

Article

Not peer-reviewed version

Research on the Construction and Application of a Water Conservancy Facility Safety Knowledge Graph Based on Large Language Models

[Cui Li](#), [Yu Wang](#)^{*}, Lei Gao, Qiaoyan Ding

Posted Date: 6 February 2026

doi: 10.20944/preprints202602.0482.v1

Keywords: knowledge graph; large language model (LLM); water conservancy facility safety



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Research on the Construction and Application of a Water Conservancy Facility Safety Knowledge Graph Based on Large Language Models

Cui Li ^{1,2}, Yu Wang ^{3,*}, Lei Gao ^{1,2} and Qiaoyan Ding ³

¹ School of Economics and Management, University of Emergency Management, Hebei Province 065201, China

² Langfang Key Laboratory of Emergency Material Security and Logistics Management, University of Emergency Management, Sanhe 065201, China

³ Department of Disciplines and Graduate Studies, University of Emergency Management, Hebei Province, 065201, China

* Correspondence: 24661711@st.cidp.edu.cn; Tel.: +86-134-6041-8601

Abstract

To address the challenges of integrating multi-source heterogeneous data and low knowledge utilization rates in water conservancy facility safety management, this study proposes a knowledge graph construction method that integrates ontology modeling with large language model enhancement. First, an ontology framework for water conservancy facility safety is constructed, encompassing four core elements: agencies and personnel, engineering equipment, risks and hidden dangers, and systems and processes. Subsequently, a KG-LLM-GraphRAG architecture is designed, which optimizes the knowledge extraction effectiveness of large language models through ontology-constrained prompt templates and utilizes the Neo4j graph database for knowledge storage and multi-hop reasoning. Experimental results demonstrate that the proposed method significantly outperforms traditional approaches in entity-relationship extraction tasks. The constructed knowledge graph not only effectively supports application scenarios such as safety hazard identification, emergency decision-making, and knowledge reuse but also provides an efficient knowledge organization and reasoning tool for water conservancy facility safety management, strongly propelling the digital transformation of the water conservancy industry.

Keywords: knowledge graph; large language model (LLM); water conservancy facility safety

1. Introduction

Water conservancy facility safety serves as a public, foundational, and strategic safeguard for national economic and social development [1]. It is directly related to the safety of people's lives and property, the stability of the ecological environment, and sustainable economic and social development. The world possesses a vast network of water infrastructure. Taking China as an example, according to publicly available statistics, China currently has nearly 95,000 reservoirs, over 320,000 kilometers of embankments, and more than 580,000 water conservancy facilities including sluices, pumping stations, and irrigation districts. These facilities play an irreplaceable role in flood control and disaster reduction, water supply security, agricultural irrigation, hydropower generation, and other aspects. However, the safe operation of water conservancy facilities faces severe challenges, and their failure or accidents can lead to catastrophic consequences. For instance, during the "7·20" extreme rainstorm disaster in Zhengzhou, Henan, China, in 2021, safety issues such as spillway blockages and gate operation failures at the Guojiazui and Changzhuang reservoirs significantly amplified the destructive power of the flood, directly resulting in 398 deaths and missing persons, with economic losses exceeding 120 billion RMB. In 2020, the Edenville Dam in Michigan, USA,

breached due to structural aging and inadequate supervision, triggering chain disasters that caused direct economic losses of 250 million USD and the emergency evacuation of thousands of people. The enormous casualties and property losses profoundly underscore the extreme importance of water conservancy facility safety management.

Globally, major water conservancy nations have built intensive regulatory systems for safety management [2]. These systems are dynamically updated. China serves as an example. As of June 2024, China has 452 water safety regulations. These regulations cover a five-tier system. The tiers are: national, river basin, province, city, and county. The average number of regulation updates each year exceeds 37 items. There is a complex regulatory framework that evolves rapidly. It is combined with massive amounts of other data. This data includes engineering equipment information, risk records, institutional responsibility descriptions, and historical case data. Together, they form the core knowledge foundation for managing water conservancy facility safety. However, as textual regulatory information grows increasingly complex, traditional manual approaches face severe challenges in managing water facility safety knowledge, creating significant pressure on safety management. Specific manifestations include: inefficient processing of massive data [3], high costs and poor timeliness in tracking and updating safety regulations, inability to represent complex dynamic relationships between elements [4], and slow and error-prone manual retrieval of scattered information during emergencies. These limitations severely impact the timeliness and accuracy of emergency decision-making. Consequently, a significant gap exists between the vast and rapidly iterating cross-document knowledge volume and the need for instant and precise emergency decision-making. Traditional manual management methods are inadequate to address these challenges, thereby constraining the accurate identification of water conservancy facility safety risks and the reliability of emergency decisions.

Knowledge Graph (KG), with their powerful semantic network expression capabilities, have opened new pathways to overcome the bottlenecks of knowledge fragmentation and relationship complexity in water conservancy facility safety management. Although this technology has matured in fields such as healthcare and finance [5], and preliminary explorations in the water conservancy domain have also made progress in scenarios like water disaster prevention [6], engineering scheduling [7], and operation and maintenance management [8], existing research still faces dual challenges: Firstly, insufficient construction depth makes it difficult to accurately model the dynamic relational chains between water conservancy facilities; secondly, traditional knowledge extraction methods are inefficient in semantically mining unstructured text. Due to the lack of domain constraint mechanisms, “hallucination” errors frequently occur, severely constraining decision-making reliability. To address these core contradictions and the limitations of existing technologies, this paper proposes leveraging Large Language Models (LLM) to enhance the information extraction capability of knowledge graphs, integrating KG and LLM technologies, and introducing an ontology-driven KG-LLM-GraphRAG collaborative architecture. This architecture first constructs a water conservancy facility safety ontology as the core framework of the ontology layer: through the systematic integration of four types of entities—agencies and personnel, engineering equipment, risks and hidden dangers, and systems and processes—and the logical generalization of six types of semantic relationships such as operates/is operated by, a structured constraint template is formed. It then adopts a Retrieval-Augmented Generation (GraphRAG) mechanism to achieve dynamic construction at the data layer—accurately extracting entity attributes through multi-source data cleansing and prompt optimization, and relying on the Neo4j graph database for efficient storage and dynamic updating of triples. Most critically, the predefined ontology serves as a generation template to real-time correct LLM outputs, thereby enhancing the accuracy of professional domain knowledge extraction and effectively suppressing the risk of “hallucination.”

Based on this, the present study aims to construct a water conservancy facility safety management knowledge graph capable of integrating complex cross-document knowledge, accurately expressing complex dynamic relationships, and supporting emergency response. This will

effectively address the prominent contradictions in current water conservancy facility knowledge management and enhance the intelligence level of water conservancy facility safety management.

2. Related Works

2.1. Large Language Model Empowerment of Knowledge Graphs

As a structured semantic knowledge base, the core of a Knowledge Graph (KG) lies in its use of an ontology-defined ontology layer to achieve standardized knowledge representation [9], while organizing knowledge at the data layer in the form of “entity-relation-entity” or “entity-attribute-attribute value” triples [10]. This structure provides a powerful framework for integrating multi-source heterogeneous data and revealing complex relationships between entities, making it critical infrastructure in the field of artificial intelligence [11]. Research on knowledge graph construction methods has undergone significant iteration. Early studies primarily focused on effectively extracting entities and relationships from textual data to build structured knowledge bases, such as employing statistical models to automatically identify and associate multi-source entities [12], laying the foundation for knowledge graph construction. However, these traditional methods relied on predefined rules and templates, limiting their ability to handle complex semantic relationships. To overcome these limitations, knowledge graph completion techniques based on deep learning have emerged, with graph neural networks seeing widespread application in complex entity-relationship graph structures [13]. Simultaneously, multi-task learning technology has been introduced to jointly process entity recognition and relationship extraction [14], effectively enhancing the accuracy and efficiency of knowledge graph construction. Nevertheless, since knowledge graphs are dynamically updated semantic networks, achieving efficient knowledge updating and completion remains an important research topic.

