective solution to overcome these limitations [48,49]. Therefore, it is essential to explore methods for constructing domain-specific KG, leveraging the strengths of both KGs and LLMs to build a QA system suitable for hydraulic structure's safety management, facilitating the intelligent generation of safety management measures.

2.2. Knowledge gaps and research objectives

In summary, hydraulic structure safety management relies heavily on domain-specific knowledge and requires rapid response capabilities. However, existing knowledge texts are fragmented, disconnected, and lack tools for efficient retrieval and utilization. Intelligent hydraulic structure safety management necessitates the construction of domain-specific knowledge graphs and the development of a fast, accurate knowledge QA system to guide user decision-making. The advantages and disadvantages of existing methods are summarized in Table 1. Specifically, the gaps in this task are as follows:

Texts related to hydraulic structure safety management are highly fragmented, low quality, and involve multiple entities and relations types. Their text structures often contain nested phenomena, making it essential to extract boundary and associative features of knowledge from the text semantics. Current knowledge extraction methods are insufficient to meet the required precision for constructing knowledge graphs in this domain.

In terms of knowledge QA, while LLMs perform excellently in text understanding and generation, they still suffer from deficiencies such as a lack of domain-specific knowledge and the generation of hallucinated responses [16,17]. These limitations necessitate the supplementation of domain knowledge to enhance their capabilities. Fine-tuning a specialized LLM requires large amounts of data, consumes significant resources, and struggles to adapt to changing field-specific information. In contrast, domain-enhanced LLMs that leverage real-time retrieval from external knowledge bases can dynamically update information, save resources, and offer greater flexibility in specialized fields. However, the use of KGs as external knowledge sources to identify user intent and extract professional answers for enhancing LLMs in the domain knowledge of safety management remains less explored. To address these gaps, this study employs partition fusion for precise text knowledge extraction and utilizes a graph database to store the knowledge. The graph structure enhances the LLM with closely related, context-rich domain knowledge, thereby generating professional and stable safety management recommendations, and providing technical support for rapid response and scientific decision-making in hydraulic structure safety management.

Table 1
Comparison of characteristics of existing QA methods.

Methods [citation]	Merits	Demerits
Traditional neural network-based QA systems ([50,51])	• Effectively captures sequence features and global information • Relatively fast execution speed	• Not suitable for complex types of questions • Lacks LLM support, limiting response flexibility and coherence
Pre-trained neural network-based QA systems ([39,40])	• Improves QA performance through semantic-based answer matching • Enhances response efficiency in construction safety management and BIM domains	• Requires large-scale annotations, leading to high costs • Weak knowledge association • Lacks LLM support, limiting response flexibility and coherence
Domain-aligned or incremental learning methods based on LLMs ([41,42,43])	• Having advantages in semantic understanding and expression • Broad applicability across various QA scenarios • Produces logically coherent and user-friendly responses	• Requires massive data and high-performance computing resources, leading to high implementation costs • Lacks domain knowledge base support, leading to hallucinations and limitations in knowledge updates
Naive retrieval-augmented generation techniques ([45,46,48,49])	• Reduces computational burden by integrating external knowledge bases • Improves answer relevance through efficient retrieval, lowering retrieval costs	• Valuable answers in the knowledge base are sparsely distributed, leading to weak correlations • Difficult to retrieve all relevant knowledge in a single query
Knowledge graph-based QA systems ([11,52])	• Establishes effective knowledge relationships, providing more comprehensive retrieval results • Supports multi-hop reasoning and complex queries	• Challenges in understanding user intent • Answer quality depends on knowledge extraction accuracy • Lacks LLM support, limiting response flexibility and coherence

3. Proposed solution

To achieve the construction of hydraulic structure safety management knowledge graph and the generation of scientific decision-making recommendations, a research framework was proposed as shown in Fig. 2, which consists of four main components: (1) data collection and ontology construction, (2) knowledge extraction and graph construction, (3) intent parsing and query conversion, and (4) answer retrieval and enhanced generation.

3.1. Data collection and ontology construction for hydraulic structure safety management

To ensure comprehensive knowledge coverage, texts were collected from diverse sources, including industry standards, risk manuals, academic papers, and web pages. Among these, standards and manuals were reviewed and published by national authoritative bodies, while academic papers underwent rigorous journal peer-review processes, ensuring high data credibility. For web data, provincial water resources society official accounts and the Baidu Encyclopedia were selected as sources. After automated extraction using scripts, the data was manually verified for accuracy and relevance to the research content, ensuring reliable data quality. Optical character recognition (OCR) technology was used for initial text extraction, followed by the application of multi-level filtering and selection methods: At the document level, content with repeated themes was removed based on chapter titles; at the sentence level, missing sentence components (such as subject restoration and pronoun resolution) were supplemented; and at the word level, garbled characters, spelling errors, and stop words were eliminated to obtain a clean dataset.

Ontology is a model that describes the concepts and relations within a specific domain, serving as the framework for constructing knowledge graphs [53]. The purpose of constructing an ontology model is to clearly define the knowledge boundaries of the domain of interest and to organize key concepts along with their relational patterns. Hydraulic structure safety management data, being unstructured text, often contains implicit relationships that are expressed semantically within the text. For example, the sentence "After a levee breach occurs, covering and reinforcing the areas at both ends of the break can prevent the breach from expanding further," implicitly conveys multiple five-tuple knowledge such as < Hydraulic structures "levee" – Existence – Hazards "breach">, <Hazards "breach" – Countermeasure – Engineering measures "covering and reinforcing">, and < Engineering measures "covering and reinforcing" – Location – Measure site "both ends of the break">, among others. The entities are diverse, and the relations are complex, making ontology construction a challenging task. In this study, combined with the needs of hydraulic structure safety management, referring to the "seven-step method" [54] of Stanford University, the establishment of domain concepts and relations were merged, and the attributes were no longer layered separately. It was simplified to the following "four-step method" to summarize the hydraulic structure safety management domain ontology.

(1) Ontology requirement analysis: the scope of the ontology is defined according to the research objectives. This study focuses on knowledge management and engineering reuse in the context of hydraulic structure safety management, so the ontology must cover a range of information including various hydraulic structure hazard types, identification methods, response measures, and resources required for disaster management. This step can be assisted by using a word segmentation tool to count text word frequency, employing a large language model to extract professional terms, and utilizing term clustering to help determine ontology concepts. For specific details, please refer to our previous work [56].

(2) Ontology reuse: existing ontologies in ontology libraries are reviewed for potential reuse. Based on previous research [55,56], nine concepts, including "Hydraulic structures", "Hazards", "Definition", "Principles", "Hazard manifestation", "Engineering measures", "Hazard level", "Supplies", and "Measure site", along with eight types of relations, such as "Include", "Existence", "Definition", "Treatment principle", "Countermeasure", "Corresponds to", "Need", and "Location", were migrated and integrated into the hydraulic structure safety management ontology construction.

(3) Ontology extension: missing concepts and relations can be supplemented based on practical needs. Previous research [56] lacked focus on risk management, particularly neglecting critical information such as the construction process, precautions, and other key information that aid decision-making. Therefore, 2 additional concepts related to "Hazards", including "Potential cause", "Feature", as well as 9 concepts related to "Engineering measures", including "Precautions", "Characteristics", "Machinery", "Function", "Scope", "Methods", "Site conditions", "Attributes", and "Requirements" were introduced. Additionally, 10 new relations, such as "Reason", "Identification", "Alias", "Caution", "Characteristic", "Application", "Function", "Encounter", "Construct", and "Require", were added to link these new concepts, providing comprehensive coverage of the hydraulic structure safety management domain.

(4) Ontology construction: based on the reuse and extension processes, a series of discussions with domain experts and researchers were conducted to analyze concept scope and organize concept relations. The ontology model tailored for hydraulic structure safety management was

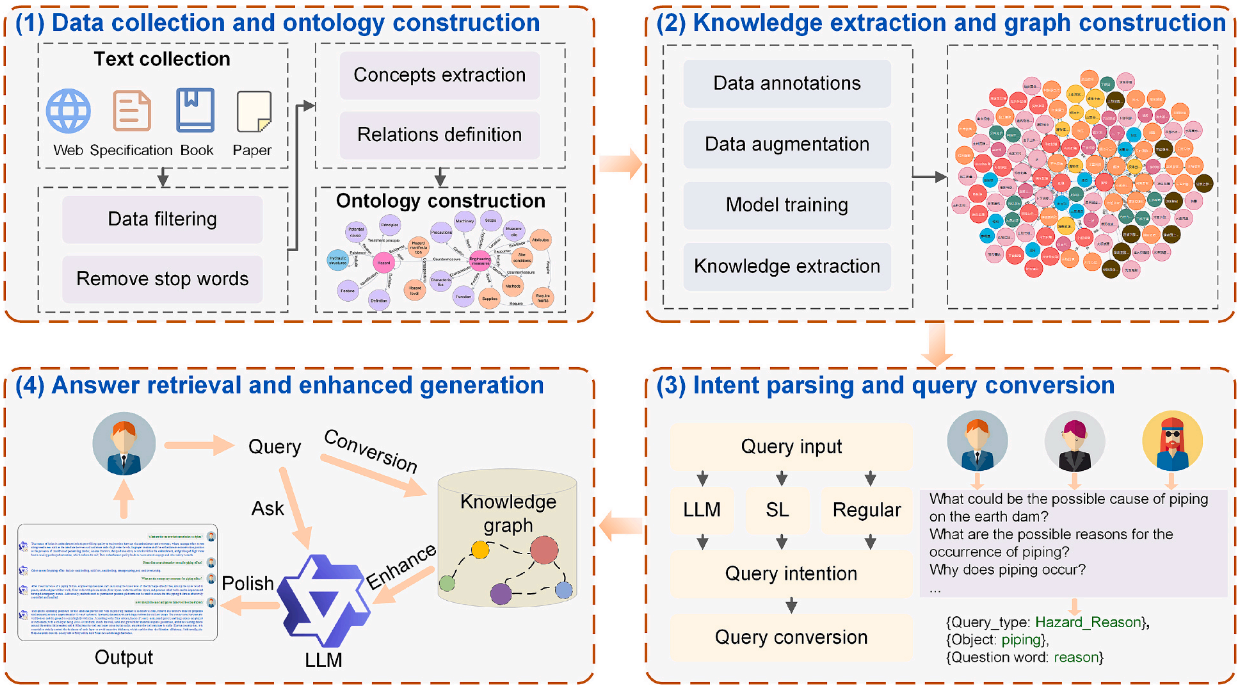


Fig. 2. Overall research framework.

collaboratively constructed. The ontology scope was continuously adjusted and updated throughout the data annotation and extraction process, resulting in the ontology model shown in Fig. 3. Annotate the collected data based on the ontology model, and then enhance the dataset using LLM's text understanding and synonym transformation capabilities to expand the data volume and improve knowledge extraction accuracy.

