ACCSCIENCE PUBLISHING

Asian Journal of Water, Environment and Pollution. Vol. 23, No. 1 (2026), pp. 166-179.
doi: 10.36922/AJWEP025330253

ORIGINAL RESEARCH ARTICLE

Dam safety emergency response decision-making method based on a graph database and language model

Shilin Gong[1,2]*, Futing Sun[2], and Keng Chen[2]

[1]Power China Huadong Engineering Corporation Limited, Hangzhou, Zhejiang, China
[2]Large Dam Safety Supervision Center, National Energy Administration, Hangzhou, Zhejiang, China
***Corresponding author:** Shilin Gon (gong_sl@hdec.com)

Received: August 13, 2025; 1st revised: September 15, 2025; 2nd revised: September 27, 2025; 3rd revised: October 10, 2025; Accepted: October 13, 2025; Published online: November 3, 2025

Abstract: Risks and hidden dangers in dam safety management may occur suddenly, such as piping effects and dam overflow. Faced with these problems, several frontline dam managers are often at risk due to a lack of experience. To address this, a dam safety emergency response decision-making method based on a graph database and language model was proposed. First, a dam safety emergency knowledge system was constructed by identifying the potential components involved in the decision-making process. Relevant data were collected, organized, and stored according to this system, forming a knowledge graph of dam safety emergencies. Then, a Siamese Bidirectional Encoder Representations from Transformers Network was used to build a semantic matching model that effectively links dam safety emergency retrieval statements with corresponding cases in the graph database. A matching and sorting method was also developed to enable precise retrieval and intelligent recommendation of the most similar cases. The practical application of this method shows that it can effectively leverage professional expertise and typical cases in the dam safety domain, automatically providing decision support to dam operation safety management personnel through the integration of subgraphs and texts.

Keywords: Graph database; Language model; Dam; Emergency; Decision-making

1. Introduction

Dam safety incidents involve significant risks arising from several factors, such as over-standard floods, engineering failures, seismic and geological disasters, large-volume floating object impacts, and terrorist activities, all of which can lead to severe consequences if not addressed promptly and effectively.[1-4] A notable example occurred between February 7 and 14, 2017, when the main spillway of the Oroville Dam sustained extensive damage during discharge operations due to melting snow and continuous rainfall. Although discharge was temporarily halted to assess the damage, and the spillway was subsequently reactivated, erosion continued to worsen despite ongoing mitigation efforts.[5-7] During the accident, there were various opportunities for human intervention to prevent the situation from deteriorating further. However, due to the mutual influence and joint action of multiple inducing factors, the best time to curb the accident was missed.[8] Another incident took place on January 12, 2022, at the No. 3 Unit of the Guanzhou Hydropower Station, where a choke plug burst due to elevated water pressure, causing upstream water to rapidly inundate the powerhouse through the turbine's inlet valve passage, ultimately submerging the lower levels of the facility. This tragic event resulted in nine fatalities and an estimated direct economic loss of approximately RMB 44 million.[9] One of the causes of the accident was that the dam safety management personnel failed to detect the abnormalities in the hydropower station's operation in time and take corresponding measures. In recent years, countries have successively formulated laws and standards to address these issues. For example, the "Emergency Management Measures for Dam Operation Safety of Hydropower Stations" issued by the National Energy Administration of China stipulates that power companies should strengthen the information construction of dam emergency management. After an emergency occurs, they should immediately initiate an emergency response, take advanced measures, and control the situation. At present, most dam safety management departments in China have developed dam safety emergency response manuals.[10,11] However, these are text documents with complex content and cannot automatically recommend specific response measures based on the occurrence of emergency events.

Dam operation safety management relies on a combination of empirical experience and theoretical knowledge.[12,13] Over several decades, China's hydropower industry has amassed extensive professional expertise and numerous typical case studies; however, this valuable knowledge is predominantly stored in unstructured formats across various media, making systematic and effective utilization challenging. In dam safety management practice, experts with rich engineering experience and specialized knowledge are often capable of accurately identifying potential hazards based on abnormal dam behavior and recommending appropriate response measures. In contrast, some frontline dam safety personnel, lacking such experience, may struggle to respond effectively to emergencies, resulting in potential serious consequences. A notable example is the Edenville Dam failure, which was partly attributed to management's inability to fully open all six spillway gates promptly.[14] The investigation team believed that the opening of the gate limited the discharge of the spillway. If the gate could be fully opened, the maximum reservoir water level would drop about 0.3 m compared to the time of the accident. For the managers of the Edenville Dam, if they had received the correct decision support in time, such as fully opening the six gates, the accident might have been avoided.[15] Therefore, it is imperative to systematically organize and leverage industry knowledge by constructing a dam safety emergency knowledge graph, which can provide decision-making support during emergencies and help ensure the safety of dam operations.

Wu *et al.*[16] were the first to propose an expert system that integrates dam safety data analysis, comprehensive evaluation, and decision support, based on the "One Engine and Four Bases" framework, comprising an integrated inference engine, including database, knowledge base, method base, and graph database.[17] While this system holds considerable theoretical value, its knowledge representation and inference processes were relatively mechanical due to the limitations of knowledge base systems and natural language processing technologies available at the time. Sheng *et al.*[18] further explored the impact mechanisms of dynamic factors on decision-making by investigating how these factors influence the decision-making process and interact through feedback loops.[10] They developed a numerical model to capture the dynamics of decision systems, enabling a quantitative description of decision behaviors and their effects on remediation outcomes. In recent years, the rapid advancement of technologies, such as knowledge graphs and large language models, has created new possibilities for enhancing decision-making in dam safety emergency response. A knowledge graph is a novel approach that uses graph-based models to represent knowledge and capture relationships among various entities,[19-21] enabling the integration of professional expertise, abnormal events, remediation strategies, and other relevant information in dam safety management. Meanwhile, large language models can intelligently interpret user inputs related to dam emergencies, accurately connect them to corresponding decision-making options within the knowledge base, and, leveraging the knowledge graph, provide informed decision support by referencing the most relevant past cases.[22-25] Yan *et al.*[26] proposed a new water quality prediction modeling approach based on a knowledge graph and deep adversarial networks. A deep adversarial joint model was used to extract water quality data from different knowledge sources and construct a knowledge graph. The contribution of parameters in water quality prediction was calculated by simultaneously considering the correlation between knowledge and data, significantly improving the effectiveness of water quality prediction.[26] Wang and Xu[27] designed a watershed-internal knowledge graph and a large language model that encompasses watershed relationships and internal information structures. The research indicated that the long short-term memory (LSTM) model, when integrated with a watershed-internal knowledge graph and a large language model, can effectively incorporate critical elements influencing water level changes. The accuracy of mountain flood forecasting was enhanced by 3% compared to the standard LSTM model.[27] Yang *et al.*[28] proposed a novel intelligent question-answering approach in hydroengineering inspection based on a synergistic knowledge graph and a large language model. The knowledge contained in the graph was integrated into an optimal clarification path and transferred to a large language model to address noise and randomness, thereby enhancing the efficiency of the clarification process. The results confirmed the method's effectiveness in enhancing the accuracy of intelligent question-answering in hydroengineering inspection.[28]