In recent years, Large Language Model (LLM) have made breakthrough advancements in natural language understanding, text generation, and contextual reasoning [15]. The powerful semantic parsing capabilities of LLM overcome the limitations of traditional rule-based/statistical models in handling professional terminology, implicit relationships, and long-text dependencies, significantly improving the automated extraction efficiency of entities, attributes, and complex relationships from unstructured data. Large Language Models (LLM) can generate triples dynamically. They do this through techniques like zero-shot prompting and hybrid frameworks. This capability lowers the barrier to adapting them to new domains. By integrating Retrieval-Augmented Generation (RAG) technology, external knowledge sources are introduced to suppress domain-specific “hallucinations,” ensuring knowledge reliability. Furthermore, leveraging frameworks like GraphRAG, context-aware knowledge graphs are constructed, enhancing multi-hop reasoning capabilities in complex scenarios and promoting deep cognitive synergy between LLM and KG [16]. Currently, the integration of large language models and knowledge graphs is being applied in domain-specific question-answering systems, personalized recommendations, and other areas [17].

2.2. Application Research of Knowledge Graphs in Water Conservancy Facility Safety

Applying knowledge graph technology to the field of water conservancy facility safety holds immense potential for integrating fragmented knowledge and enhancing risk early warning and intelligent decision-making capabilities [18]. Water conservancy facility safety management involves multi-dimensional elements such as “institutions-equipment-hidden dangers-regulations” and their complex dynamic interactions, urgently requiring structured knowledge representation and reasoning support. Currently, research on knowledge graphs in the water conservancy domain remains in the exploratory stage [19]. Although general knowledge graph construction techniques (including knowledge extraction [20], fusion [21], and storage [22]) are relatively mature and have achieved significant results in fields such as finance [23], healthcare [24], and education [25], the water conservancy domain’s strong specialization and task-specific nature present unique challenges. Existing research on water conservancy facility safety knowledge management primarily focuses on:

emphasizing the importance of ontology construction—recognizing that building ontologies covering core concepts (e.g., institutions, equipment, hidden dangers, regulations) in the water conservancy field is a prerequisite for knowledge reuse and sharing [26], and utilizing tools (e.g., Protégé) for formal modeling; and addressing the difficulties in information extraction—acknowledging that water conservancy texts are highly unstructured and specialized, leading to challenges in domain-specific entity and relationship recognition [27].

Despite the groundbreaking advancements in LLM technology creating opportunities and demonstrating significant potential for the deep application of knowledge graphs in water conservancy facility safety, existing research on LLM-driven knowledge graph construction in this domain still faces critical limitations: First, general-purpose LLM exhibit insufficient understanding of domain-specific terminology and dynamic interactions between entities, leading to complex semantic parsing deviations [28]; second, the lack of robust domain-constrained mechanisms fails to effectively suppress knowledge “hallucinations,” compromising the reliability of professional knowledge; simultaneously, the semantic framework of ontologies and LLM extraction capabilities lack effective coordination, hindering accurate cross-source knowledge fusion; furthermore, current methods struggle to support real-time dynamic tracking of temporally evolving risks, limiting proactive early warning capabilities.

Therefore, this study focuses on addressing the aforementioned core challenges by proposing a synergistically enhanced knowledge graph construction method that integrates domain ontologies with large language models (LLM).

3. Methods

3.1. Overall Research Framework

Given the massive scale of water conservancy facilities and the critical safety responsibilities involved, facility failures could trigger severe consequences. However, safety knowledge in this domain is characterized by highly unstructured, multi-source, and heterogeneous data formats. Traditional technologies struggle to effectively integrate and model the complex dynamic relationships among “institutions-equipment-hidden dangers-regulations,” which constrains risk early warning and intelligent decision-making capabilities. Meanwhile, existing knowledge graphs lack depth in water conservancy safety applications, and large language models (LLM) face challenges such as professional “hallucinations” and instability in long-text comprehension. To address these challenges, this study proposes an LLM-enhanced knowledge graph construction method aimed at systematically resolving core issues of complex knowledge, unstructured data, and utilization difficulties in water conservancy facility safety. The main innovative contributions of this method are threefold: First, constructing a comprehensive ontology system covering “institutions-equipment-hidden dangers-regulations” to provide a unified and rigorous semantic framework for domain knowledge; Second, integrating LLM and GraphRAG technologies for efficient knowledge extraction, enabling precise extraction of triple knowledge from unstructured texts while introducing ontology templates to effectively suppress LLM hallucinations, thereby enhancing the model’s adaptability to specialized domains; Third, designing a knowledge graph architecture with dynamic update capabilities to support real-time updates of the graph.

This study followed the methodology described above. We constructed a water conservancy facility safety knowledge graph. This graph was built in the Neo4j graph database. We also developed an intelligent question-answering system. This system was created by integrating GraphRAG indexing with hybrid retrieval technology. The overall research framework is shown in Figure 1, which comprises three core modules: 1) Data Integration and Preprocessing: automated collection and cleansing of multi-source heterogeneous data; 2) Knowledge Graph Construction: defining the ontology model, utilizing LLM+GraphRAG technology for knowledge extraction, and implementing graph storage and visualization; 3) Question-Answering System Research: achieving multi-hop

reasoning and precise answer generation to serve practical scenarios such as hidden danger management and emergency decision-making.

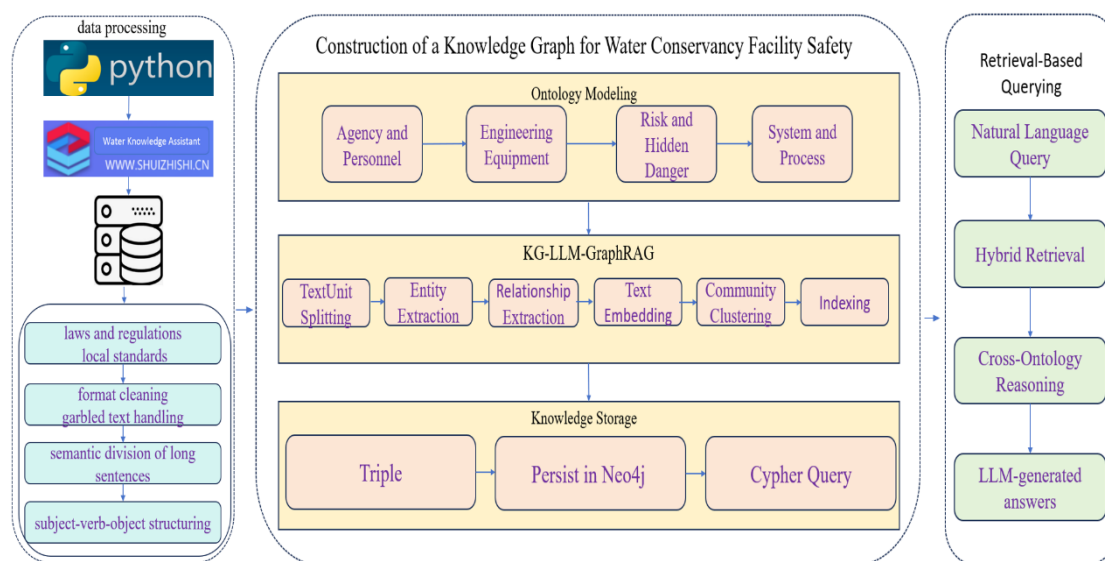


Figure 1. Construction framework of a water conservancy facility safety knowledge graph enhanced by large language models.