3.2. Knowledge extraction based on PFKAN

Knowledge extraction is the primary method for acquiring knowledge to build a knowledge graph. Based on the differences in processing objects, it can be divided into two tasks: Named Entity Recognition (NER) and Relation Extraction (RE). The distinctions between these two tasks are illustrated in Table 2.

Hydraulic structure safety management texts contain numerous entity concepts, with many nested entities and complex relations, many of which are implicitly conveyed in the text's semantics. Existing knowledge extraction methods [9,56] struggle to handle such complex texts. To extract implicit knowledge from unstructured text effectively, semantic representation serves as the foundation, feature extraction is the key, and knowledge prediction is the core. In view of this, we proposed an entity-relation joint extraction model, PFKAN, based on feature partition fusion and the Kolmogorov-arnold network (KAN), as shown in Fig. 4.

In the feature extraction layer, we use the BERT module, which has strong semantic representation capabilities, to encode each input word W_i into a semantically rich text vector V_i. These text vectors are then passed through the feature partition module, where entity and relation features are extracted separately. During the feature partition process,

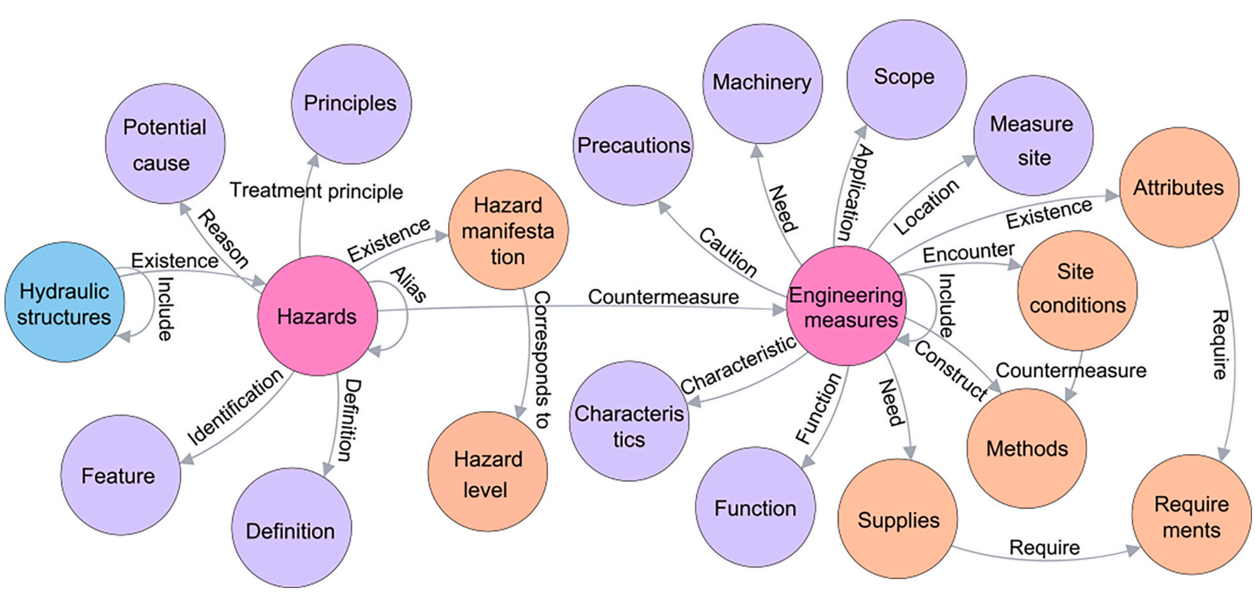


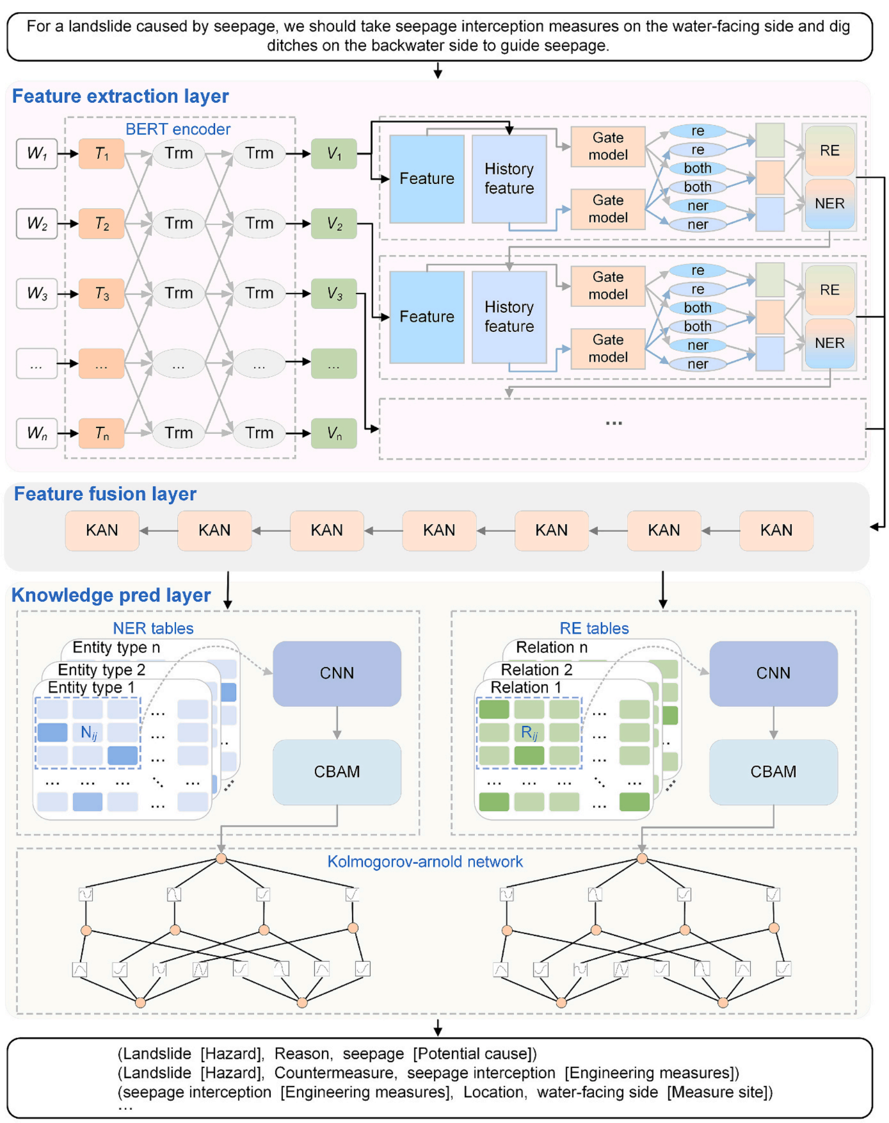
Fig. 3. Hydraulic structure safety management ontology model.

Table 2
NER task VS. RE task.

	NER task	RE task
Raw text	After a levee breach occurs, covering and reinforcing the areas at both ends of the break can prevent the breach from expanding further.	
<Head entity-Relation-Tail entity>	<Levee – Existence – Breach> <Breach – Countermeasure – Covering and reinforcing> <Covering and reinforcing – Location – Both ends of the break>	
Concept	Identify entity knowledge with specific meaning from text	Identify semantic relationships between entities from text
Role	The knowledge nodes that make up the knowledge graph	The relationship edges that make up the knowledge graph
Examples	Levee, Breach, Covering and reinforcing, Both ends of the break	Existence, Countermeasure, Location
Characteristic	It is directly included in the raw text and can be obtained by slicing from the raw text	It is often implicitly embedded in the semantics and requiring determination based on semantic analysis

the partition information from previous words as historical features and input into the partition module together with the current feature. The gating unit respectively splits the historical features and current features into three feature categories: "re" features, "ner" features, and shared "both" features. The division of "re" and "ner" features decouples the characteristics required for relation extraction and entity extraction, avoiding interference from irrelevant features in a single task. The "both" features preserve the information interaction between the two tasks, representing the correlation between them. Then, the historical and current features are added together, injecting contextual information into the current feature to enrich its representation. Subsequently, the "both" features are concatenated with the task-specific features to form the "RE" and "NER" features, which are used in subsequent computations.

In the feature fusion layer, since the feature extraction layer only considers local context features, the KAN network [57] was introduced to enhance the fusion of global knowledge features. Unlike traditional neural network architectures, KAN abandons the use of fixed activation functions at network nodes and instead employs dynamically learnable spline functions on the edges, providing improved flexibility. This characteristic makes KAN particularly effective in handling complex



Notes: "Trm" means Transformer, "KAN" means Kolmogorov-arnold network, "RE" means relation extraction, "NER" means named entity recognition, "CNN" means convolutional neural network, and "CBAM" means convolutional block attention module.

Fig. 4. Flowchart of proposed PFKAN model.

data patterns and nonlinear relationships. By utilizing the B-spline basis functions of the KAN network to construct nonlinear mapping functions, the knowledge extraction task first applies conventional linear transformations to extract basic text features, and then captures nonlinear features through spline transformations. Through feature addition and fusion, the “RE” and “NER” features are globally integrated respectively to obtain their comprehensive feature representation.

In the knowledge prediction layer, tables are constructed based on the feature fusion results using the broadcasting mechanism, with the table-filling task serving as knowledge encoding. The NER tables are used to determine the start and end positions of entity knowledge, while the RE tables identify the starting and ending entities of relations. The table’s rows and columns correspond to the length of the input text. Specifically, let i represent the row and j represent the column. In the NER tables, the cell $N_{i,j}$ represents the probability that the text between the i-th and j-th characters is a specific type of entity. To extract entities of multiple categories, the number of NER tables is equal to the number of entity categories. Similarly, in the RE tables, the cell $R_{i,j}$ represents the probability that a relation of a specific type exists between the entity starting at the i-th character and the entity starting at the j-th character. In order to extract multiple types of relations, the number of RE tables is equal to the number of relation categories. After knowledge encoding, CNNs are used to compress the knowledge features, with further refinement achieved using the convolutional block attention module (CBAM) [58] attention mechanism, which focuses on both spatial and channel attention. During decoding, the KAN network is employed for accommodating the complex feature partitioning and table encoding. The encoded entity and relation information from the tables is input into the KAN network, which uses its dynamic activation functions to handle issues related to feature sparsity and uneven distribution in the table encoding. The nonlinear modeling ability of B-spline basis functions in KAN improves the precision of knowledge prediction. After the hydraulic structure safety management knowledge text is processed by the PFKAN model, the knowledge can be directly extracted in the form of $<$ head entity, head entity type, relation, tail entity, tail entity type$>$, providing critical information for constructing the KG.