Building on these foundations, this study proposes a dam safety emergency response decision-making method that integrates the knowledge storage and visualization strengths of knowledge graphs with the text understanding capabilities of large language models. A structured knowledge system for dam safety emergencies was developed based on key decision-making elements, which guided the systematic collection and organization of relevant data into a comprehensive knowledge graph. To enable intelligent case retrieval, a semantic matching model based on Siamese bidirectional encoder representations from transformers (SBERT) was constructed, allowing for the effective association of user queries with relevant cases in the graph. This was further supported by a case-matching and -ranking mechanism designed to ensure accurate retrieval and intelligent recommendation of decision-support information.

2. Methods

2.1. Knowledge graph

A knowledge graph is a type of graph database that organizes and stores various forms of knowledge in a structured network graph format.[29,30] As an efficient means of knowledge representation, it models relationships between entities by connecting nodes to form either directed or undirected graphs.[31] In a knowledge graph, nodes may represent tangible entities, such as dams or rivers, and abstract concepts, including emergency issues or decision-making solutions. The edges between nodes denote attributes, such as river names or dam elevations, or relationships, such as "located in" or "belongs to," thereby linking entities and constructing a comprehensive knowledge network. At its core, a knowledge graph is composed of triplets, each consisting of a head entity, a relationship, and a tail entity. These triplets represent discrete knowledge units from the real world. When aggregated in large numbers, they enable the modeling of complex inter-entity relationships, thus realizing a true interconnection of all relevant entities.

2.2. Semantic matching model

Semantic matching is a subfield of natural language processing that focuses on determining the degree of similarity between two textual expressions. Early methods for semantic matching primarily relied on basic word-level comparisons, where sentences were segmented into individual characters or words and then matched against entries in a knowledge base. The emergence of large language models, most notably bidirectional encoder representations from transformers (BERT), has introduced advanced solutions for semantic understanding and matching.[32-34] However, when using a single BERT model to calculate sentence similarity, the typical approach involves concatenating two statements as a single input, which can lead to increased computational complexity, especially with longer texts. Furthermore, BERT's multi-head attention mechanism merges the contextual information of both statements, potentially causing information loss. To overcome these limitations, this study integrated the BERT model with a Siamese network, constructing a semantic similarity matching model known as the SBERT network. This approach employed two separate BERT models to independently extract semantic features from each statement, calculate their similarity, and then intelligently match them within the knowledge graph based on the degree of semantic similarity.[35,36] As illustrated in Figure 1, the architecture of the SBERT network consisted of three main components: the input layer, the feature extraction layer, and the output layer.

To accommodate the specific data format required by the BERT model for text-based tasks, the input layer must first convert the two statements into a format suitable for model processing. Initially, the input statements were tokenized into word-level units using a tokenizer, resulting in two distinct sets of word tokens $A = (a_1, a_2, \cdots, a_m)$ and $B = (b_1, b_2, \cdots, b_n)$. Special tokens, classification token (CLS) and separator token (SEP), were then added to structure the input properly, where (CLS) was placed at the beginning of each token sequence and (SEP) was used to denote the end. This formatting ensured that the input was compatible with the BERT architecture and enabled effective extraction of semantic features. The final input structure for each statement followed the format in Equations (1) and (2).

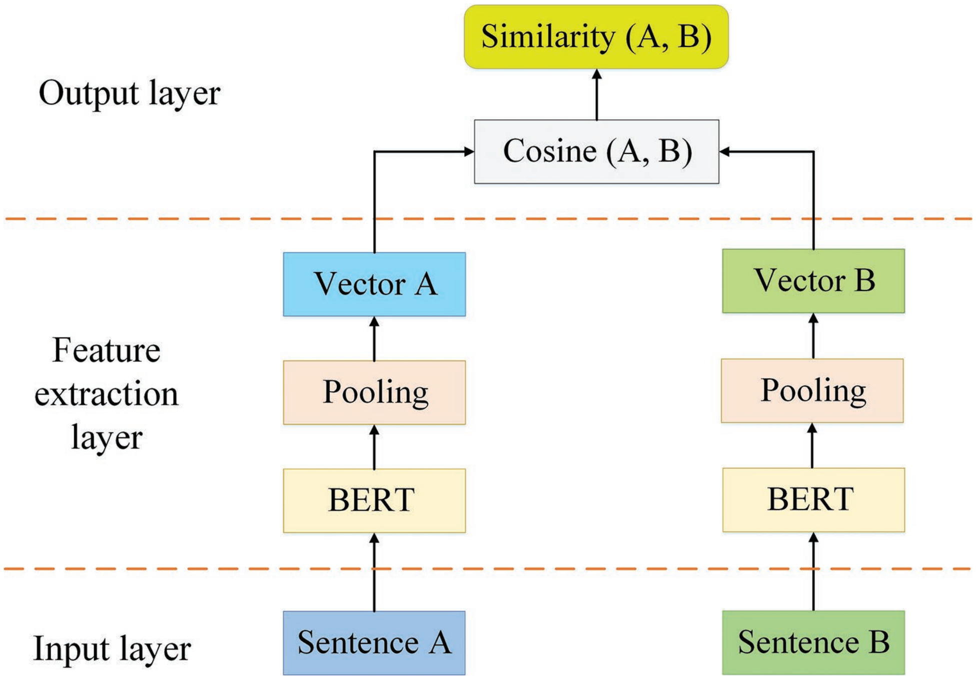

Figure 1. Structure of the Siamese bidirectional encoder representations from transformers (BERT) network

$$A = \{[CLS]\ a_1, a_2, \cdots, a_m [SEP]\} \tag{1}$$

$$B = \{[CLS]\ b_1, b_2, \cdots, b_n [SEP]\} \tag{2}$$

The feature extraction layer employed two separate BERT models to learn the textual features of the two input statements, using a Siamese network architecture that enabled parameter sharing between the models. Once the textual features were extracted, a mean pooling strategy was applied to reduce dimensionality while preserving the most significant semantic features. Subsequently, a bidirectional LSTM model equipped with an attention mechanism assigned varying weights to different components of each statement, enhancing the model's ability to capture interaction information both within individual statements and between them. After passing through the feature extraction layer, the two input sentences were converted into two feature vectors.