3.2. Model Construction Process

3.2.1. Multi-Source Heterogeneous Data Processing

Texts in the field of water conservancy facility safety are characterized by high degrees of unstructured content and specialization. Direct knowledge extraction often leads to suboptimal results due to semantic ambiguity and contextual deficiencies. Therefore, data preprocessing is essential to transform raw text into a normalized corpus with clear structure and unambiguous semantics, thereby laying the foundation for high-quality entity and relationship recognition in subsequent stages.

- It is necessary to remove formatting errors and garbled characters from the text, eliminate content that fails to meet quality and relevance requirements, while preserving the original format and expression forms of domain-specific materials [29]. This ensures the purity and consistency of the textual data.
- This step is performed using Python programs for regex-based cleansing.
- This phase involves manual curation, segmentation of long sentences, and removal of low-value information to provide robust data support for subsequent construction of the water conservancy facility safety knowledge graph.

3.2.2. Domain Ontology Modeling

Given that this study aims to achieve fine-grained representation of water conservancy facility safety knowledge, and considering the current lack of reusable ontology libraries, this paper comprehensively analyzes the content structure of data sources, systematically reviews a series of normative documents, and references existing cases and research to extract core elements. Consequently, a comprehensive water conservancy facility safety ontology library is constructed, with “Agency and Personnel,” “Engineering Equipment,” “Risks and Hidden Dangers,” and “Systems and Processes” as its core components. The structural representation of this ontology library is as follows:

$$O_{Wcf} = \{O_{Apo}, O_{Eeo}, O_{Rhd}, O_{Spo}, O_{Rbo}\} \quad (1)$$

In the formula,

- O_{Wcf} denotes the Comprehensive Water Conservancy Facility Safety Ontology;
- O_{Apo} represents the Agency and Personnel Ontology;
- O_{Eeo} refers to the Engineering Equipment Ontology;
- O_{Rhd} indicates the Risk and Hidden Danger Ontology;
- O_{Spo} signifies the System and Process Ontology;
- O_{Rbo} captures the Inter-Relationships among these ontologies.

This study integrates water conservancy facility safety management regulations and guidelines as primary knowledge sources for constructing the safety management ontology. Utilizing Stanford University's Protégé tool, we formally built the water conservancy facility safety knowledge ontology [30]. The constructed ontology comprises four major categories: Agency and Personnel, Engineering Equipment, Risks and Hidden Dangers, and Systems and Processes. Based on comprehensive consideration of conceptual logic, predefined semantic relationships between ontologies were established, including: "operates/is operated by", "executes/is executed by", "identifies/is identified by", "complies with/regulates", "triggers/affects", and "prevents/exposes".

3.2.3. Retrieval-Augmented Knowledge Extraction with Large Language Models

To address the limitations of traditional Retrieval-Augmented Generation (RAG) in the water conservancy facility safety domain, specifically its low retrieval precision [31] and poor intent alignment [32], this paper proposes a KG-LLM-GraphRAG architecture. This framework deeply integrates the structured semantic relationships of Knowledge Graphs (KG) with the generalized reasoning capabilities of Large Language Models (LLM), while incorporating a vector database as a supplementary retrieval source. The architecture supports natural language queries from users, automatically parses query intent, leverages graph relationships for multi-hop reasoning, and generates accurate, interpretable decision-support answers.

GraphRAG leverages large language models to extract and construct knowledge graphs from raw text, addressing the needs for cross-context retrieval and complex reasoning. Its construction process transforms unstructured text into a semantically enriched entity-relationship network, forming a structured knowledge representation that is more expressive than plain text segments. This significantly enhances the system's ability to handle complex multi-hop questions. During this process, a prompt-based optimization strategy guides the LLM to extract structured triples from unstructured text. By introducing an ontology layer-based contextual prompting method, the model's generalization capability in the water conservancy facility safety domain is effectively enhanced, improving the accuracy of entity and relationship extraction. Building upon the constructed water conservancy facility safety ontology, this study further designs tailored extraction templates and employs a contextual prompting mechanism to create standardized prompt information suitable for LLM processing. The construction workflow can be divided into the following key steps, as illustrated in Figure 2 below.

- Text Segmentation

Normative documents in the water conservancy domain are typically lengthy and complex, while large language models (LLM) have inherent input length limits. Direct input of full text would cause loss of critical information due to truncation, necessitating text segmentation. The original input documents are partitioned along semantic boundaries into manageable textual units (TextUnits), ensuring each text block carries a relatively complete semantic segment while avoiding fragmentation of entities or relationships. This process lays the foundation for subsequent in-depth parsing.

- Prompt Design And Entity-Relationship Extraction.
- Based on the partitioned TextUnits and leveraging the optimization strategy of contextual prompts, the large language model (LLM) is invoked for sequential processing to automatically identify and extract involved entities and their semantic relationships, generating preliminary triple structures.

- Knowledge Graph Construction And Community Generation.

The extracted entities and relationships are constructed into a preliminary knowledge graph. Subsequently, graph clustering algorithms are introduced to aggregate semantically closely-related entities and relationships into communities, thereby revealing knowledge substructures and their intrinsic connections. Furthermore, a structured summary report is generated for each community, outlining its core themes, key entities, relationships, and their influence within the overall knowledge network.

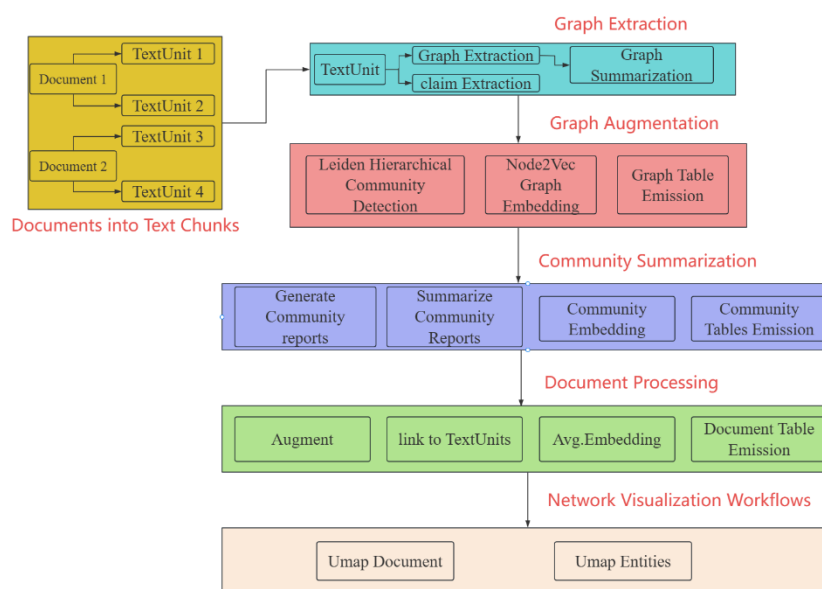


Figure 2. Construction workflow of GraphRAG.