3.3. Voting-based intent parsing

The knowledge graph stores a vast amount of specialized knowledge related to hydraulic structure safety management. In contrast to fine-tuning LLMs using domain knowledge, using KG for graph-based retrieval-augmented generation to enhance LLMs for knowledge QA is a promising approach to generating hazard mitigation strategies. Intent parsing is a critical step in identifying the query subject and core issue embedded in the user’s question. Rule-based intent parsing requires analyzing the linguistic features of the input query and setting up rule templates for specific question types. Although this method requires reconfiguration when switching domains, it remains effective in certain tasks due to its simplicity, interpretability, and fast parsing speed. On the other hand, SL-based intent parsing can automatically learn the style and expression patterns of questions through sample training, making it adaptable to linguistic variations and offering higher accuracy and generalization capability. LLMs with billions or even tens of billions of parameters, have the potential to handle complex tasks, including intent parsing. However, their output stability is often suboptimal, and the responses frequently reflect the model’s understanding rather than directly outputting the specified intent. Considering the strengths and weaknesses of these different methods, it was found that there are significant differences and complementary advantages between them. Therefore, a voting strategy was proposed, as shown in Fig. 5, where model weights are allocated based on their performance. Weighted voting is then used to achieve accurate intent parsing by aggregating the results from different models.

(1) Rule-based intent parsing. First, the trained PFKAN model was used to extract the entity types in the query, which serve as the question subject. Next, the syntactic structure of the query was analyzed to identify the specific question words and questioning methods associated with the sentence. By combining the question subject and the questioning words, the intent of the query is determined. For example, in the sentence “What is the best construction location for backfilling and compaction measures?”, after entity extraction, it was identified that “backfilling and compaction measures” belongs to the category of “Engineering measures,” and “location” was classified as a question word related to location. As a result, the intent of the query can be classified as “measure location.” The ontology model constructed in Fig. 3 contains a total of 23 types of $<$ head entity, relation, end entity $>$ triples. By merging similar types, 20 distinct intent categories were obtained. A list of the various intents and corresponding question words is provided in Table 3.

(2) LLM-based intent parsing. To fully leverage the semantic understanding capabilities of LLMs while ensuring that the generated intent type aligns with the intent in Table 3, three control measures were implemented. First, the trained PFKAN model was used to extract the entity and its type from the query, and this information was provided as prior input to assist the LLM in parsing the query’s intent. Second, a small set of intent parsing examples was used as prompt samples to stimulate the LLM’s capability in the domain of hydraulic structures hazard intent parsing. This enables the model to choose the most appropriate intent from a predefined list of intent categories. Finally, for some outputs that do not meet the expected criteria, a normalization mapping process was applied. For instance, both “measure advantage”

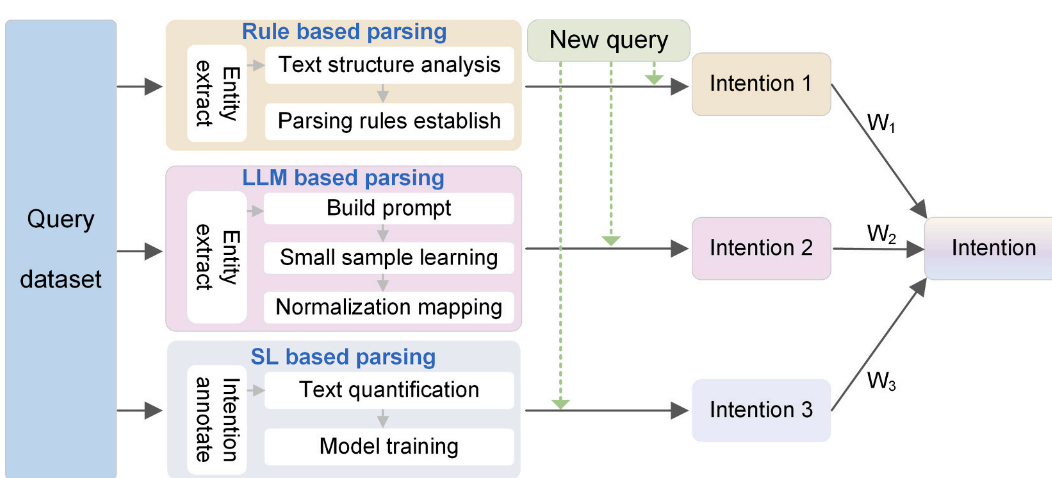


Fig. 5. Intent parsing process.

Table 3
Intent types and corresponding question words.

Code	Intent type	Question words	Code	Intent type	Question words
a	other	None	**K**	subclass of structures	contain, unit, which parts…
A	measure location	position, area, location…	**L**	structures hazard	risk, danger, emergency…
B	measure function	usage, goal, benefits…	**M**	hazard severity	severity, level, significant…
C	subclass of measure	subclass, unit, which parts…	**N**	hazard identify	recognize, judge, feature…
D	measure construction	operate, how to do…	**O**	hazard alias	vernacular name, epithet…
E	measure precaution	attention, comply, focus…	**P**	subclass of hazards	contain, unit, which parts…
F	measure supplies	need, prepare, materials…	**Q**	hazard cause	inducement, origin, cause…
G	measure feature	strength, weakness…	**R**	principle of response	guiding principle, policy…
H	measure method	special cases, handle…	**S**	method of handling	respond, take, handle…
J	measure scope	applicable range, situation…	**T**	hazard definition	concept, mean, explain…

and “measure disadvantage” were mapped to “measure feature,” ensuring that the LLM’s output remains controlled and consistent.

(3) SL-based intent parsing. First, the query dataset was annotated with intent to create an intent parsing dataset, which was then split into a training set and a test set. The training set was used to train the SL model, while the test set was used to evaluate the accuracy of the models and calculate their weights. The voting weight w_i for each model can computed using Equation (1).

$$w_i = \frac{f_i}{\sum_1^n f_i} \tag{1}$$

where f_i represents the F1 score of the i-th voting model on the test set, which can be calculated by equations (2), (3), and (4).

$$precision = \frac{correct\_recognize}{all\_recognize} \tag{2}$$

$$recall = \frac{correct\_recognize}{total} \tag{3}$$

$$\mathrm{F1} = 2 \times \frac{precision \times recall}{precision + recall} \tag{4}$$

where *correct_recognize* denotes the number of true positives correctly identified by the model, *all_recognize* refers to the total number of positives identified by the model, and *total* represents the total actual number of positive instances, *precision* reflects the proportion of correctly identified positive samples among all samples classified as positive by the model; *recall* indicates the proportion of correctly identified positive samples out of all actual positive instances. The F1 score is the harmonic mean of precision and recall, providing a comprehensive measure of model performance. During voting, I_i represents the intent obtained by the i-th voting model, and the weighted vote for a candidate intent I is the sum of the weights assigned by each model that classifies the intent as I, as shown in Equation (5). The final intent I_o is the candidate intent that receives the highest number of votes, as indicated in Equation (6).

$$votes(I) = \sum_1^n w_i if I_i = I \tag{5}$$

$$I_o = \mathrm{argmax}(votes(I)) \tag{6}$$

3.4. Knowledge extraction and answer generation

KGs cannot directly process natural language queries. To facilitate natural language to natural language QA, it is necessary to convert natural language queries into a format that the graph can understand. In this study, the Neo4j graph database was used as the storage platform for the hydraulic structure safety management knowledge graph. This database offers flexible capabilities for adding, deleting, querying, and modifying data, with fast response times and support for multi-hop reasoning [59], making it convenient for knowledge retrieval. Neo4j uses the Cypher query language; therefore, for each intent category in Table 3, specific conversion rules were designed to convert natural language queries into Cypher queries, preparing the system for real-time feedback from KG.

To generate meaningful and convincing answers based on domain knowledge, a knowledge QA process was designed, as shown in Fig. 6. For user-initiated queries related to hydraulic structure safety management, the system sends processing requests to both the KG and LLMs, facilitating a comprehensive integration of the two. In the KG branch, intent parsing and query conversion are employed to extract domain-specific hydraulic structure safety management knowledge, enhancing the LLM’s understanding of the query and answer while mitigating hallucination effects. In the LLM branch, prompt engineering is designed to enable the model to prioritize domain knowledge from the KG while supplementing it with the knowledge stored within the LLM itself. This approach ensures the generation of accurate, professional, and logically coherent safety management recommendations.

4. Application scenario

4.1. Data preparation

This section takes the safety management of levee engineering as an application scenario to verify the effectiveness of the proposed framework. Extreme rainfall events can trigger risk incidents in levee infrastructure. For instance, in 2024, extreme precipitation in the Dongting Lake basin led to a series of incidents involving piping, collapse, and eventually breach of the levee (Fig. 7(a)). Although timely evacuation prevented casualties, the event caused great economic losses. In addition to these risks, levees are also vulnerable to a range of other hazards, such as structural cracks, overtopping, and wave erosion, with some of these hazards illustrated in Fig. 7(b).