The output layer evaluated the feature vector representations of the two input statements to determine their semantic similarity. Among various methods for measuring vector distances, cosine similarity is the most widely used. It calculates the cosine of the angle between the two statement vectors, thereby quantifying how similar the texts are in the vector space. A cosine similarity value closer to 1 indicates that the vectors are nearly aligned, meaning the semantic content of the statements is highly similar, with the angle between them approaching 0°. In contrast, a cosine value closer to 0 implies that the vectors are nearly orthogonal, suggesting low semantic similarity with an angle approaching 90°. Assuming that the feature vectors of the two input statements were X and Y, with a vector feature dimension of n, that is, $X = (x_1, x_2, \cdots, x_n)$, $Y = (y_1, y_2, \cdots, y_n)$, the semantic similarity coefficient between the two statements could then be calculated as shown in Equation (3).

$$s = \frac{\sum_{i=1}^{n}(x_i \times y_i)}{\sqrt{\sum_{i=1}^{n} x_i^2} \times \sqrt{\sum_{i=1}^{n} y_i^2}} \tag{3}$$

2.3. Dam safety emergency knowledge system

Based on the key components involved in the dam safety emergency response decision-making process, dam safety emergency knowledge can be categorized into eight entity types: Dam, river, dam type, emergency issues, abnormal manifestations, causal factors, emergency measures, and treatment effectiveness. Among these, the dam serves as the central element of the knowledge system, acting as the carrier for emergency issues and related components. Its entity attributes describe engineering characteristics, such as dam name, height, storage capacity, installed capacity, and others. The dam depends on a river for its development, and rivers at different levels form a natural hydrological network. River attributes include the river name and the basin to which it belongs. The dam type represents a fundamental classification, as different dam types may experience distinct emergency issues and require varied response strategies. Emergency issues encompass representative incidents that have occurred in dams, such as wide-span tensile stress in concrete arch dams or sudden increase in leakage at the base of earth-rockfill dams, and include details, including issue name and description. Abnormal manifestations refer to physical or monitoring-based indicators associated with emergency issues, such as increased leakage pressure at arch dam abutments or a strong correlation between leakage and reservoir water level. Causal factors represent the root causes of emergencies, examples being a collision with a sand carrier or a long-term operation of gates with minimal opening. Emergency measures include detailed solutions for addressing these issues. Treatment effectiveness describes the outcomes of interventions, evaluated across four levels: significant, good, moderate, and poor. Entity relationships define the semantic links between pairs of entities, serving as vital components of the structured knowledge. Based on the eight proposed entity types, a total of 14 types of entity relationships are established to support comprehensive knowledge modeling. The detailed structure of the dam safety emergency knowledge system is illustrated in Figure 2.

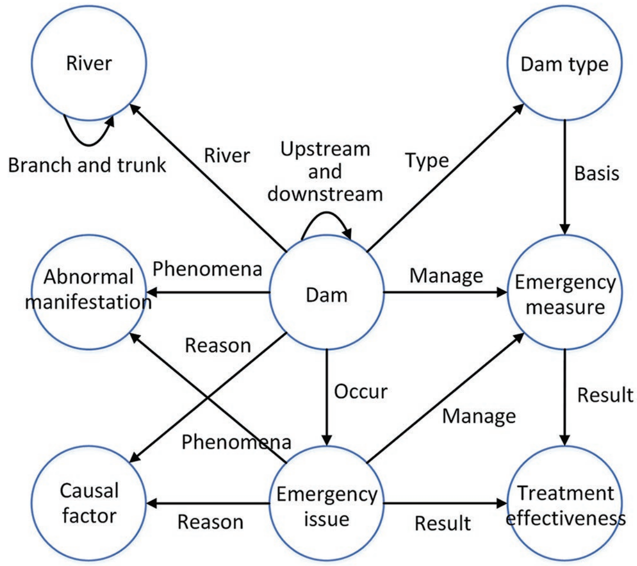

Figure 2. Dam safety emergency knowledge system

3. Experiment

3.1. Construction of the dam safety emergency knowledge graph

The construction of a dam safety emergency knowledge graph served as a foundational step in enabling effective decision support. Based on the dam safety emergency knowledge system outlined in Section 2.3, relevant knowledge was extracted from diverse sources, such as documents, web pages, and books, and then organized into a standardized, unified format. Once standardized, this structured knowledge was stored in bulk to form a comprehensive dam safety emergency knowledge graph, enabling systematic representation and retrieval of critical information for emergency decision-making.

3.1.1. Collection and compilation of dam safety emergency knowledge

The data sources used to construct the dam safety emergency knowledge graph in this study were primarily composed of three parts. The first part included dam engineering characteristic data stored in various dam safety management systems constructed by national supervisory agencies and power companies, covering several attributes, such as dam height, storage capacity, and installed capacity. These datasets were stored in a structured format and were primarily collected using an automated batch export method. The second part consisted of semi-structured data presented in tables or lists, containing information on domestic rivers and their tributaries, as well as upstream and downstream relationships of dams. These data were collected from China's River and Lake Dictionary, which is China's first professional reference book that systematically sorts out the current status of river and lake systems. After proper organization, these data could also be processed in batches. The third part involved unstructured textual data derived from professional reports, technical books, academic papers, and specialized websites, encompassing various details, such as emergency issues, abnormal manifestations, causal factors, emergency measures, and effectiveness of intervention. Since this information was primarily unstructured, research and testing were conducted using automatic text knowledge extraction models, such as TechGPT-2.0, OneKE, and Unified Structure Generation for Universal Information.