3.2.4. Graph Database Storage and Visualization

Based on the constructed triples [33], Cypher statements are employed to batch import them into the Neo4j graph database, achieving efficient knowledge storage. In the graph database, entities, relationships, and attributes are mapped to nodes and edges respectively, collectively forming a semantically rich topological network structure [34]. Compared with relational databases, graph databases demonstrate significant advantages in storing topologically interconnected data. Their native graph structure supports efficient multi-hop queries and complex relational reasoning, making them particularly suited for the application requirements of water conservancy facility safety knowledge graphs. This provides a reliable foundation for knowledge organization and storage in this vertical domain.

4. Construction of Water Conservancy Facility Safety Knowledge Graph

4.1. Data Sources

This study selects representative multi-source normative texts from the water conservancy facility safety domain as the core data foundation for constructing the knowledge graph. The partial data adopted (see Table 1) covers multiple levels and types including national laws, administrative regulations, departmental rules, local regulations, national standards, and local standards. Key examples include the “Water Law of the People’s Republic of China” and the “Regulations on Work Safety Management of Water Conservancy Projects”. The content comprehensively covers institutional responsibilities, engineering equipment management, risk hazard identification, and emergency response procedures. These authoritative and structurally diverse texts collectively form the core knowledge source for water conservancy facility safety management, providing reliable and diversified data support for the deep integration of semantic understanding and structured knowledge reasoning.

Table 1. Statistics of Data Sources for Water Conservancy Facility Safety Knowledge Graph (Partial).

No.	Name	Category	Level/Source
S01	Water Law of the People's Republic of China	National Law	National People's Congress
S02	Flood Control Law of the People's Republic of China	National Law	National People's Congress
S03	Regulations on the Safety Management of Reservoir Dams	Administrative Regulation	State Council
S04	Provisions on Work Safety Management of Water Conservancy Projects	Departmental Rule	Ministry of Water Resources
S05	Provisions on Quality Management of Water Conservancy Projects	Departmental Rule	Ministry of Water Resources
S06	Sichuan Province Water Conservancy Engineering Management Regulations	Local Regulation	Sichuan People's Congress
S07	Chongqing City Water Conservancy Engineering Management Regulations	Local Regulation	Chongqing People's Congress
S08	GB/T 40582-2021 Basic Terminology for Hydropower Stations	National Standard	Standardization Administration
S09	DB11/T 2193-2023 Specification for Investigation and Management of Flood Prevention Hidden Dangers—Water Conservancy Projects	Local Standard	Beijing Municipality
S10	Guide to the List of Major Hidden Dangers for Production Safety in Water Conservancy Projects (2021 Edition)	Departmental Normative Document	Ministry of Water Resources
S11	Standard for Post Setting of Water Conservancy Project Management Units (Pilot) and Quota Standard for Water Conservancy Project Maintenance (Pilot)	Departmental Normative Document	Ministry of Water Resources
S12	Measures for the Assessment and Management of Work Safety for Principal Responsible Persons, Project Responsible Persons, and Full-time Work Safety Management Personnel of Water Conservancy and Hydropower Construction Enterprises	Departmental Normative Document	Ministry of Water Resources
S13	Guide for Identification and Risk Assessment of Operational Hazard Sources for Water Conservancy and Hydropower Projects (Reservoirs, Sluices) (Trial)	Departmental Normative Document	Ministry of Water Resources
S14	Guide to the List of Major Hidden Dangers for Production Safety in Water Conservancy Projects (2023 Edition)	Departmental Normative Document	Ministry of Water Resources
S15	Guidelines for Risk Assessment of Dams (ICOLD)	International Organization Guide	International Commission on Large Dams (ICOLD)
S16	Hebei Province Water Conservancy Engineering Management Regulations	Local Regulation	Hebei People's Congress

4.2. Knowledge Graph Construction

4.2.1. Construction of Water Conservancy Facility Safety Ontology

Based on the ontology layer, the comprehensive water conservancy facility safety ontology library O_{Wcf} is constructed. This system framework explicitly defines four core ontologies: "Agency and Personnel" O_{Apo} , "Engineering Equipment" O_{Eeo} , "Risks and Hidden Dangers" O_{Rhd} , "Systems and Processes" O_{Spo} , and presets six categories of semantic relationships O_{Rbo} : "operates/is operated by", "executes/is executed by", "identifies/is identified by", "complies with/regulates", "triggers/affects", and "prevents/exposes".

Firstly, the specific construction process of the four core ontologies is as follows:

- (1) Regarding the construction of the Agency and Personnel Ontology: At the institutional level, entities are classified into five categories based on hierarchical relationships, business scenarios, and emergency roles: government regulatory departments, project management units, construction and operation enterprises, technical support agencies, and emergency coordination agencies. At the personnel level, individuals are categorized into five types according to qualification constraints, duty associations, and cross-industry mappings: unit responsible persons, technical management personnel, operation and maintenance staff, administrative support personnel, and emergency response personnel. Given the complex and dynamic nature of organizational arrangements for institutions and personnel, this paper constructs separate models for institutions and personnel based on this classification to enhance the targeting and efficiency of institutional and personnel arrangements in water conservancy projects. Table 2 demonstrates the construction of the institution ontology using government regulatory departments as an example, while Table 3 illustrates the personnel ontology construction with technical management personnel as an example.

Table 2. Conceptual Hierarchy of Water Conservancy Agency Ontology – Example of Government Regulatory Agencies.

Level 1 Concept	Level 2 Concept	Instances
Government Regulatory Agencies	National Regulatory Agencies	Ministry of Water Resources, Ministry of Emergency Management, Ministry of Finance
	Provincial Regulatory Agencies	Provincial Water Resources Department, Provincial Emergency Management Department, Provincial Finance Department
	Municipal Regulatory Agencies	Municipal Water Resources Bureau, Municipal Emergency Management Bureau, Municipal Finance Bureau
	County Regulatory Agencies	County Water Resources Bureau, County Emergency Management Bureau, County Finance Bureau
	Inter-basin Management Agencies	River Basin Management Agencies, Regional Coordination Agencies
	Specialized Regulatory Agencies	Hydrology Bureau, Water Conservancy Project Quality Supervision Station, Water Administration Supervision Detachment

Table 3. Conceptual Hierarchy of Water Conservancy Personnel Ontology – Example of Technical Management Personnel.

Level 1 Concept	Level 2 Concept	Instances
Technical Management Personnel	Safety Engineer	Beiyun River Levee and Gate Safety, Yangzhuang Reservoir Water Quality Collaborative Management

Quality Supervisor	Chaobai River Levee Project Quality Control, Pipe Material Quality Dispute Handling, Ad-hoc Quality Inspection During Flood Season Construction
Hydrological Monitor	Jiyun River Salt-Fresh Water Interaction Monitoring, Storm Surge Red Warning Response

- (2) Regarding the construction of the Engineering Equipment Ontology: Based on ISO 55000 (Asset Management Standards) and water conservancy engineering systems theory, core concepts mentioned in various standard documents are refined. Engineering equipment is classified into five categories: water-retaining engineering equipment, water-discharging engineering equipment, water-diversion engineering equipment, monitoring and control engineering equipment, and auxiliary engineering equipment. Table 4 demonstrates the construction of the engineering equipment ontology.

Table 4. Conceptual Hierarchy of Engineering Equipment Ontology.