To construct a hydraulic structure safety management knowledge QA system, a variety of relevant texts, including standards, manuals, web pages, and academic papers, were collected. These texts were processed and cleaned using the methods described in Section 3 to yield high-quality safety management texts. Subsequently, two domain experts with extensive expertise in hydraulic engineering and proficiency in entity annotation, were selected to annotate the text. The annotations included entities within the text and the relations between entities, where relations were directed from head entities to tail entities. Part of the annotation results is shown in Fig. 8. To ensure the quality of the annotations, the first 10 % of the text was annotated collaboratively, and once consensus was reached, the remaining texts were annotated individually. Throughout the annotation process, cross-checking was conducted. Upon completion, all texts were categorized according to the relation type, and the consistency of each type of annotation was reviewed. Then, samples were selected using a stratified sampling method based on relation categories in a 3:1:1 ratio, creating the training, validation, and test datasets. After deduplication, the data was

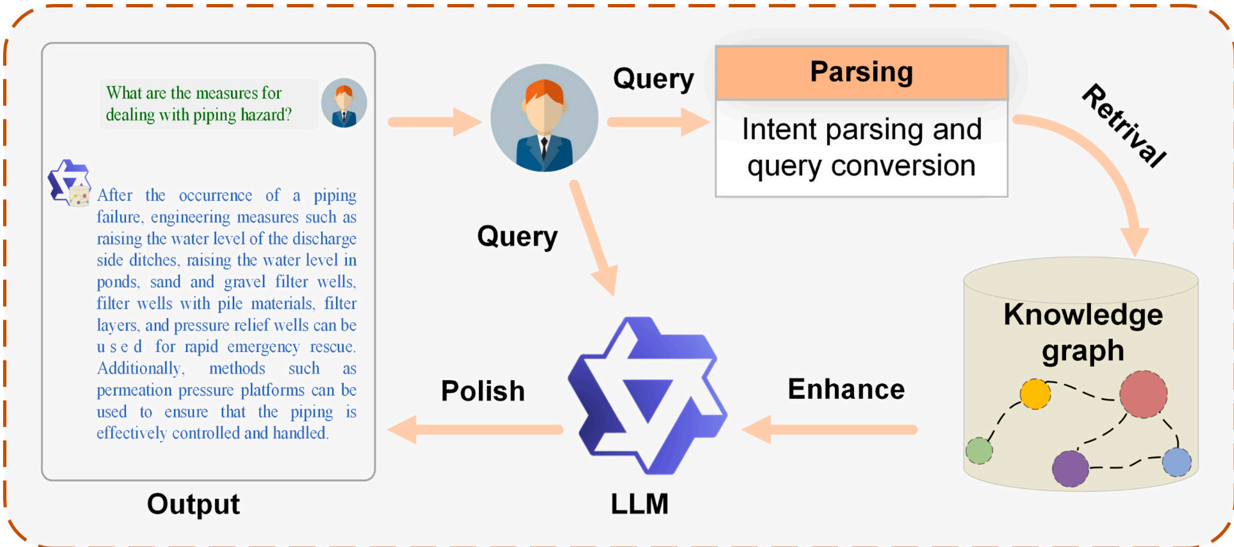


Fig. 6. Knowledge-based question answering process.

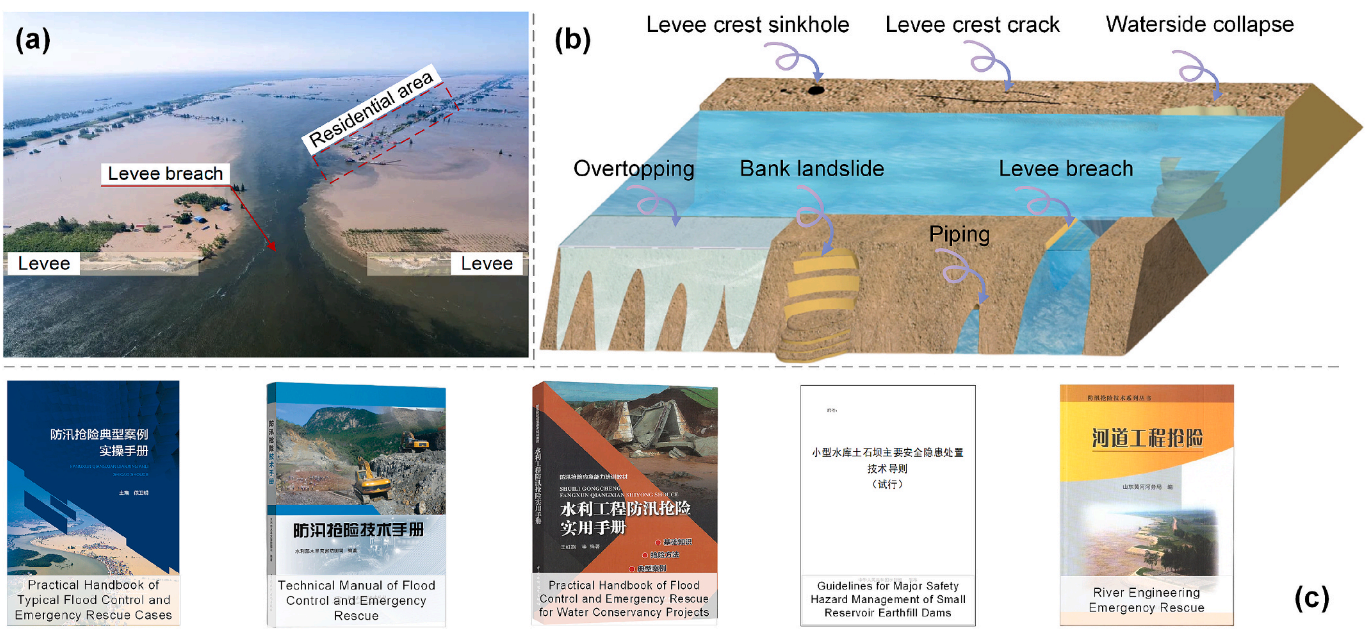


Fig. 7. Levee hazards and safety management knowledge texts.

doubled using the synonym conversion capability of LLM, and a total of 4468 gold standards were obtained. To complete the intent parsing task, a questionnaire was designed, and some answers were added to the questionnaire according to the intent category. Eight participants were asked to generate both technical and colloquial queries from different perspectives based on these answers. After expert screening, 440 valid queries were obtained and split into training and test sets in a 3:1 ratio for training the SL intent parsing model. To evaluate the model's performance, *precision*, *recall*, and F1 score were selected as metrics for assessment.

4.2. Performance of PFKAN model

4.2.1. Comparison of multiple models for entity extraction

The PFKAN model was built using the PyCharm programming platform with PyTorch 1.13.0 as the deep learning framework. During training, the learning rate was set to 3×10^{-5}, and the Adam optimizer was used to prevent getting stuck in local optima. The binary cross-entropy loss function was applied to guide model training, while L1 regularization and the DropConnect mechanism were employed to accelerate model convergence. To evaluate the model's performance in the entity extraction task, the widely used W2NER [60], BERT_BiLSTM_CRF (BBC) [24], and BiLSTM_CRF [61] models were selected as comparison models. The training was set for 200 epochs, and the performance of each model on the validation set is shown in Fig. 9(a). As observed from Fig. 9(a), the W2NER model requires a longer warm-up period, while other models show a rapid improvement in precision from the start of training. Notably, the F1 curve of the proposed PFKAN model has the sharpest slope, reaching its inflection point first and constantly maintaining the highest level, indicating optimal convergence during training. This can be attributed to the KAN module's flexible feature extraction capability and strong nonlinear modeling power, which allows the model to consistently improve precision during training, reaching convergence at an accelerated rate. Fig. 9(b)

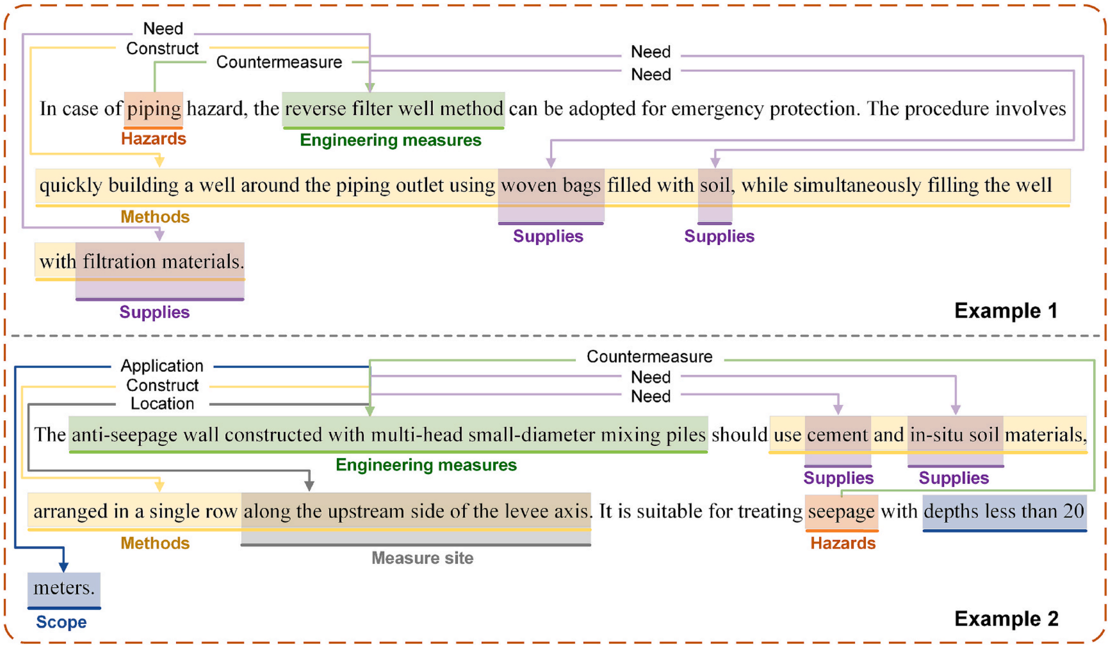


Fig. 8. Examples of dataset annotation results.

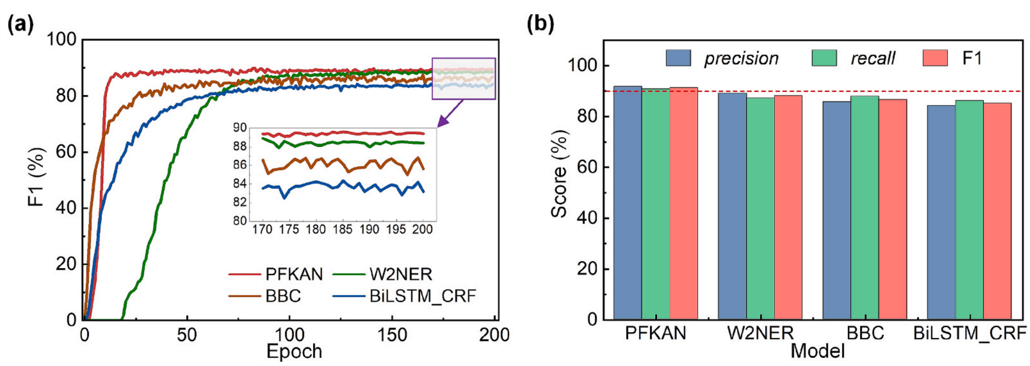


Fig. 9. Performance of each model on the (a) validation set and (b) test set in entity extraction.

compares the performance of each model on the test set, showing that the proposed model achieves an F1 score of 91.27, outperforming BiLSTM_CRF by 6.11 points and W2NER, the second-best model, by 3.18 points, demonstrating the best performance.