While these models demonstrated high flexibility and accuracy in extracting short text, they were less effective in handling long texts that required contextual judgment. Consequently, long textual data, such as detailed emergency case descriptions, were manually extracted and organized. The dam safety emergency cases were the key to making emergency response decisions. A total of 1,177 dam safety emergency cases were collected in this study, encompassing almost all possible event categories, including over-standard floods, engineering failures, seismic and geological disasters, large-volume floating object impacts, and flooded powerhouses. Each emergency case contained information, such as emergency issues, emergency process, abnormal manifestations, causal factors, emergency measures, and remediation effectiveness, which could be used for emergency decision support. Part of the data sources for constructing the dam safety emergency knowledge graph is shown in Table A1. Ultimately, all collected data were transformed into a unified triplet format to illustrate the dam safety emergency knowledge graph. Examples of such triplets include “Goupitan, dam height, 230.5 m” and “Goupitan, emergency issue, erosion damage to the plunge pool.”

3.1.2. Storage of dam safety emergency knowledge based on Neo4j database

After the knowledge collection phase, the target data used for constructing the dam safety emergency knowledge graph was organized in the form of triplets, making it well-suited for storage in a graph database. At present, the primary data models employed for knowledge graphs include the resource description framework (RDF) graph and the property graph.

The RDF Graph is a standardized data model primarily utilized in the Semantic Web domain and developed by the World Wide Web Consortium standardization organization. It is designed to describe entities, resources, and the relationships between them, with a core structure based on triplets composed of a subject, predicate, and object, where the subject and object represent entities and the predicate defines the semantic relationship between them. While the RDF graph is recognized for its formal rigor and comprehensive functionality, it has notable limitations in practical implementation. Its usage is mainly confined to academic research, with limited adoption in industrial applications. In contrast, the property graph model, introduced more recently, offers greater flexibility and better compatibility with existing storage systems. This model emphasizes the relationships between nodes and edges and is composed of elements such as nodes, edges, properties, identifiers, and labels. Neo4j, the most widely adopted open-source property graph database, effectively illustrates the interconnected characteristics of data and is valued for its fast read/write performance, highly flexible structural design, and support for distributed, high-availability deployment.

In this study, the Neo4j database was employed to construct the dam safety emergency knowledge graph, with the py2neo toolkit serving as an interface between Python and the Neo4j environment. Python programming was utilized to convert a large volume of triplets into the property graph format. These triplets were then imported into the Neo4j database in batches to complete the graph construction process, as illustrated in Figure 3. The subgraph is shown in Figure 4.