Concept Classification	Instances
Water-Retaining Engineering Equipment	Dam, Levee, Gate
Water-Discharging Engineering Equipment	Spillway, Flood Discharge Tunnel, Drainage Valve
Water-Diversion Engineering Equipment	Diversion Channel, Pipeline, Pump Station
Monitoring and Control Engineering Equipment	Water Level Sensor, Stress Monitor, SCADA System
Auxiliary Engineering Equipment	Hoist, Trash Rake, Emergency Power Supply

- (3) Regarding the construction of the Risk and Hidden Danger Ontology: Based on disaster chain theory, the evolution of hidden dangers exhibits temporality, requiring distinction between latent, trigger, and outbreak phases. It is worth noting that these three phases of hidden danger evolution occur sequentially: the latent phase accumulates risks, the trigger phase is initiated by external conditions, and ultimately leads to disasters in the outbreak phase. Table 5 demonstrates the construction of the risk and hidden danger ontology.

Table 5. Conceptual Hierarchy of Risk and Hidden Danger Ontology.

Phase	Characteristics	Instances
Latent	Hidden danger exists but not triggered	Concrete Carbonation, Metal Fatigue
Trigger	External conditions exceed critical threshold	Water Level Exceeds Warning Line, Peak Ground Acceleration Exceeds Limit
Outbreak	System instability leads to disaster	Dam Breach, Pipeline Burst

- (4) Regarding the construction of the system and process ontology, based on legal hierarchy, effectiveness, and management process stages, system processes are categorized into three types according to legal hierarchy: national laws, administrative regulations and departmental rules, and local regulations. It is important to note that logical consistency in classification must be maintained, avoiding overlaps or omissions. Table 6 demonstrates the construction of the system and process ontology.

Table 6. Conceptual Hierarchy of System and Process Ontology.

Concept Classification	Instances
National Laws	Water Law of the People's Republic of China, Flood Control Law of the People's Republic of China
Administrative Regulations and Departmental Rules	Regulations on the Safety Management of Reservoir Dams, Provisions on Work Safety Management of Water Conservancy Projects, Provisions on Quality Management of Water Conservancy Projects

Local Regulations	Sichuan Province Water Conservancy Engineering Management Regulations, Chongqing City Water Conservancy Engineering Management Regulations
-------------------	--

Secondly, the definition of the six types of semantic relationships. After completing the construction of the four core ontologies, it is necessary to further clarify the semantic associations among them to form a knowledge network capable of supporting complex relational reasoning. Based on the inherent logic and interaction patterns among “institutions, equipment, hidden dangers, and regulations” in water conservancy facility safety management, this paper predefines six pairs of top-level relationship sets. Table 7 illustrates the definition of these six types of top-level semantic relationships.

Table 7. Definition of Six Types of Top-Level Semantic Relationships.

Top-Level Semantic Relation	Integrated Similar Expressions
Operates / Is Operated By	Uses / Is Used By, Controls / Is Controlled By, Manages / Is Managed By, Manipulates / Is Manipulated By, Runs / Is Run By, Operates / Is Controlled By
Executes / Is Executed By	Implements / Is Implemented By, Carries Out / Is Carried Out By, Fulfills / Is Fulfilled By, Performs / Is Performed By, Executes / Is Commanded By, Responsible For / Is Responsibility Of
Identifies / Is Identified By	Discovers / Is Discovered By, Detects / Is Detected By, Monitors / Is Monitored By, Diagnoses / Is Diagnosed By, Determines / Is Determined By, Assesses / Is Assessed By
Complies With / Regulates	Obeys / Is Obeyed By, Based On / Is Basis For, Conforms To / Is Conformed To, Follows / Is Followed By, Regulates / Is Regulated By, Constrains / Is Constrained By
Triggers / Affects	Causes / Is Caused By, Induces / Is Induced By, Activates / Is Activated By, Results In / Is Resulted In By, Affects / Is Affected By, Exacerbates / Is Exacerbated By
Prevents / Exposes	Prevents / Is Prevented By, Avoids / Is Avoided By, Mitigates / Is Mitigated By, Controls / Is Controlled By, Exposes / Is Exposed By, Reveals / Is Revealed By

In summary, this paper systematically constructs a comprehensive ontology library for the water conservancy facility safety domain, which includes four core ontologies and six predefined semantic relationships among them. This provides a rigorous semantic framework and structural constraints for the subsequent accurate knowledge extraction and graph construction based on LLM.

4.2.2. LLM-Integrated Prompt Engineering and Ontology-Constrained Entity-Relationship Extraction

After completing the construction of the domain ontology, we follow a structured semantic framework and carry out systematic operational procedures to accurately extract entities and relationships from unstructured texts, thereby completing the construction of the knowledge graph. The specific processing flow is as follows:

- Text Segmentation

The text must be segmented to avoid fragmenting key entities or relationships. This is done by dividing the text into segments called TextUnits. The segmentation follows semantic boundaries. This process ensures that each TextUnit contains a complete semantic fragment.

- Prompt Design And Entity-Relationship Extraction

On the basis of text segmentation, a prompting strategy is employed to guide the large language model in extracting entities and relationships from unstructured text, thereby generating structured triples. Based on the water conservancy facility safety ontology, tailored extraction templates are constructed, and a contextual prompting method is integrated to design standardized prompt information for input into the large language model. The specific prompt content is illustrated in Figure 3.

```

-Goal-
Suppose you are an expert in the field of water system and water conservancy engineering, and adopt a step-by-step approach to identify all entities and relationships from the given text, directly extracting the text content. Your answer should not include the thought process. The extracted results can only be output according to the template, and the content in the template is output based on the prompts.
-Steps-
1. Identify all entities. For each identified entity, extract the following information:
- entity_name: Name of the entity, capitalized
- entity_type: One of the following types: [Agency and Personnel, Engineering Equipment, Risks and Hidden Dangers, Systems and Processes]
- entity_description: Comprehensive description of the entity's attributes and activities
Format each entity as ("entity"(tuple_delimiter)<entity_name>(tuple_delimiter)<entity_type>(tuple_delimiter)<entity_description>)
2. From the entities identified in step 1, identify all pairs of (source_entity, target_entity) that are *clearly related* to each other.
For each pair of related entities, extract the following information:
- source_entity: name of the source entity, as identified in step 1
- target_entity: name of the target entity, as identified in step 1
- relationship_description: explanation as to why you think the source entity and the target entity are related to each other
- relationship_strength: an integer score between 1 to 10, indicating strength of the relationship between the source entity and target entity
Format each relationship as ("relationship"(tuple_delimiter)<source_entity>(tuple_delimiter)<target_entity>(tuple_delimiter)<relationship_description>(tuple_delimiter)
<relationship_strength>)
3. Return output in English
It should be noted that there are some possible non - standard sentences or typos in the text, but overall, the text is mainly composed of English words and follows English grammar and expression patterns. as a single list of all the entities and relationships identified in steps 1 and 2. Use **{record_delimiter}** as the list delimiter.
4. If you have to translate into English
It should be noted that there are some possible non - standard sentences or typos in the text, but overall, the text is mainly composed of English words and follows English grammar and expression patterns. just translate the descriptions, nothing else!
5. When finished, output {completion_delimiter}.

```

Figure 3. Prompt template content.

After completing the design of the prompt, a large model (such as doubao-1-5-pro-32k-250115) is used to extract the entities therein. These entities typically refer to information such as persons, locations, institutions, and concepts appearing in the document. The purpose of entity recognition is to construct an Entity Graph, extracting all entities and preparing for subsequent relationship mining and queries.

On the basis of identifying entities, the text units are further analyzed to mine the semantic relationships existing between entities. These relationships define how entities are interrelated.