Fig. 10 illustrates the detailed performance of the PFKAN model in extracting 20 types of entities of hydraulic structure safety management.

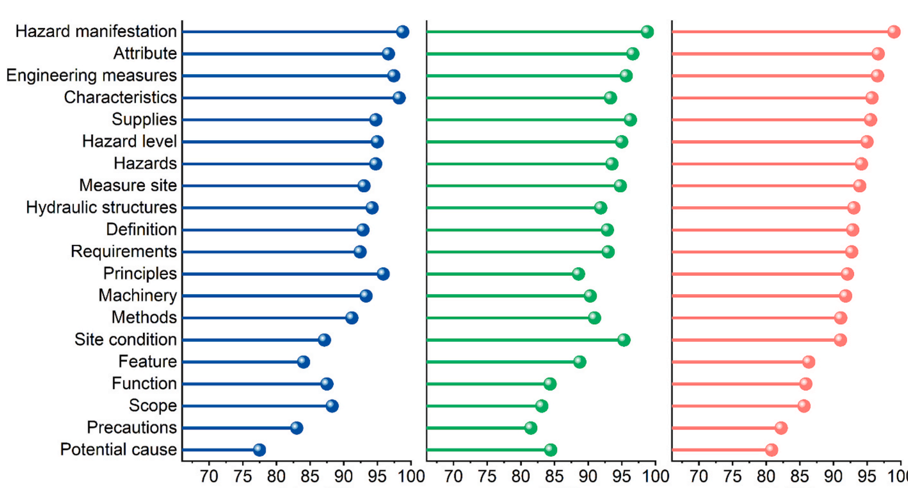


Fig. 10. Performance of the PFKAN model in each entity category.

Despite the large number of entity types and significant variations in entity lengths, the results show that the F1 score for each entity category exceeds 80, demonstrating competitive performance. This confirms that the model provides an effective solution for domain entity knowledge extraction.

4.2.2. Comparison of multiple models for relation extraction

To evaluate the performance of the PFKAN model in the relation extraction task, GPLink [62], PRGC [63], and SPN4RE [64] were selected as comparison models. The changes in the F1 scores of each model during training on the validation set are shown in Fig. 11(a). As seen from Fig. 11(a), although the PRGC model exhibits a rapid increase in F1 scores during the early stages of training, its performance fluctuates throughout the training process. In contrast, the PFKAN model steadily improves and surpasses PRGC in F1 score after the 12th epoch. To further assess the actual use performance of the models in relation extraction, Fig. 11(b) compares their results on the test set. The results show that the PFKAN model achieves an F1 score of 89.92, outperforming PRGC by 4.33 points, thus demonstrating the most competitive performance.

Fig. 12 illustrates the specific extraction results of PFKAN on 18 types of relations defined in the ontology model. The results indicate that even the "Reason" relation, which has the lowest F1 score, still achieves a score close to 80, and more than half of the category's *precision* values exceed 90 %. When analyzing Fig. 9(b) and 11(b) together, it shows that the overall accuracy of relation extraction is similar to that of entity extraction, i.e. the performance of relation extraction is not substantially affected by entity extraction. Suggesting that the PFKAN model effectively avoids error propagation in the joint entity-relation extraction task. This can be attributed to the fact that PFKAN separates entity and relation features early in the feature extraction process while incorporating global feature interactions. This strategy effectively prevents the error propagation caused by the sequential extraction process in traditional models and retains the necessary feature relationships between entities and relations, resulting in optimal performance.

4.2.3. Construction of levee safety management knowledge graph

The PFKAN model can convert scattered and unstructured texts of various levee safety management into well-structured five-tuple knowledge. Subsequently, the Neo4j graph database was used to sequentially read these knowledge chains, merge nodes with identical values, and gradually expand the knowledge chain into a knowledge graph. Using the PFKAN model, a total of 3,516 entity knowledge and 4,121 relation knowledge were extracted. These knowledge elements were sequentially populated into Neo4j to form a levee safety management knowledge graph (LSM-KG). The local visualization of the KG structure is shown in Fig. 13(a), with some detailed visualizations presented in Fig. 13(b). These visualizations show how LSM-KG stores safety management information using nodes and edges to map relationships between hazard types, response measures, construction methods, locations, materials, and so on. This structured knowledge enables effective support in the question-answering system.

4.3. Performance of domain knowledge-enhanced LLM

4.3.1. Intent parsing performance

In the intent parsing dataset, most queries are brief (under 15 words), making complex models unnecessary. Here SVM is chosen in intent parsing tasks for their proven effectiveness in text classification [10]. First, a BERT model was used to encode query semantics, and then an SVM model was trained and saved on this encoded data. After training, F1 scores were calculated on the test set using three intent parsing methods: rule-based, SL-based (SVM), and LLM-based. The performance of each method is presented in the form of confusion matrices, as shown in Fig. 14. The rows of the confusion matrix represent the true labels, while the columns represent the predicted labels for each model. The letters on the matrix coordinate axes correspond to the intent categories in Table 3.

Due to the broad range of intents in the "other" category, which makes it difficult to effectively define in the SL intent parsing training set, the "other" category was excluded from the classification results. As seen in Fig. 14, the rule-based intent parsing method achieves the lowest F1 score because it misclassifies a significant number of samples as the "other" category. The SL-based and LLM-based methods both achieve identical F1 scores. Substitute the F1 values of the three into Eq. (1) to calculate their respective weights, and then use Eqs (4) and (5) to design a voting strategy. The intent parsing results using the voting strategy are shown in the last image of Fig. 14. The results demonstrate that, after applying the voting strategy, the precision for intent parsing increases to 0.97, and the F1 score reaches 0.95, showing a significant improvement compared to individual models. The confusion matrix generated by the voting strategy indicates a notable reduction in misclassifications, confirming the effectiveness of the proposed voting strategy.

4.3.2. Comparison of safety management knowledge question answering performance

To validate the enhanced performance of the LSM-KG in safety management question answering, the large language model Qwen2.5-7B was selected as the base model. This model is the 2.5th generation of Alibaba's open-source Qwen [15] model, with 7 billion parameters, and offers strong support for Chinese. Four randomly selected queries of different types were used to test the model's performance before and after LSM-KG enhancement, with Question 4 being an extension of Question 3. The results, translated into English, are shown in Fig. 15(a) for the enhanced responses and in Fig. 15(b) for the unenhanced ones.

A comparison of Fig. 15(a) and 15(b) reveals that LSM-KG enriches answers with more technical detail and expertise. For instance, in Question 1, the enhanced response lists potential causes for levees through seepage, offering practical guidance for managers to identify and fix issues, which is of significant value for safety management. In

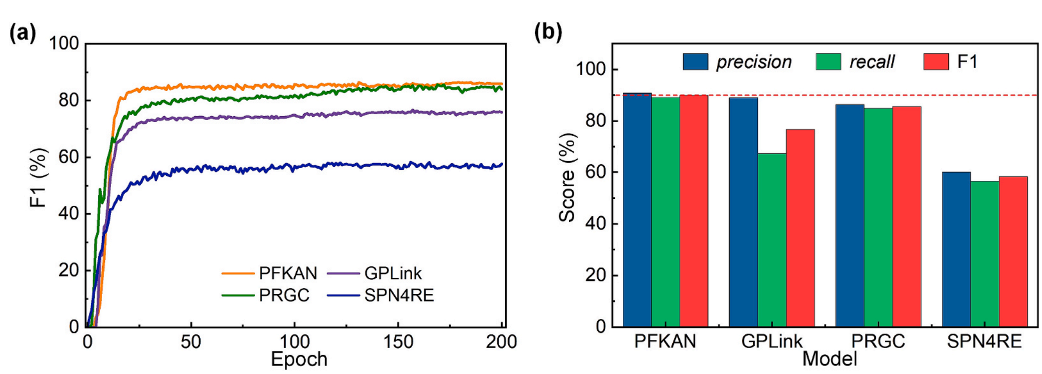


Fig. 11. Performance of each model on the (a) validation set and (b) test set in relation extraction.

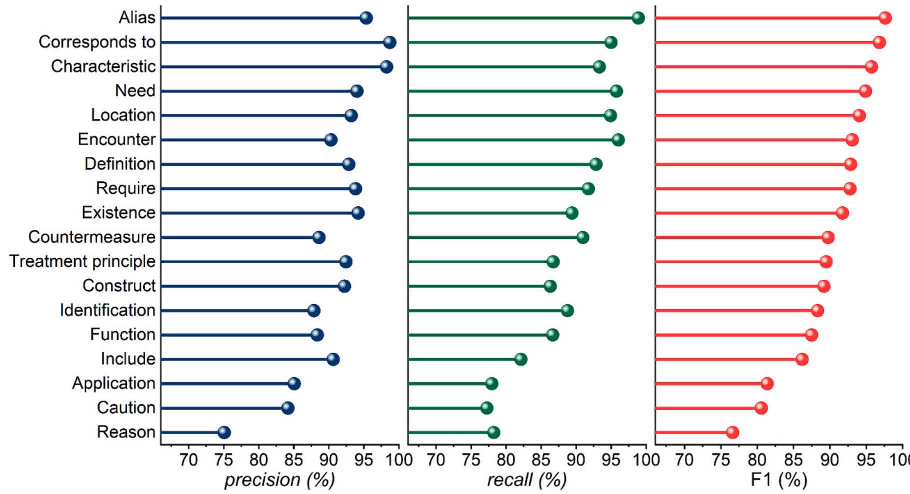


Fig. 12. Performance of the PFKAN model in each relation category.