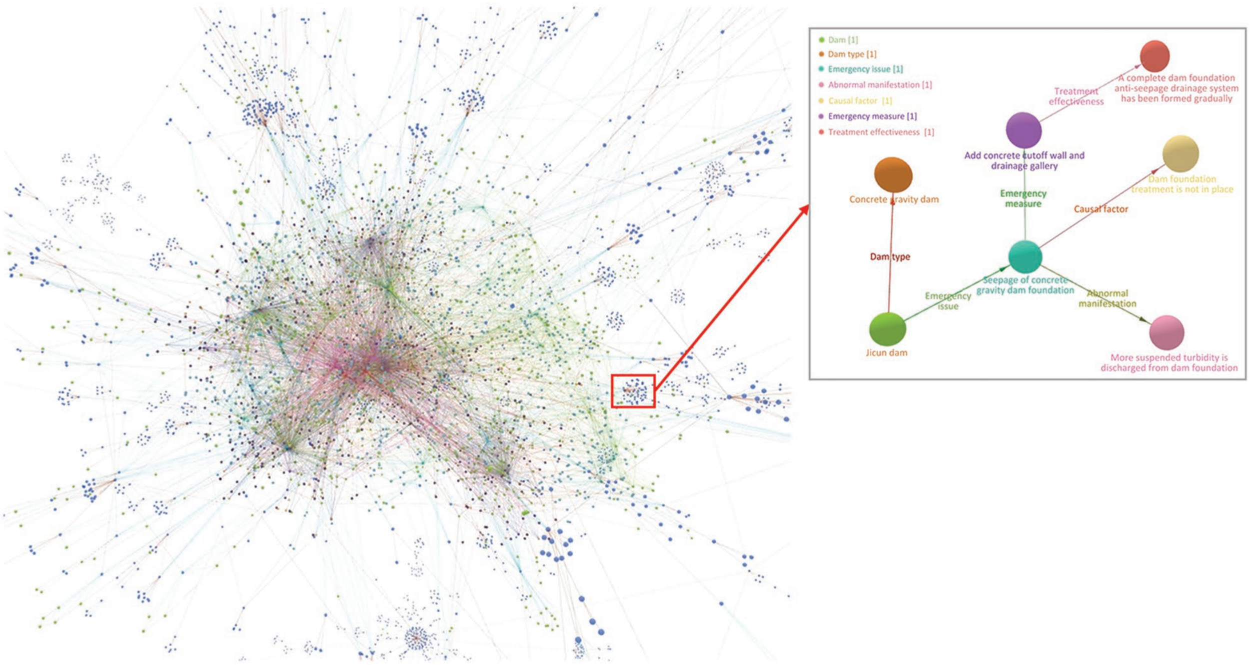

Figure 3. Dam safety emergency knowledge graph

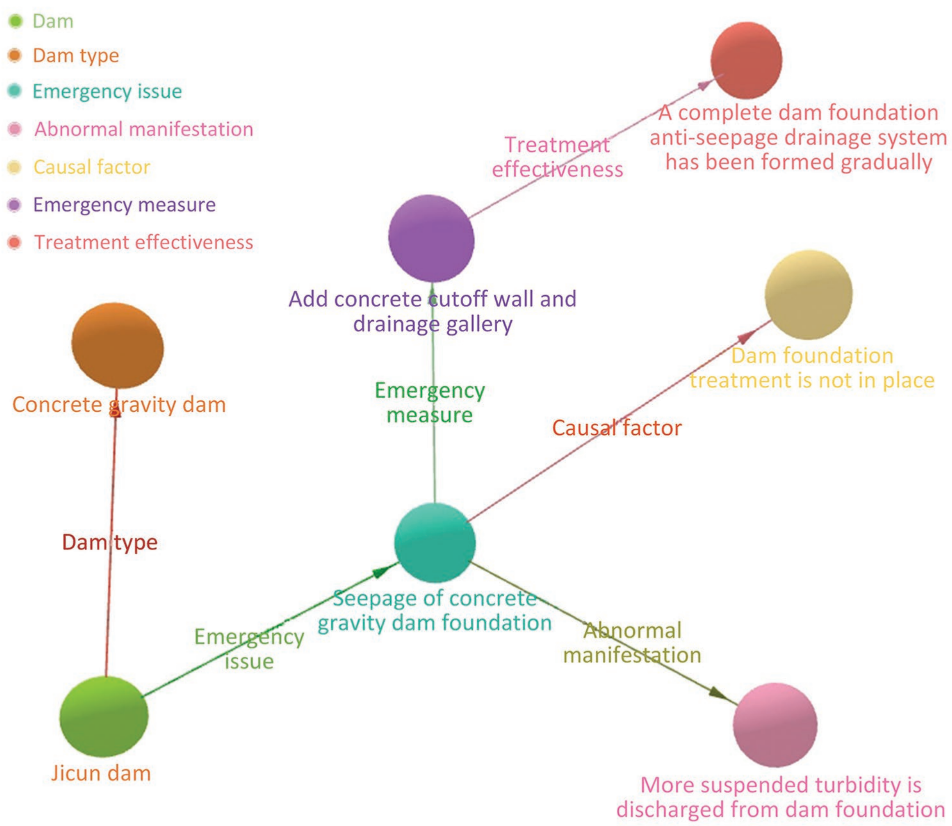

Figure 4. Subgraph of the dam safety emergency case

3.2. Construction of the dam safety emergency semantic matching model

When frontline dam safety management personnel require decision support during emergencies, they often seek to retrieve similar past cases along with their corresponding response measures. This necessitates the development of an association matching model that links input statements to the trigger statements of dam safety emergency scenarios within the knowledge graph. To achieve semantic matching, the SBERT model constructed in Section 2.2 was utilized, where the effectiveness of the matching process largely depended on the choice of the underlying BERT model. Since the introduction of the BERT architecture, numerous organizations have developed a wide range of pre-trained BERT models and BERT-based variants tailored to various corpora, resulting in differences in model applicability. Among these, the bge-large-zh-v1.5 model demonstrated superior performance as a leading semantic vector model compared to other large language models, having outperformed comparable models in terms of Chinese semantic representation capabilities. Consequently, this study adopted the bge-large-zh-v1.5 model as the pre-trained foundation for the SBERT network, enabling accurate semantic parsing and vector transformation of input statements.

3.3. Intelligent matching of dam safety emergency cases based on a knowledge graph

Dam safety emergency cases were stored within the knowledge graph as subgraphs, with each emergency case represented as an individual subgraph. A complete subgraph of a dam safety emergency case is illustrated in Figure 4. Each subgraph included seven key entities: Dam, dam type, emergency issues, abnormal manifestations, causal factors, emergency measures, and remediation effectiveness. The entity attributes within each subgraph contained detailed information specific to the corresponding emergency case. For example, the entity attributes of the emergency issue included both the name and a comprehensive description of the issue. The issue name played a critical role in case retrieval, as it served as the basis for semantic matching with the input query related to an emergency issue. Consequently, intelligent matching of dam emergency issue cases entailed retrieving relevant subgraphs from the knowledge graph and extracting the associated entity attribute content for decision support.

As shown in Figure 5, matching dam safety emergency cases involved the following four steps:

(i) Collection of dam safety emergency cases: All emergency issue entities were retrieved from the dam safety emergency knowledge graph using Cypher query language. The names of these entities were then extracted and compiled into a collection of emergency issue names, which served as the key reference for semantic matching. This collection of emergency issue names is formally represented in Equation (4).

$$C = [c_1, c_2, c_3, \cdots, c_n] \quad \text{(IV)}$$

(ii) Matching of candidate emergency issues: Based on the dam safety emergency semantic matching model developed in Section 3.2, the semantic similarity between the input dam emergency issue retrieval statement and the collection of emergency issue names was calculated. Similarity scores were computed between the input statement and each element within the collection. The top 5 elements with the highest similarity coefficients were then selected as candidate emergency issue entities for further processing or retrieval.

(iii) Subgraph of dam safety emergency case: For each candidate emergency issue entity obtained from step 2, the associated entities in the dam safety emergency knowledge graph, such as dam, dam type, abnormal manifestations, causal factors, emergency measures, and remediation effectiveness, were retrieved using single-hop and multi-hop query methods. These related entities were then used to generate the corresponding subgraph representing the complete dam safety emergency case.

(iv) Text representation of dam safety emergency case: The textual content of the dam safety emergency case was retrieved by querying the entity attributes of the subgraph entities within the graph database. These retrieved attributes were then assembled to form the complete textual representation of the dam safety emergency case.

4. Engineering case

4.1. Engineering introduction and emergency issue

A domestic pumped-storage power station consists of an upper reservoir, a lower reservoir, a waterway system,