- Knowledge Graph Construction And Community Generation

Based on the extracted entities and their relationships, an initial knowledge graph is constructed. Graph clustering algorithms are then applied to group closely related entities and relationships into different communities. For each community, a community report is generated, summarizing the community's core themes, key entities and relationships contained within, and the community's importance or influence within the overall knowledge network.

By combining the contextual prompts of the large language model with ontological constraints, the accuracy and consistency of entity and relationship extraction by the LLM in the field of water conservancy safety are effectively enhanced, laying a reliable data foundation for the subsequent construction of the knowledge graph.

4.3. Model Performance

To systematically validate the feasibility and effectiveness of the proposed method for constructing a water conservancy facility safety knowledge graph, this study evaluates the approach from two dimensions: ontology construction quality and knowledge extraction performance. First, regarding ontology construction, experts in the water conservancy field and experienced frontline practitioners were invited to conduct a professional review of the constructed ontology [35]. By assessing the rationality of the ontology structure, the necessity of core concepts, and the similarity and distinction between concepts, it was unanimously concluded that the ontology demonstrates validity and feasibility in terms of semantic coverage and logical consistency. Second, in terms of knowledge extraction effectiveness, precision (P), recall (R), and F1-score were used to quantitatively evaluate the entity and relationship extraction results driven by the large language model [36]. Here,

precision measures the proportion of true positive instances among the predicted positive cases, recall reflects the proportion of true positive instances correctly predicted, and the F1-score, as the harmonic mean of the two, comprehensively evaluates the overall performance of the model. The average F1-score, calculated as the arithmetic mean of the F1-scores across all categories, is used to assess the model's generalization capability across different types of entities and relationships. To ensure the reliability of the results, all extraction experiments were repeated three times, with the average values taken as the final performance metrics. By comparing the results with actual safety assessment reports, the P, R, and F1 values for the entity and relationship extraction tasks were calculated.

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (4)$$

In the provided formulas, TP (True Positive), FP (False Positive), TN (True Negative), and FN (False Negative) represent the four possible outcomes of the classifier's target identification. When describing these four outcomes for the classifier's target identification, it is assumed that there are only two categories: Positive and Negative.

Table 8. Model Evaluation Results.

Prediction Actual	Actual Positive	Actual Negative
Predicted Positive	TP	TN
Predicted Negative	FP	FN

Among these metrics, TP (True Positive) refers to instances where the model correctly predicts positive cases as positive; FP (False Positive) refers to instances where the model incorrectly predicts negative cases as positive; TN (True Negative) refers to instances where the model correctly predicts negative cases as negative; and FN (False Negative) refers to instances where the model incorrectly predicts positive cases as negative.

To validate the feasibility of the proposed method, this study designed and conducted three sets of comparative experiments. These experiments employed different extraction methods to comprehensively evaluate the performance and effectiveness of each approach. The first experiment used a widely adopted large language model for direct information extraction. This method relies on the model's pre-trained knowledge and capabilities, generating extraction results directly from the input text. The second experiment introduced a template prompting method, which involves providing example templates manually extracted from a small number of samples to enhance the model's understanding of specific domain contexts. The third experiment adopted the context-integrated method proposed in this study, which combines prompt technology with ontological knowledge for information extraction. This approach guides the model through designed prompts and incorporates domain ontology knowledge to improve the accuracy and interpretability of the extraction results.

First, no prompting method was used during extraction, and the large model was directly applied for extraction. Second, without using the ontological template, extraction was performed directly using manually designed templates. Finally, both the ontological template and the prompt method were input into the large language model to obtain the extraction results. The comparative results of the three experiments are shown in Table 9.

From the table, it can be observed that the strategy combining the ontological template with the prompt method significantly outperforms the other two extraction methods. Commonly used large language model content extraction methods often face challenges such as insufficient semantic consistency and unclear expression of complex relationships when processing massive data. In contrast, the method proposed in this study, by constructing a knowledge graph and combining it

with ontology-driven large model knowledge extraction, can more accurately capture complex entity relationships and semantic information. This approach not only addresses the shortcomings of existing direct extraction and template extraction methods but also significantly improves extraction efficiency and the completeness of information expression, demonstrating notable superiority. This indicates that large language models possess strong learning capabilities and can effectively capture useful information or features from templates and prompts. These results also confirm that the combined use of ontological templates and prompt-based recognition is entirely feasible.

Table 9. Entity and Relationship Extraction Results.

Type	Precision (P)	Recall (R)	F1
Direct Extraction	0.435	0.560	0.490
Template Extraction	0.643	0.728	0.683
Prompt + Ontology	0.840	0.948	0.891

5. Knowledge Graph Visualization and Application

Based on the ontological framework, this study aligns the entities and entity attribute data identified by the large language model with the entities and entity attributes in the ontological layer to complete the matching of entities and attributes. This process enables the construction of all entities, entity attributes, and relationships, ultimately forming a complete knowledge graph [37]. The constructed knowledge graph can be visually displayed in the Neo4j graph database.

The types of graph nodes are consistent with those in the ontology and can be broadly categorized into four major classes: agencies and personnel, engineering equipment, risks and hidden dangers, and systems and processes. The relationships can be divided into six pairs: “operates/is operated by,” “executes/is executed by,” “identifies/is identified by,” “complies with/regulates,” “triggers/affects,” and “prevents/exposes.” Specifically, the relationship between personnel and engineering equipment should be an operational relationship; the relationship between agencies and personnel and systems and processes should be an execution relationship; the relationship between agencies and personnel and risks and hidden dangers should be an identification relationship; the relationship between engineering equipment and systems and processes should be a compliance and regulation relationship; the relationship between engineering equipment and risks and hidden dangers should be a triggering and affecting relationship; and the relationship between systems and processes and risks and hidden dangers should be a prevention and exposure relationship. The completed safety knowledge graph for water conservancy facilities, using the Huangzhuangwa Flood Diversion Sluice as an example, is shown in Figure 4. The extracted entity information is shown in Figure 5.

For example, to understand the newly built ancillary structures in the hazard mitigation and reinforcement project of the Huangzhuangwa Flood Diversion Sluice, one only needs to perform the corresponding query operation in the Neo4j graph database to retrieve the data. The results, as shown in Figure 6, indicate that the newly built ancillary structure in the Huangzhuangwa Flood Diversion Sluice hazard mitigation and reinforcement project is a management building.

To gain a clearer understanding of the data chain, one can also query the entire graph data chain containing the “field safety inspection” node to obtain relevant information about the field safety inspection, as shown in Figure 7.

The graph database, built on the graph structure, can clearly display the relationships between nodes through its inherent graph architecture[38]. Taking the field safety inspection as the starting point of the data chain, upward reasoning reveals that the Huangzhuangwa Flood Diversion Sluice was required to undergo field safety inspection. Downward reasoning indicates that the inspection aimed to identify concrete structural appearance defects (such as damage, cracks, exposed reinforcement, etc.), providing a more intuitive understanding of the overall facility management. The complete knowledge description of the facility management process is: “During the field safety

Figure 5. Entity information.

```

{
  "identity": 41,
  "labels": [
    "Entity",
    "Equipment"
  ],
  "properties": {
    "human_readable_id": 13,
    "text_unit_ids": [
      "227a6616d0e42014c006e18b8230cff917bb5983e6661a7bf07d00561934970686bc37e5bf64185940934b1c192b382620168cba2f7ee4bbc33efed73a163193"
    ],
    "description": "Ancillary facilities built during the Huangzhuangwa flood diversion gate removal and reinforcement project, which support the normal operation and management of the gate  
Ancillary facilities built during the Huangzhuangwa flood diversion gate removal and reinforcement project",
    "id": "46f1e61b-a1ed-4a54-98bc-6f0a45c877fb",
    "category": "Institutional personnel",
    "type": "EQUIPMENT",
    "title": "MANAGEMENT ROOMS AND OTHER ANCILLARY FACILITIES"
  },
  "elementId": "4:593568d8-ec9b-4f7e-926c-501b2c97a447:41"
}

```

Figure 6. Query results.