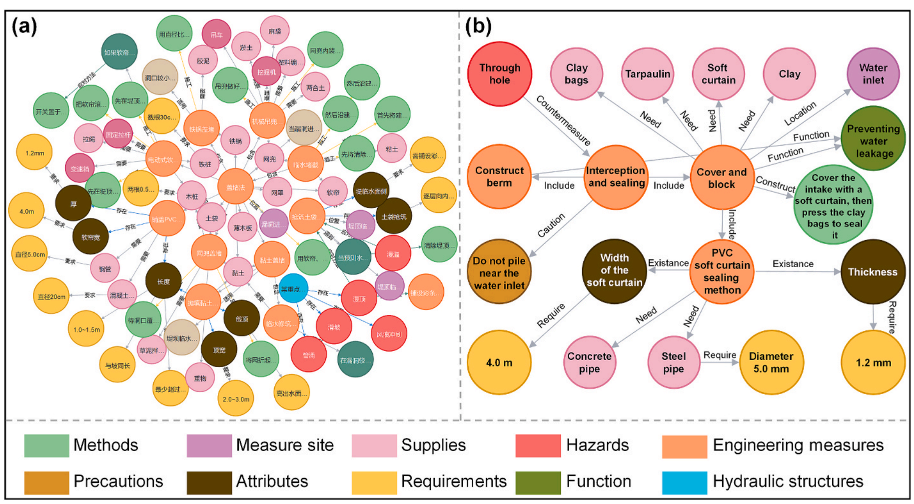


Fig. 13. Visualization of (a) the structure and (b) the details of LSM-KG (English translation from Chinese).

contrast, the unenhanced answer only offers general categories like "Geological conditions" and "Human factors," which suit general audiences but lack practical guidance. Furthermore, the unenhanced answer contains clear hallucinations—subtle errors that are difficult to detect (highlighted in red). For example, in Fig. 15(b), answer 2 mistakenly refers to "cracks" as an alias for "piping," although they actually represent different types of hazards. In point 4 of answer 3, the suggested measure does not apply to the piping rescue; in fact, the piping rescue must follow the principle of "blocking above and draining below." The "block the piping outlet" strategy could exacerbate the issue by generating new piping hazards. Additionally, grouting methods, which require pressure, are prohibited for piping emergencies. In answer 4, the basic response contains two errors: Point 1 wrongly suggests rescue efforts can be moved from the piping exit, and Point 5 incorrectly requires response height to exceed the levee top. These mistakes could mislead non-experts and increase overall risks. Moreover, the unenhanced responses are inconsistent across queries. In contrast, the LSM-KG-enhanced answers are stable and professional, offering concrete operational suggestions and numerical details, enabling better risk analysis and decision-making.

5. Discussion

Analysis of Figs. 10 and 12 reveals that the lower precision in certain categories of knowledge extraction with the PFKAN model often occurs within the related triple categories. For example, in relation extraction, the categories "Caution" and "Reason," as well as the entity categories "Precautions" and "Potential cause," exhibit lower precision. These categories frequently appear together in triples such as < Engineering measures – Caution – Precautions > and < Hazards – Reason – Potential cause > . To find out the reasons, a further statistical analysis of the entity data distribution and average entity length in the training set was performed, as shown in Table 4. It was observed that entities with F1 scores below 90, such as "Precautions," "Potential cause," and "Function," tend to have smaller data proportions and longer average entity lengths, which complicates knowledge extraction and reduced precision. However, the categories "Definition" and "Hazard manifestation" were exceptions. Despite having fewer data points and longer lengths, they exhibit consistent expression patterns that make them easier for the model to learn. Thus, it can be concluded that, in most cases, fewer training samples and longer entity lengths hinder text feature extraction, resulting in reduced model precision. Overall, the F1 scores for most knowledge extraction tasks exceed 80, with half surpassing 90, indicating that the proposed PFKAN model demonstrates satisfactory performance.

Additionally, the causes of errors in intent parsing were analyzed in conjunction with Fig. 14. Since both rule-based and LLM-based intent parsing methods rely on entity extraction results as prior information,

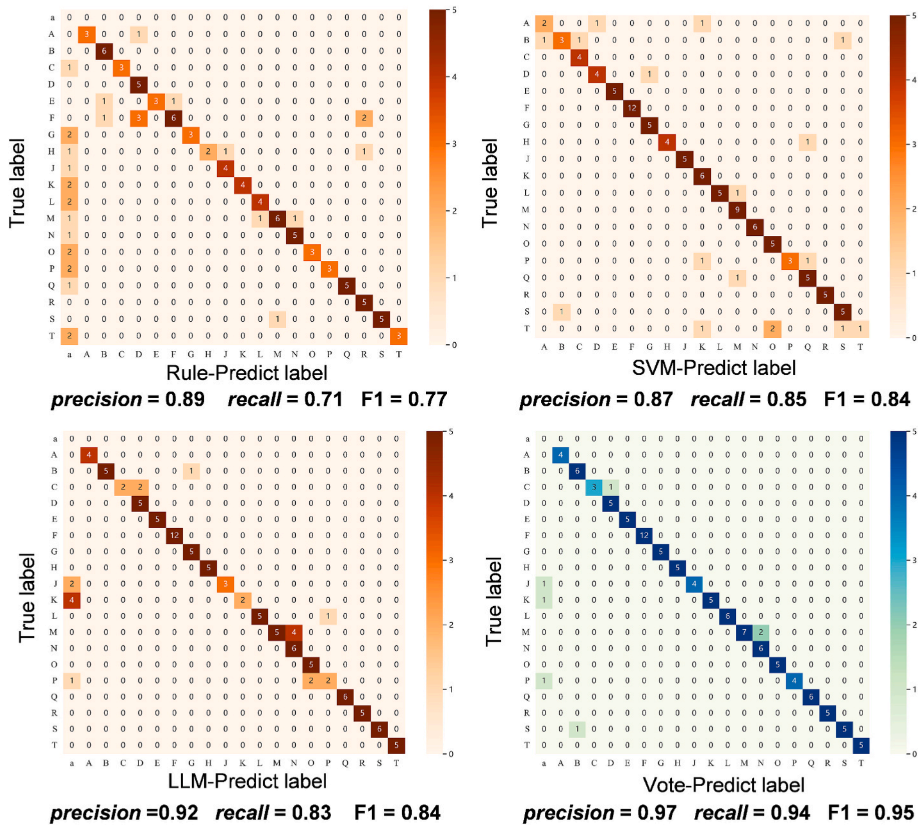


Fig. 14. Results of the single intent parsing model and the voting model.

the accuracy of entity recognition directly impacts their performance. For example, in LLM recognition, the "subclass of structures" class (K-class) exhibits lower precision due to errors in entity recognition. For example, in the question "What are the components of the gate equipment?", "gate equipment" should be the "Hydraulic structures" entity, but it was mistakenly recognized as the "Machinery" entity, causing LLM to categorize the query as "other" intent. Moreover, rule-based intent recognition errors also suffer from incomplete keyword coverage. In the query "What is overtopping?", the phrase "what is" cannot be set as a question word because it is too common, resulting in misclassification as "other" due to the absence of clear question words during rule parsing. In contrast, SL-based intent parsing shows a more even error distribution, primarily due to random model errors. The voting strategy in SL-based intent parsing mitigates the impact of entity recognition accuracy, thereby compensating for the shortcomings of the other models to a certain extent. Overall, the results indicate that the combined voting strategy of the three models effectively improves parsing accuracy.

This proposed strategy first analyzes the scope of hydraulic structure's safety management concepts and constructs an ontology model to guide knowledge extraction. A feature partitioning and fusion approach was adopted to develop the PFKAN knowledge extraction model, which demonstrates superior performance in hydraulic structure safety management knowledge extraction tasks, leading to the creation of a hydraulic structure safety management knowledge graph. Subsequently, a voting strategy was employed to parse user intent. By combining the domain knowledge graph with LLM, the domain knowledge enhances the expertise of the LLM, resulting in stable and professional expert-level safety management responses. This approach improves the efficiency of domain knowledge reuse and enhances the hydraulic structure's safety management capabilities. The proposed method has the following significant advantages: (1) Improved accuracy in domain knowledge extraction. The idea of partition fusion decouples entity features and relation features while taking into account the correlation between them. Combined with the feature modeling ability of the KAN network, it eliminates error propagation and improves knowledge extraction accuracy. This approach is also applicable across various domains. (2) Enhanced flexibility and credibility in domain-specific QA. The knowledge graph supports flexible expansion and modification. With the addition of new knowledge, the knowledge base can be incrementally stored and directly expanded. Additionally, with specialized domain knowledge, LLM provides in-depth, detailed, and reliable answers, which significantly improves the credibility of the answer. (3) Increased efficiency in hydraulic structure safety management. Traditional hydraulic structure safety management requires managers to read texts individually and interpret their semantics, which is time-consuming and inefficient. With the proposed method, managers can simply ask questions in natural language, while the model automatically handles intent parsing, knowledge retrieval, and response generation. This automation significantly reduces response time, enhances decision reliability, and improves the efficiency of hydraulic structure's safety management.

In terms of scalability, the method proposed in this paper demonstrates strong generalizability. (1) The ontology model used in this study itself possesses scalability and reusability characteristics. The concepts and relationships contained in the proposed ontology are suitable for safety management and operation maintenance processes across various engineering domains. (2) The KG offers flexible knowledge management and expansion capabilities, enabling the extension of its knowledge coverage by extracting relevant knowledge from adjacent fields. (3) The LLM supports multiple languages, meaning that texts written in different languages can also be transformed by the proposed framework into domain-specific knowledge that enhances the LLM's expertise, and this can provide professional and efficient risk management decision support for managers in different countries.

Fig. 15. Question answering performance before and after KG-enhanced LLM (English translation from Chinese).

Table 4
Entity F1 and its influencing factors.

Entity class	F1	Count (proportion)	Average length	Entity class	F1	Count (proportion)	Average length
Definition	92.86	244 (1.64 %)	34.64	Principles	92.08	304 (2.05 %)	6.92
Methods	91.08	1351 (9.11 %)	30.15	Characteristics	95.73	437 (2.95 %)	6.67
Hazard manifestation	99.39	140 (0.94 %)	21.79	Measure site	93.90	534 (3.60 %)	6.05
Precautions	**82.24**	**334 (2.25 %)**	**21.57**	Engineering measures	96.54	2740 (18.47 %)	5.69
Feature	**86.34**	**206 (1.39 %)**	**17.89**	Hazard level	95.00	136 (0.92 %)	4.03
Scope	**85.62**	**441 (2.97 %)**	**17.27**	Attribute	96.64	431 (2.91 %)	3.93
Function	**85.89**	**518 (3.49 %)**	**15.64**	Hydraulic structures	93.04	396 (2.67 %)	3.32
Potential cause	**80.82**	**694 (4.68 %)**	**14.04**	Machinery	91.80	229 (1.54 %)	3.31
Site condition	91.04	357 (2.41 %)	12.26	Hazards	94.15	1543 (10.40 %)	3.11
Requirements	92.71	1044 (7.04 %)	9.62	Supplies	95.52	2756 (18.58 %)	2.69

Note: The bolded parts are entity categories with F1 scores below 90.