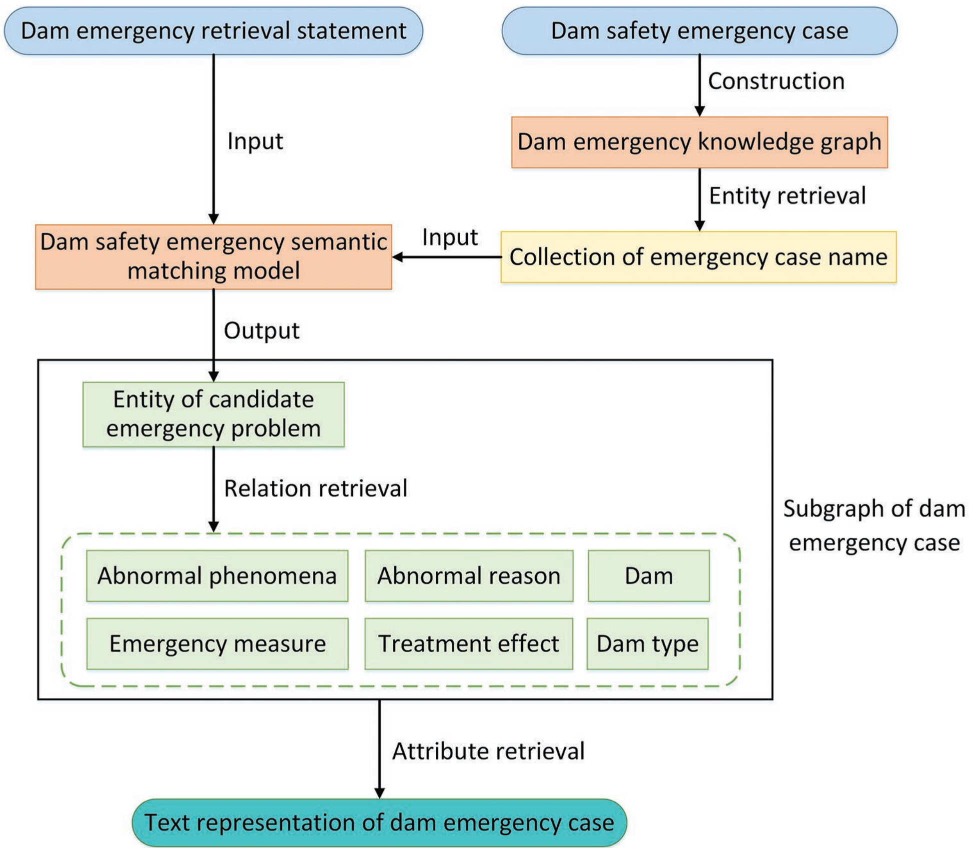

Figure 5. Dam safety emergency case matching process

and an underground powerhouse system. The upper reservoir is an asphalt concrete-faced rockfill dam with a maximum height of 43.9 m. The lower reservoir is a roller-compacted concrete gravity dam with a maximum height of 73.0 m, equipped with spillways and other facilities. During a routine dam safety inspection, the following issues were discovered: High leakage in the concrete-faced rockfill dam body, erosion damage to the stilling-basin slab, and tensile stress existed in the concrete gravity dam heel.

4.2. Semantic matching of dam safety emergency cases

According to the dam safety emergency case matching method in Section 3.3 and the semantic matching model in Section 3.2, emergency cases similar to the current emergency issues of the pumped storage power station were matched from the dam safety emergency knowledge graph and sorted according to the similarity coefficient, as shown in Table 1. As demonstrated by statement pairs 1, 6, and 11, corresponding to the input statements of high leakage in the concrete-faced rockfill dam body, erosion damage to the stilling-basin slab, and tensile stress exists in the concrete gravity dam heel, respectively, the model successfully identified the most semantically similar emergency issue cases based on the calculated similarity coefficients, all of which exceeded 0.95 for correct matches. Furthermore, a comparison between statement pairs 1 and 2 revealed that the similarity coefficient of pair 2 was significantly lower than that of pair 1, indicating that the semantic matching algorithm developed in this study demonstrated strong discriminative ability between synonyms and antonyms. For instance, it accurately distinguished between high leakage, severe leakage, and low leakage. In addition, the similarity coefficient of statement pair 6 was higher than that of pair 8, suggesting that the model recognized greater semantic similarity between erosion and scouring than between erosion and cavitation. This observation

Table 1. Semantic matching test results for dam safety emergency issues

S/N	Current emergency issues	Candidate emergency cases	Similarity coefficient
1	High leakage in the concrete-faced rockfill dam body	Severe leakage in the concrete-faced rockfill dam body	0.968
2	High leakage in the concrete-faced rockfill dam body	Low leakage in the concrete-faced rockfill dam body	0.891
3	High leakage in the concrete-faced rockfill dam body	High leakage in the concrete gravity dam	0.787
4	High leakage in the concrete-faced rockfill dam body	Obvious deformation in the concrete-faced rockfill dam body	0.883
5	High leakage in the concrete-faced rockfill dam body	Collapse of the earth-rockfill dam body	0.732
6	Erosion damage to the stilling-basin slab	Scouring damage to the stilling-basin slab	0.966
7	Erosion damage to the stilling-basin slab	Erosion damage to the plunge pool slab	0.748
8	Erosion damage to the stilling-basin slab	Cavitation damage to the stilling-basin slab	0.925
9	Erosion damage to the stilling-basin slab	Exposed reinforcement on the sidewall of the stilling basin	0.631
10	Erosion damage to the stilling-basin slab	Erosion damage to the spillway chute	0.633
11	Tensile stress exists in the concrete gravity dam heel	Tensile stress appears in the concrete gravity dam heel	0.986
12	Tensile stress exists in the concrete gravity dam heel	Tensile stress exists in the concrete arch dam heel	0.925
13	Tensile stress exists in the concrete gravity dam heel	Excessive tensile stress in the concrete gravity dam body	0.886
14	Tensile stress exists in the concrete gravity dam heel	Compressive stress exists in the concrete gravity dam heel	0.942
15	Tensile stress exists in the concrete gravity dam heel	A scouring pit exists at the concrete gravity dam heel	0.852

Abbreviation: S/N: Serial number.