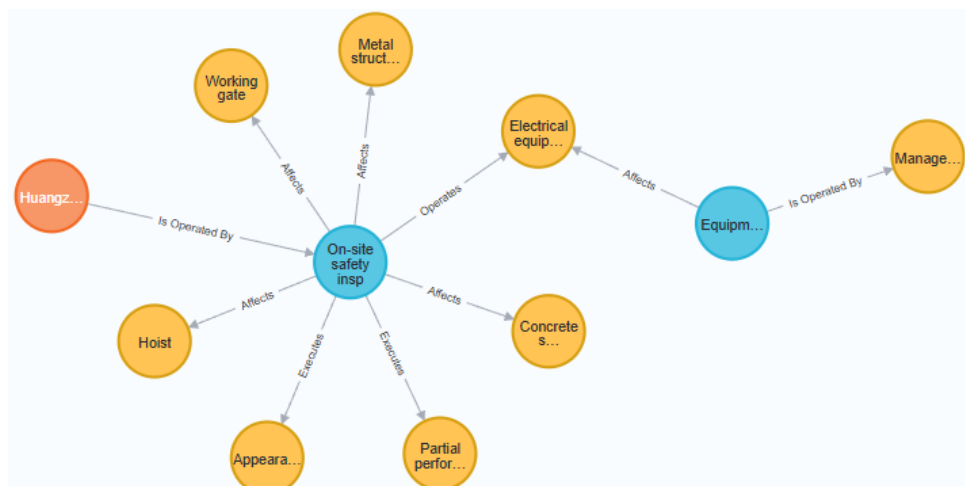


Figure 7. Data chain query.

6. Conclusions

To address issues such as multi-source heterogeneous data barriers and difficulties in expressing complex entity relationships in water conservancy facility safety management, this paper proposes a novel method for constructing a safety knowledge graph for water facilities based on large language models (LLM), and conducts research and discussions on related safety management issues. By introducing an ontological framework, the recognition accuracy of LLM in specialized domains has been effectively improved. To tackle the complexity of entity relationships, LLM are utilized to significantly enhance the identification capability of entities and relationships in water facility data. The constructed knowledge graph provides support for establishing an intelligent safety management system, contributing to a comprehensive improvement in safety management levels.

This study still has certain limitations: Firstly, although domain ontology constraints were introduced to enhance the extraction performance of LLM, the accuracy of entity and relationship extraction in extremely complex contexts (such as multiple nested relationships and professional terminology ambiguity) still needs improvement. Future improvements should focus on enhancing the model's precision in understanding implicit semantics and long-text dependencies to ensure the completeness and consistency of the knowledge graph. Secondly, the current research primarily

relies on textual normative documents for knowledge extraction and graph construction, and has not yet integrated multimodal data sources such as images, real-time sensor data, and geographic information systems (GIS). Additionally, the dynamic update mechanism for the knowledge graph has not been systematically implemented, resulting in limitations in the comprehensiveness and timeliness of knowledge types in the constructed graph, which restricts its completeness and evolutionary capability. This consequently affects its supporting effectiveness in practical scenarios such as proactive early warning and emergency response. Future research will focus on exploring deep integration mechanisms for multi-source heterogeneous data, incorporating non-textual data such as images, real-time monitoring data, and social sensing information to construct a more comprehensive and dynamically evolving knowledge graph.

Author Contributions: All the authors contributed to the study conception and design. Literature curation and investigation, Yu Wang, Xinhua Cui and Chenyu Zhu; Formal analysis and synthesis, Cui Li and Xinhua Cui; Writing original draft preparation, Yu Wang and Cui Li; Writing review and editing, Cui Li, Xinhua Cui and Chenyu Zhu; Supervision and project administration, Cui Li. All the authors read and approved the final manuscript.

Funding: This work was supported by the Langfang Science and Technology Support Program Project and the Science(2023013202) and Technology Innovation Program for Postgraduates at the IDP subsidized by the Fundamental Research Funds for the Central Universities (ZY20260319).

Conflicts of Interest: The authors have no relevant financial or nonfinancial interests to disclose.

References

1. Wang, Y.; Hu, A. China's Water Conservancy: Review and Outlook (1949–2050). *J. Tsinghua Univ. (Philos. Soc. Sci.)* 2011, *26*, 99–112. <https://doi.org/10.13613/j.cnki.qhdz.002012>.
2. Ge, W. Editorial: Risk assessment and management of water conservancy projects. *Front. Earth Sci.* 2023, *11*, 1330621. <https://doi.org/10.3389/feart.2023.1330621>.
3. Liu, Y.; Tang, Y.; Jing, L.; Chen, F.; Wang, P. Remote Sensing-Based Dynamic Monitoring of Immovable Cultural Relics, from Environmental Factors to the Protected Cultural Site: A Case Study of the Shunji Bridge. *Sustainability* 2021, *13*, 6042. <https://doi.org/10.3390/su13116042>.
4. Lu, J.; Feng, J.; Tang, Z.; Zhang, P. Research on Key Technologies of Water Conservancy Big Data Directory Service and Resource Sharing. *Water Resour. Inform.* 2017, *4*, 17–20+27. 10.19364/j.1674-9405.2017.04.004.
5. Qiu, L.; Zhang, A.; Li, S.; Zhang, Y.; Shen, M.; Zhou, P. A Review on Knowledge Graph Construction in Aviation Manufacturing. *Appl. Res. Comput.* 2022, *39*, 968–977. 10.19734/j.issn.1001-3695.2021.09.0367.
6. Huang, Y.; Yu, S.; Luo, B.; Li, R.; Li, C.; Huang, W. Exploring the Digital Twin Yangtze River for Joint Intelligent Scheduling of Basin Water Engineering Disaster Prevention. *J. Hydraul. Eng.* 2022, *53*, 253–269. 10.13243/j.cnki.slx.20210865.
7. Chen, Y.; Zhang, T.; Niu, W.; Qin, H. Research on Key Technologies for Digital Twin Construction of the Three Gorges Reservoir Area. *Yangtze River* 2023, *54*, 19–24. 10.16232/j.cnki.1001-4179.2023.08.003.
8. Xie, A.; Wu, Q.; Liu, F. Exploring Intelligent Operation and Maintenance Approaches for Pump Station Projects Based on Voiceprint Recognition and Knowledge Graph Technology. *Yangtze River Technol. Econ.* 2021, *5*, 88–92. 10.19679/j.cnki.cjjsj.2021.0617.
9. Huang, H.; Yu, J.; Liao, X.; Xi, Y. A Survey of Knowledge Graph Research. *Comput. Syst. Appl.* 2019, *28*, 1–12. 10.15888/j.cnki.csa.006915.
10. Zhou, Y.; Liu, Z.; Su, X.; Jin, T. Construction of a Q&A Knowledge Graph Ontology Model Integrating Multi-level Data. *Lib. Inf. Serv.* 2022, *66*, 125–132. 10.13266/j.issn.0252-3116.2022.05.013.
11. Ibrahim, N.; Aboulela, S.; Ibrahim, A.; Kashef, R. A survey on augmenting knowledge graphs (KGs) with large language models (LLMs): models, evaluation metrics, benchmarks, and challenges. *Discov. Artif. Intell.* 2024, *4*, 76. 10.1007/s44163-024-00175-8.
12. Lv, W.; Liao, Z.; Liu, S.; Zhang, Y. MEIM: A Multi-source Software Knowledge Entity Extraction Integration Model. *Comput. Mater. Contin.* 2020, *66*, 1027–1042. 10.32604/cmc.2020.012478.

13. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* 2021, *32*, 4–24. 10.1109/TNNLS.2020.2978386.
14. Chen, W.; Tian, J.; Xiao, L.; He, H.; Jin, Y. Exploring Logically Dependent Multi-task Learning with Causal Inference. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 16–20 November 2020; pp. 2213–2225. 10.18653/v1/2020.emnlp-main.173.
15. Zhao, W.X.; et al. A Survey of Large Language Models. *arXiv* 2023, arXiv:2303.18223. 10.48550/arXiv.2303.18223.
16. Pan, S.; Luo, L.; Wang, Y.; Chen, C.; Wang, J.; Wu, X. Unifying Large Language Models and Knowledge Graphs: A Roadmap. *IEEE Trans. Knowl. Data Eng.* 2024, *36*, 3580–3599. 10.1109/tkde.2024.3352100.
17. Chen, H.; Xie, R.; Cui, X.; Yan, Z.; Wang, X.; Xuan, Z.; Zhang, K. LKPNR: Large Language Models and Knowledge Graph for Personalized News Recommendation Framework. *Comput. Mater. Contin.* 2024, *79*, 10.32604/cmc.2024.049129.
18. Liu, X.; Lu, H.; Li, H. Intelligent generation method of emergency plan for hydraulic engineering based on knowledge graph---take the South-to-North Water Diversion Project as an example. *LHB* 2022, *108*, 2153629. 10.1080/27678490.2022.2153629.
19. Duan, H.; Han, K.; Zhao, H.; Jiang, Y.; Li, H.; Mao, W. Research on the Construction of Comprehensive Water Conservancy Knowledge Graph. *J. Hydraul. Eng.* 2021, *52*, 948–958. 10.13243/j.cnki.slxh.20200924.
20. Abdullah, M.H.A.; Aziz, N.; Abdulkadir, S.J.; Alhussian, H.S.A.; Talpur, N. Systematic Literature Review of Information Extraction From Textual Data: Recent Methods, Applications, Trends, and Challenges. *IEEE Access* 2023, *11*, 10535–10562. 10.1109/ACCESS.2023.3240898.
21. Zhang, J.; Zhang, X.; Wu, C.; Zhao, Z. A Survey of Knowledge Graph Construction Technology. *Comput. Eng.* 2022, *48*, 23–37. 10.19678/j.issn.1000-3428.0061803.
22. Liu, S.; Yang, H.; Li, J.; Kolmanič, S. Preliminary Study on the Knowledge Graph Construction of Chinese Ancient History and Culture. *Information* 2020, *11*, 186. 10.3390/info11040186.
23. Wang, W.; Xu, Y.; Du, C.; Chen, Y.; Wang, Y.; Wen, H. Data Set and Evaluation of Automated Construction of Financial Knowledge Graph. *Data Intell.* 2021, *3*, 418–443. 10.1162/dint_a_00108.
24. Abu-Salih, B.; AL-Qurishi, M.; Alweshah, M.; AL-Smadi, M.; Alfayez, R.; Saadeh, H. Healthcare knowledge graph construction: A systematic review of the state-of-the-art, open issues, and opportunities. *J. Big Data* 2023, *10*, 81. 10.1186/s40537-023-00774-9.
25. Dang, F.-R.; Tang, J.-T.; Pang, K.-Y.; Wang, T.; Li, S.-S.; Li, X. Constructing an Educational Knowledge Graph with Concepts Linked to Wikipedia. *J. Comput. Sci. Technol.* 2021, *36*, 1200–1211. 10.1007/s11390-020-0328-2.
26. Cheng, Q.; Wang, J.; Lu, W.; Huang, Y.; Bu, Y. Keyword-citation-keyword network: a new perspective of discipline knowledge structure analysis. *Scientometrics* 2020, *124*, 1923–1943. 10.1007/s11192-020-03576-5.
27. Lin, J.; Zhao, Y.; Huang, W.; Liu, C.; Pu, H. Domain knowledge graph-based research progress of knowledge representation. *Neural Comput. Appl.* 2021, *33*, 681–690. 10.1007/s00521-020-05057-5.
28. Tariq, A.; et al. Domain-specific LLM Development and Evaluation---A Case-study for Prostate Cancer. *J. Biomed. Inform.* 2024, *154*, 104650. 10.1101/2024.03.15.24304362.
29. SENTiVENT: enabling supervised information extraction of company-specific events in economic and financial news. *Lang. Resour. Eval.* 2020, *54*, 1077–1106. <https://link.springer.com/article/10.1007/s10579-021-09562-4>
30. Shu, X.; Yang, H. Ontology-driven intelligent assessment system for dam structural safety based on spatiotemporal anomaly detection framework. *Comput.-Aided Civ. Infrastruct. Eng.* 2025, *40*, 1–20. 10.1111/mice.70115.
31. Ong, Q.C.; et al. Advancing health coaching: A comparative study of large language model and health coaches. *Artif. Intell. Med.* 2024, *157*, 103004. 10.1016/j.artmed.2024.103004.
32. Sawarkar, K.; Mangal, A.; Solanki, S.R. Blended RAG: Improving RAG (Retriever-Augmented Generation) Accuracy with Semantic Search and Hybrid Query-Based Retrievers. In Proceedings of the 2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval (MIPR), Orlando, FL, USA, 15–17 August 2024; pp. 155–161. 10.1109/MIPR62202.2024.00031.

33. Li, J.; Hu, J.; Zhang, G. Enhancing Relational Triple Extraction in Specific Domains: Semantic Enhancement and Synergy of Large Language Models and Small Pre-Trained Language Models. *CMC-Comput. Mater. Contin.* 2024, *79*, 2481–2503. 10.32604/cmc.2024.050005.
34. Amit, A.; Chakraborty, R.; Patra, M.R.; Jadhav, A. Data profiling in property graph databases. *J. Supercomput.* 2023, *79*, 20056–20073.10.1007/s41060-023-00454-0.
35. Jiang, L.; Shi, J.; Wang, C. Multi-ontology fusion and rule development to facilitate automated code compliance checking using BIM and rule-based reasoning. *Adv. Eng. Inform.* 2022, *51*,101449. 10.1016/j.aei.2021.101449.
36. Benchmarking Large Language Models: Opportunities and Challenges. *Artif. Intell. Rev.* 2024, *57*, 72. https://link.springer.com/chapter/10.1007/978-3-031-68031-1_6.
37. Yang, F.; Meng, B. Design of Computer-Aided Instruction Model Based on Knowledge GraphConstruction and Learning Path Recommendation. *Int. J. Web-Based Learn. Teach. Technol.*10.4018/IJWLTT.366803.
38. Dudáš, A.; Kleinedler, A. Effective Visualization of Data Structures in Graph Databases. *J. Image Graph.* 2024, *12*, 283–291.10.18178/joig.12.3.283-291.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.