6. Concluding remarks

Engineering safety management requires substantial professional knowledge and is a labor and knowledge-intensive task. This study extends the ontology framework in the field of engineering safety management. It not only achieves the accurate extraction of safety management knowledge through innovative knowledge extraction methods but also proposes effective knowledge retrieval strategies, enabling the feedback of knowledge into engineering practices. By integrating advanced artificial intelligence technologies, this work formalizes fragmented and complex engineering knowledge for engineers, facilitating the transition from an empiricist safety management approach to knowledge-driven safety management. The main contributions are as follows:

(1) A new framework that combines KG with LLMs is proposed to improve specialized knowledge QA. This framework fixes the limitations of LLMs in providing incorrect or misleading answers, by leveraging the expertise of KG. Our results show that this method produces accurate, expert-level advice, outperforming LLMs alone in offering hydraulic structure safety management decisions.

(2) A domain knowledge extraction model is constructed by using a BERT pretraining module to represent textual semantics and a KAN network to integrate global features. The proposed approach decouples entity and relation features to prevent errors from affecting precision. Experimental results demonstrate that the proposed model achieves outstanding results in entity and relation extraction tasks, with F1 scores of 91.27 and 89.92, respectively.

(3) A voting strategy-based intent parsing method is introduced by combining three strategies: rule-based, supervised learning, and LLMs. This strategy effectively mitigates the limitations of models and the impact of entity recognition errors, accurately parses user intent, and reduces instances of irrelevant or off-target responses, thereby providing effective support for domain-knowledge-enhanced LLM.

Despite remarkable progress, there are also limitations. Future work will focus on two areas: (1) Collecting more domain-specific textual data and improving knowledge extraction models by combining LLM to extend the applicability of the proposed method, and (2) enhancing response organization and usage of domain knowledge in external models. Particularly, fine-tuning specialized LLMs using domain knowledge will be a key goal to improve overall quality.

CRediT authorship contribution statement

Dongliang Zhang: Writing – original draft, Visualization, Validation, Methodology, Data curation. **Gang Ma:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Tongming Qu:** Writing – review & editing, Validation. **Xudong Wang:** Investigation, Data curation. **Wei Zhou:** Supervision, Funding acquisition. **Xiaomao Wang:** Resources, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (Grant No.2022YFC3005505), and the National Natural Science Foundation of China (Grant Nos. 52322907, 52179141, U23B20149). The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

Data availability

Data will be made available on request.

References

[1] C. Chu, H. Su, C. Wei, G. You, Semi-quantitative evaluation of levees risk within the context of hydrology uncertainty, J. Hydrol. 617 (2023) 128750, https://doi.org/10.1016/j.jhydrol.2022.128750.
[2] Association of State Dam Safety, Dam failures and incidents, (2024). https://damsafety.org/dam-failures. (accessed November 20, 2024).
[3] D. Blockley, The Nature of Structural Design and Safety (1981), https://doi.org/10.1002/nme.1620170218.
[4] N. Adamo, N. Al-Ansari, V. Sissakian, J. Laue, S. Knutsson, Dam safety and overtopping, (2020) https://www.diva-portal.org/smash/get/diva2:1443995/FULLTEXT01.pdf.
[5] K. Chen, Z. Zhang, J. Long, H. Zhang, Turning from TF-IDF to TF-IGM for term weighting in text classification, Expert Syst. Appl. 66 (2016) 245–260, https://doi.org/10.1016/j.eswa.2016.09.009.
[6] S. Kim, J. Gil, Research paper classification systems based on TF-IDF and LDA schemes, Hum-Cent. Comput. Info. 9 (2019) 30, https://doi.org/10.1186/s13673-019-0192-7.
[7] Q. Guo, S. Cao, Z. Yi, A medical question answering system using large language models and knowledge graphs, Int. J. Intell. Syst. 37 (2022) 8548–8564, https://doi.org/10.1002/int.22955.
[8] E. Zhu, J. Zhang, J. Yan, K. Chen, C. Gao, N-gram MalGAN: Evading machine learning detection via feature n-gram, Digit Commun. Netw. 8 (2022) 485–491, https://doi.org/10.1016/j.dcan.2021.11.007.
[9] D. Tian, M. Li, Y. Shen, S. Han, Intelligent mining of safety hazard information from construction documents using semantic similarity and information entropy, Eng. Appl. Artif. Intell. 119 (2023) 105742, https://doi.org/10.1016/j.engappai.2022.105742.
[10] D. Zhang, M. Li, D. Tian, L. Song, Y. Shen, Intelligent text recognition based on multi-feature channels network for construction quality control, Adv. Eng. Inform. 53 (2022) 101669, https://doi.org/10.1016/j.aei.2022.101669.
[11] K. Cao, S. Chen, C. Yang, Z. Li, L. Luo, Z. Ren, Revealing the coupled evolution process of construction risks in mega hydropower engineering through textual semantics, Adv. Eng. Inform. 62 (2024) 102713, https://doi.org/10.1016/j.aei.2024.102713.
[12] Z. Du, Y. Qian, X. Liu, M. Ding, J. Qiu, Z. Yang, J. Tang, GLM: General language model pretraining with autoregressive blank infilling, ArXiv Prepr. (2022), https://doi.org/10.48550/arXiv.2103.10360.
[13] J. OpenAI, S. Achiam, S.A. Adler, et al., GPT-4 technical report, ArXiv Prepr. (2024), https://doi.org/10.48550/arXiv.2303.08774.
[14] H. Touvron, T. Lavril, G. Izacard, et al., LLaMA: Open and efficient foundation language models, ArXiv Prepr. (2023), https://doi.org/10.48550/arXiv.2302.13971.
[15] J. Bai, S. Bai, Y. Chu, et al., Qwen technical report, ArXiv Prepr. (2023), https://doi.org/10.48550/arXiv.2309.16609.
[16] J. Yao, K. Ning, Z. Liu, M. Ning, Y. Liu, L. Yuan, L.L.M. Lies, Hallucinations are not bugs, but features as adversarial examples, ArXiv Prepr. (2024), https://doi.org/10.48550/arXiv.2310.01469.
[17] C. Feng, X. Zhang, Z. Fei, Knowledge Solver: Teaching LLMs to search for domain knowledge from knowledge graphs, ArXiv Prepr. (2023), https://doi.org/10.48550/arXiv.2309.03118.
[18] M. Marzouk, M. Enaba, Text analytics to analyze and monitor construction project contract and correspondence, Autom. Constr. 98 (2019) 265–274, https://doi.org/10.1016/j.autcon.2018.11.018.
[19] H. Gadekar, N. Bugalia, Automatic classification of construction safety reports using semi-supervised YAKE-Guided LDA approach, Adv. Eng. Inform. 56 (2023) 101929, https://doi.org/10.1016/j.aei.2023.101929.
[20] W. Fang, H. Luo, S. Xu, P. Love, Z. Lu, C. Ye, Automated text classification of near-misses from safety reports: An improved deep learning approach, Adv. Eng. Inform. 44 (2020) 101060, https://doi.org/10.1016/j.aei.2020.101060.
[21] G. Fantoni, E. Coli, F. Chiarello, R. Apreda, F. Dell'Orletta, G. Pratelli, Text mining tool for translating terms of contract into technical specifications: Development and application in the railway sector, Comput. Ind. 124 (2021) 103357, https://doi.org/10.1016/j.compind.2020.103357.
[22] J. Zhang, N. El-Gohary, Semantic NLP-based information extraction from construction regulatory documents for automated compliance checking, J. Comput. Civ. Eng. 30 (2016) 04015014, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000346.
[23] Y. Zhou, Z. Zheng, J. Lin, X. Lu, Integrating NLP and context-free grammar for complex rule interpretation towards automated compliance checking, Comput. Ind. 142 (2022) 103746, https://doi.org/10.1016/j.compind.2022.103746.
[24] Q. Ren, D. Zhang, M. Li, S. Chen, D. Tian, H. Li, L. Liu, Automatic quality compliance checking in concrete dam construction: Integrating rule syntax parsing and semantic distance, Adv. Eng. Inform. 60 (2024) 102409, https://doi.org/10.1016/j.aei.2024.102409.
[25] R. Mihalcea, P. Tarau, TextRank: Bringing order into text, in: D. Lin, D. Wu (Eds.), Proc. 2004 Conf. Empir. Methods Nat. Lang. Process., Association for Computational Linguistics, Barcelona, Spain, (2004) pp. 404–411. https://aclanthology.org/W04-3252.
[26] J. Santoso, E. Setiawan, C. Purwanto, E. Yuniarno, M. Hariadi, M. Purnomo, Named entity recognition for extracting concept in ontology building on Indonesian language using end-to-end bidirectional long short term memory, Expert Syst. Appl. 176 (2021) 114856, https://doi.org/10.1016/j.eswa.2021.114856.
[27] Y. An, X. Xia, X. Chen, F. Wu, J. Wang, Chinese clinical named entity recognition via multi-head self-attention based BiLSTM-CRF, Artif. Intell. Med. 127 (2022) 102282, https://doi.org/10.1016/j.artmed.2022.102282.
[28] M. Xiong, H. Wang, Y. Wong, Z. Hou, Enhancing aviation safety and mitigating accidents: A study on aviation safety hazard identification, Adv. Eng. Inform. 62 (2024) 102732, https://doi.org/10.1016/j.aei.2024.102732.

[29] L. Zhang, J. Wang, Y. Wang, H. Sun, X. Zhao, Automatic construction site hazard identification integrating construction scene graphs with BERT based domain knowledge, Autom. Constr. 142 (2022) 104535, https://doi.org/10.1016/j.autcon.2022.104535.

[30] A. Hogan, E. Blomqvist, M. Cochez, C. D'amato, G. Melo, C. Gutierrez, S. Kirrane, J. Gayo, R. Navigli, S. Neumaier, A. Ngomo, A. Polleres, S. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, A. Zimmermann, Knowledge graphs, ACM Comput. Surv. 54 (2021) 1–37, https://doi.org/10.1145/3447772.