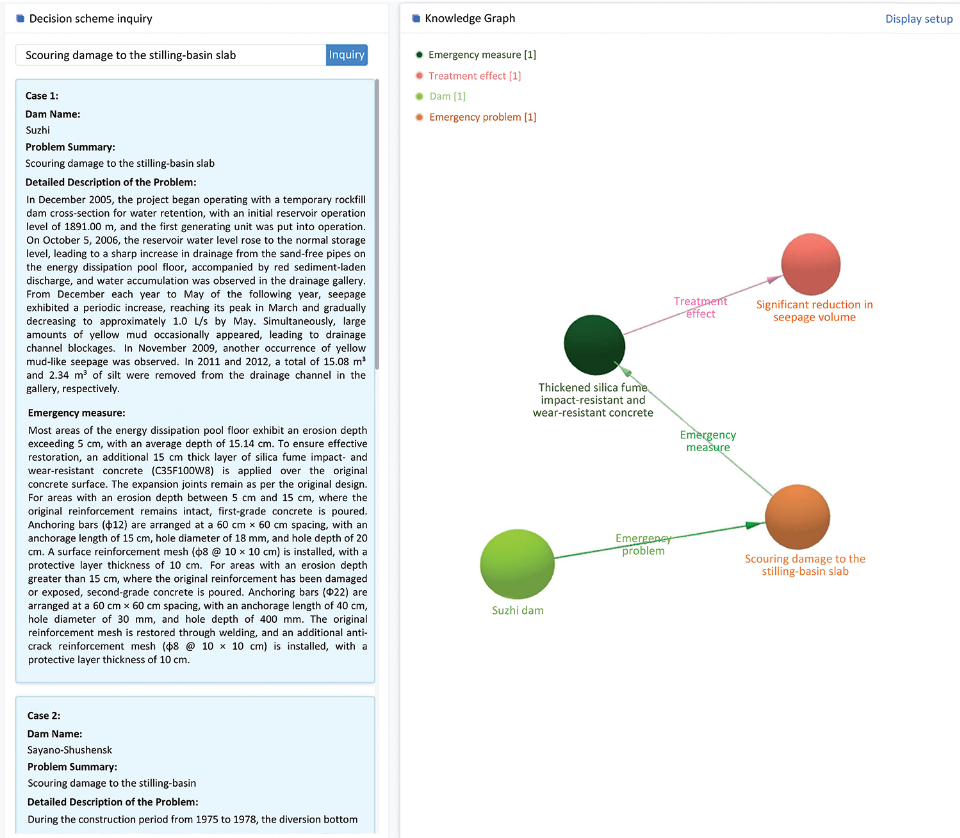

Figure 6. Results of auxiliary decision support for dam operation safety

aligned with domain knowledge in the dam safety management industry, demonstrating the model's ability to understand specialized terminology. In summary, the dam emergency knowledge semantic matching model, based on the SBERT network developed in this study, can effectively identify semantic similarities in dam safety emergency statements and reliably retrieve the most relevant emergency cases from the knowledge graph.

4.3. Subgraph and text representation of the dam safety emergency case

Based on the method proposed in this study, a dam safety emergency decision-making system was constructed to display recommended similar cases and decision-making solutions. The system interface is shown in Figure 6. For each candidate emergency issue entity in Table 1, the associated entities in the dam safety emergency knowledge graph, such as dam, dam type, abnormal manifestations, causal factors, emergency measures, and remediation effectiveness, were retrieved using single-hop and multi-hop query methods. These related entities were used to generate the corresponding subgraph representing the complete dam safety emergency case, as shown on the right side of Figure 6. Then, the textual content of the dam safety emergency case was retrieved by querying the entity attributes of the subgraph entities within the graph database. These retrieved attributes were assembled to form the complete textual representation of the dam safety emergency case, as shown on the left side of Figure 6. These retrieved cases included similar case descriptions and detailed emergency response measures. The dam safety manager could make decisions based on the disposal measures recommended by the dam safety emergency decision-making system.

5. Conclusion

This study proposes a dam safety emergency response decision-making system that integrates a knowledge graph, large language models, and a substantial collection of historical cases. Compared to existing methods, the proposed system enables users to query decision-making solutions through natural language inputs and delivers results in a combined format of subgraphs and textual descriptions, thereby offering effective support to technical personnel involved in dam safety management. The specific conclusions are as follows:

(i) The knowledge graph is well-suited for constructing dam safety emergency knowledge bases, as representing and storing emergency cases in graphical form facilitates effective retrieval, clear presentation, and efficient utilization of industry-specific knowledge.

(ii) The SBERT network, integrated with the bge-large-zh-v1.5 vector model, demonstrates strong performance in the semantic matching of dam safety emergency knowledge and accurately interprets specialized vocabulary within the dam safety domain.

(iii) By integrating knowledge graph technology with the SBERT network, this study achieves the systematic utilization of professional expertise and typical cases. The proposed method not only supports emergency decision-making during critical events but also offers valuable decision support for potential hazard management in routine dam operations.

(iv) The method proposed in this study is highly dependent on dam safety emergency cases stored in the knowledge graph. Continuous improvement of the categories and number of cases is necessary to obtain better decision-making support effects in future work. In addition, the decision-making system serves solely as an auxiliary tool, offering case-based support and a reference framework for dam safety managers; the final decision-making plan should be determined through further expert consultation and discussion.

Acknowledgments

None.

Funding

This research was funded by the China Postdoctoral Science Foundation (2023M733315) and the National Key Research and Development Program of China (2021YFC3090101).

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Shilin Gong
Data curation: Shilin Gong
Funding acquisition: Shilin Gong
Formal analysis: Keng Chen
Investigation: Futing Sun
Methodology: Shilin Gong
Project administration: Keng Chen
Resources: Shilin Gong
Software: Futing Sun
Supervision: Keng Chen
Validation: Keng Chen
Visualization: Futing Sun
Writing–original draft: Shilin Gong
Writing–review & editing: Futing Sun

Availability of data

Data are available from the corresponding author on reasonable request.

References

1. Cai YJ, Yang XG, Zhang LM, Zhou JW, Peng WX, Fan G. Research framework and anticipated results of the rapid detection of risk assessment, research and development of emergency disposal technology and equipment of dammed lakes. *Adv Eng Sci*. 2020;52(2):10-18. doi: 10.15961/j.jsuese.201900918
2. Tran CK. Sensitivity analysis of dam breach parameters for variation capacity earthen dams. *Asian J Water Environ.* 2021;18(3):1-9. doi: 10.3233/AJW210023
3. Wang XH, Xu H, Zhou KF. Analysis of key techniques for dam break emergency management. *Water Resour Power*. 2010;28(5):54-55, 173.
4. Johnpaul V, Venkatesan G, Vinoth V. Feasibility assessment of medium-scale hydro-power plant - Case study in Kerala, India. *Asian J Water Environ*. 2024;21(2):39-47. doi: 10.3233/AJW240020
5. Vahedifard F, AghaKouchak A, Ragno E, Shahrokhabadi S, Mallakpour I. Lessons from the Oroville dam. *Science*. 2017;355(6330):1139-1140. doi: 10.1126/science.aan0171
6. Zhou XB, Zhou JP, Du XH. Lessons from the Oroville Dam spillway failure analysis. *J Hydraul Eng ASCE*. 2019;50(5):650-660. doi: 10.13243/j.cnki.slxb.20181009
7. Henn B, Musselman KN, Lestak L, Ralph FM, Molotch NP. Extreme runoff generation from atmospheric river driven snowmelt during the 2017 Oroville dam spillways incident. *Geophys Res Lett*. 2020;47(14):e2020GL088189. doi: 10.1029/2020GL088189
8. Koskinas A, Tegos A, Tsira P, *et al.* Insights into the