[31] Q. Wang, Z. Mao, B. Wang, L. Guo, Knowledge graph embedding: a survey of approaches and applications, IEEE Trans. Knowl. Data Eng. 29 (2017) 2724–2743, https://doi.org/10.1109/TKDE.2017.2754499.

[32] J. Chen, Q. Song, C. Zhao, Z. Li, Graph database and relational database performance comparison on a transportation network, in: M. Singh, P. Gupta, V. Tyagi, J. Flusser, T. Ören, G. Valentino (Eds.), Adv. Comput. Data Sci., Springer, Singapore, (2020) pp. 407–418. DOI: 10.1007/978-981-15-6634-9_37.

[33] Y. Luan, D. Wadden, L. He, A. Shah, M. Ostendorf, H. Hajishirzi, A general framework for information extraction using dynamic span graphs, in: J. Burstein, C. Doran, T. Solorio (Eds.), Proc. 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol., Association for Computational Linguistics, Minneapolis, Minnesota, (2019) pp. 3036–3046. DOI: 10.18653/v1/N19-1308.

[34] Y. Chan, D. Roth, Exploiting syntactico-semantic structures for relation extraction, in: D. Lin, Y. Matsumoto, R. Mihalcea (Eds.), Proc. 49th Annu. Meet. Assoc. Comput. Linguist. Hum. Lang. Technol., Association for Computational Linguistics, Portland, Oregon, USA, (2011) pp. 551–560. https://aclanthology.org/P11-1056.

[35] F. Ren, L. Zhang, X. Zhao, S. Yin, S. Liu, B. Li, A simple but effective bidirectional framework for relational triple extraction, in: Proc. Fifteenth ACM Int. Conf. Web Search Data Min, Association for Computing Machinery, New York, NY, USA, 2022, pp. 824–832, https://doi.org/10.1145/3488560.3498409.

[36] Q. Li, H. Ji, Incremental joint extraction of entity mentions and relations, in: K. Toutanova, H. Wu (Eds.), Proc. 52nd Annu. Meet. Assoc. Comput. Linguist. Vol. 1 Long Pap., Association for Computational Linguistics, Baltimore, Maryland, (2014) pp. 402–412. DOI: 10.3115/v1/P14-1038.

[37] M. Miwa, M. Bansal, End-to-end relation extraction using LSTMs on sequences and tree structures, in: K. Erk, N.A. Smith (Eds.), Proc, 54th Annu., Association for Computational Linguistics, Berlin, Germany, 2016, pp. 1105–1116, https://doi.org/10.18653/v1/P16-1105.

[38] X. Wang, N. El-Gohary, Deep learning-based relation extraction and knowledge graph-based representation of construction safety requirements, Autom. Constr. 147 (2023) 104696, https://doi.org/10.1016/j.autcon.2022.104696.

[39] D. Tian, M. Li, Q. Ren, X. Zhang, S. Han, Y. Shen, Intelligent question answering method for construction safety hazard knowledge based on deep semantic mining, Autom. Constr. 145 (2023) 104670, https://doi.org/10.1016/j.autcon.2022.104670.

[40] T. Lin, Y. Huang, A. Putranto, Intelligent question and answer system for building information modeling and artificial intelligence of things based on the bidirectional encoder representations from transformers model, Autom. Constr. 142 (2022) 104483, https://doi.org/10.1016/j.autcon.2022.104483.

[41] W. Liao, X. Chu, Y. Wang, TPO: Aligning Large Language Models with Multi-branch & Multi-step Preference Trees, (2024). DOI: 10.48550/arXiv.2410.12854.

[42] Y. Feng, Z. Tian, Y. Zhu, Z. Han, H. Luo, G. Zhang, M. Song, CP-Prompt: Composition-Based Cross-modal Prompting for Domain-Incremental Continual Learning, in: Proc. 32nd ACM Int. Conf. Multimed., Association for Computing Machinery, New York, USA, 2024: pp. 2729–2738. DOI: 10.1145/3664647.3681481.

[43] K. Shi, X. Sun, D. Wang, Y. Fu, G. Xu, Q. Li, LLaMA-E: Empowering E-commerce Authoring with Object-Interleaved Instruction Following, in: O. Rambow, L. Wanner, M. Apidianaki, (Eds.), Proc. 31st Int. Conf. Comput. Linguist., Association for Computational Linguistics, Abu Dhabi, UAE, 2025: pp. 870–885. https://aclanthology.org/2025.coling-main.58/ (accessed March 10, 2025).

[44] Q. Lin, Y. Zhu, X. Mei, L. Huang, J. Ma, K. He, Z. Peng, E. Cambria, M. Feng, Has multimodal learning delivered universal intelligence in healthcare? A comprehensive survey, Inf. Fusion 116 (2025) 102795, https://doi.org/10.1016/j.inffus.2024.102795.

[45] J. Álvaro, J. Barreda, An advanced retrieval-augmented generation system for manufacturing quality control, Adv. Eng. Inform. 64 (2025) 103007, https://doi.org/10.1016/j.aei.2024.103007.

[46] T. Han, R. Cong, B. Yu, B. Tang, Y. Wei, Integrating local knowledge with ChatGPT-like large-scale language models for enhanced societal comprehension of carbon neutrality, Energy AI 18 (2024) 100440, https://doi.org/10.1016/j.egyai.2024.100440.

[47] M. Therasa, G. Mathivanan, ARNN-QA: Adaptive recurrent neural network with feature optimization for incremental learning-based question answering system, Appl. Soft Comput. 124 (2022) 109029, https://doi.org/10.1016/j.asoc.2022.109029.

[48] S. Siriwardhana, R. Weerasekera, E. Wen, T. Kaluarachchi, R. Rana, S. Nanayakkara, Improving the domain adaptation of retrieval augmented generation (RAG) models for open domain question answering, Trans. Assoc. Comput. Linguist. 11 (2023) 1–17, https://doi.org/10.1162/tacl_a_00530.

[49] R. Taiwo, I. Bello, S. Abdulai, A. Yussif, B. Salami, A. Saka, T. Zayed, Generative ai in the construction industry: a state-of-the-art analysis, ArXiv Prepr. (2024), https://doi.org/10.48550/arXiv.2402.09939.

[50] T. Shao, Y. Guo, H. Chen, Z. Hao, Transformer-based neural network for answer selection in question answering, IEEE Access 7 (2019) 26146–26156, https://doi.org/10.1109/ACCESS.2019.2900753.

[51] L. Cai, M. Wei, S. Zhou, X. Yan, Intelligent question answering in restricted domains using deep learning and question pair matching, IEEE Access 8 (2020) 32922–32934, https://doi.org/10.1109/ACCESS.2020.2973728.

[52] X. Ye, S. Yavuz, K. Hashimoto, Y. Zhou, C. Xiong, RNG-KBQA: Generation Augmented Iterative Ranking for Knowledge Base Question Answering, in: S. Muresan, P. Nakov, A. Villavicencio (Eds.), Proc. 60th Annu. Meet. Assoc. Comput. Linguist. Vol. 1, Association for Computational Linguistics, Dublin, Ireland, 2022: pp. 6032–6043. DOI: 10.18653/v1/2022.acl-long.417.

[53] S. Cardoso, M. Silveira, C. Pruski, Construction and exploitation of an historical knowledge graph to deal with the evolution of ontologies, Knowl.-Based Syst. 194 (2020) 105508, https://doi.org/10.1016/j.knosys.2020.105508.

[54] N. Noy, D. McGuinness, Ontology development 101: A guide to creating your first ontology, (2001) pp. 1–25. .

[55] L. Wang, X. Liu, Y. Liu, H. Li, J. Liu, L. Yang, Multimodal knowledge graph construction for risk identification in water diversion projects, J. Hydrol. 635 (2024) 131155, https://doi.org/10.1016/j.jhydrol.2024.131155.

[56] Zhang D., Zhou W., Ma G., Wang X., Liu Y., Wang X., Construction and application of knowledge graph for flood defense and rescue of water infrastructure, J. Hydraul. Eng. (2024) preprint 1-13. DOI: 10.13243/j.cnki.slxb.20240268. (In Chinese).

[57] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T.Y. Hou, M. Tegmark, KAN: Kolmogorov-arnold networks, ArXiv Prepr. (2024), https://doi.org/10.48550/arXiv.2404.19756.

[58] S. Woo, J. Park, J. Lee, I. Kweon, CBAM: Convolutional block attention module, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), Comput. Vis. ECCV 2018, Springer International Publishing, Cham, (2018) pp. 3–19. DOI: 10.1007/978-3-030-01234-2_1.

[59] Z. Li, H. Wang, W. Zhang, Translational relation embeddings for multi-hop knowledge base question answering, J. Web Semant. 74 (2022) 100723, https://doi.org/10.1016/j.websem.2022.100723.

[60] J. Li, H. Fei, J. Liu, S. Wu, M. Zhang, C. Teng, D. Ji, F. Li, Unified named entity recognition as word-word relation classification, ArXiv Prepr. (2021), https://doi.org/10.48550/arXiv.2112.10070.

[61] M. Wang, T. Zhou, H. Wang, Y. Zhai, X. Dong, Chinese power dispatching text entity recognition based on a double-layer BiLSTM and multi-feature fusion, Energy Rep. 8 (2022) 980–987, https://doi.org/10.1016/j.egyr.2022.02.272.

[62] J. Su, A. Murtadha, S. Pan, J. Hou, J. Sun, W. Huang, B. Wen, Y. Liu, Global Pointer: Novel efficient span-based approach for named entity recognition, ArXiv Prepr. (2022), https://doi.org/10.48550/arXiv.2208.03054.

[63] H. Zheng, R. Wen, X. Chen, Y. Yang, Y. Zhang, Z. Zhang, N. Zhang, B. Qin, X. Ming, Y. Zheng, PRGC: Potential relation and global correspondence based joint relational triple extraction, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), Proc. 59th Annu. Meet. Assoc. Comput. Linguist. 11th Int. Jt. Conf. Nat. Lang. Process., Association for Computational Linguistics, (2021) pp. 6225–6235. DOI: 10.18653/v1/2021.acl-long.486.

[64] D. Sui, X. Zeng, Y. Chen, K. Liu, J. Zhao, Joint entity and relation extraction with set prediction networks, IEEE Trans. Neural Netw. Learn. Syst. 35 (2024) 12784–12795, https://doi.org/10.1109/TNNLS.2023.3264735.