Oroville dam 2017 spillway incident. *Geosciences*. 2019;9(1):37.
doi: 10.3390/geosciences9010037
9. Wang XG, Cheng RY, Zhao D. Warning significance from the "1·12" water inrush accident of Guanzhou hydropower station. *Dam Saf*. 2022;(5):1-3.
10. Shen GZ, Lu Y, Zhang SC, *et al*. Risk dynamics modeling of reservoir dam break for safety control in the emergency response process. *Water Supply*. 2021;21(3):1356-1371.
doi: 10.2166/ws.2021.004
11. Li SM, Yuan LW, Yang H, An HM, Wang GJ. Tailings dam safety monitoring and early warning based on spatial evolution process of mud-sand flow. *Saf Sci*. 2020;124:104579.
doi: 10.1016/j.ssci.2019.104579
12. Ding SL, Pan JJ, Wang YL, Xu H, Li DQ, Liu X. Developing a digital twin for dam safety management. *Comput Geotech*. 2025;180:107120.
doi: 10.1016/j.compgeo.2025.107120
13. Martac R, Milivojevic N, Despotovic-Zrakic M, Bogdanovi Z, Barac D. Enhancing large dam safety using IoT technologies: A case of a smart dam. *J Univers Comput Sci*. 2020;26(5):583-603.
doi: 10.3897/jucs.2020.031
14. France JW, Alvi IA, Miller AC, *et al*. Final report-Investigation of failures of Edenville and Sanford Dams. Independent Forensic Team; 2022. Available from: https://damsafety.org/MI-Final-Report [Last accessed on 2025 Oct 01].
15. Babaki AP, Tannant DD. Numerical evaluation of static liquefaction-induced flowslide causing the Edenville dam failure, Michigan. *J Geotech Geoenviron*. 2024;150(5):04024034.
doi: 10.1061/JGGEFK.GTENG-11998
16. Wu ZR, Gu CS, Hu QG, Li J. An investigation on the dam safety comprehensive appraisal expert system. *Water Resour Power*. 2000;(2):1-5.
doi: 10.20040/j.cnki.1000-7709.2000.02.001
17. Gu CS, Wu ZR. On-time analysis expert system for Long Yangxia Dam. *J Manag Sci China*. 1997;???(1):72-79.
18. Sheng JB, Xiang Y, Yang DW, Yan JH, Dong K. Key technologies and applications of safety diagnosis and smart management of reservoir dams. *Chin J Geotech Eng*. 2022;44(7):1351-1366.
doi: 10.11779/CJGE202207013
19. Padia A, Kalpakis K, Ferraro F, Finin T. Knowledge graph fact prediction via knowledge-enriched tensor factorization. *J Web Semant*. 2019;59:100497.
doi: 10.1016/j.websem.2019.01.004
20. Tamasauskaite G, Groth P. Defining a knowledge graph development process through a systematic review. *ACM T Softw Eng Meth*. 2023;31(1):27.
doi: 10.1145/3522586.
21. Bi Z, Cheng SY, Chen J, Liang XZ, Xiong FY, Zhang NY. Relphormer: Relational graph transformer for knowledge graph representations. *Neurocomputing*. 2024;566:127044.
doi: 10.1016/j.neucom.2023.127044
22. Floridi L, Chiriatti M. GPT-3: Its nature, scope, limits, and consequences. *Mind Mach*. 2020;30(4):681-694.
doi: 10.1007/s11023-020-09548-1
23. Kumar R, Kaur A, Dangi H. Role of artificial intelligence in modernisation of fire risk management. *Asian J Water Environ*. 2024;21(6):213-219.
doi: 10.3233/AJW240089
24. Feng YY, Wang Q, Xie MH, Shu JW. From BERT to ChatGPT: Challenges and technical development of storage systems for large model training. *J Comput Res Dev*. 2024;61(4):809-823.
25. Senthilmahesh H, Adavala KM, Thangamariappan T, Kumar PS, Tumaati R, Nagappan MP. Enhancing swimming pool hygiene: A robotic approach to debris removal and water quality monitoring. *Asian J Water Environ*. 2025;22(2):19-31.
doi: 10.36922/ajwep.6564
26. Yan JZ, Gao QC, Yu YCA, Chen LH, Xu Z, Chen JH. Combining knowledge graph with deep adversarial network for water quality prediction. *Environ Sci Pollut R*. 2023;30(4):10360-10376.
doi: 10.1007/s11356-022-22769-4
27. Wang SS, Xu OG. Exploring a long short-term memory for mountain flood forecasting based on watershed-internal knowledge graph and large language model. *PLoS One*. 2025;20(3):e0318644.
doi: 10.1371/journal.pone.0318644
28. Yang YR, Chen SS, Zhu YP, Liu XM, Pan SF, Wang X. Intelligent question answering for water conservancy project inspection driven by knowledge graph and large language model collaboration. *LHB*. 2024;110(1):2397337.
doi: 10.1080/27678490.2024.2397337
29. Chai X. Diagnosis method of thyroid disease combining knowledge graph and deep learning. *IEEE Access*. 2020;(8):149787-149795.
doi: 10.1109/ACCESS.2020.3016676
30. Ji S, Pan S, Cambria E, Marttinen P, Yu PS. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transc Neural Netw Learn Syst*. 2022;33(2):494-514.
doi: 10.1109/TNNLS.2021.3070843
31. Li X, Liu H, Zhao X, Zhang GG, Xing CX. Automatic approach for constructing a knowledge graph of knee osteoarthritis in Chinese. *Health Inf Sci Syst*. 2020;8(1):1-8.
doi: 10.1007/s13755-020-0102-4
32. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. *In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language*

Technologies; 2019. p. 4171-4186.

33. Nicolae DC, Yadav RK, Tufiş D. A lite romanian BERT: ALR-BERT. *Computers*. 2022;11(4):57. doi: 10.3390/computers11040057
34. Reimers N, Gurevych I. Sentence-BERT: Sentence embeddings using siamese BERT Networks. *In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*; 2019. p. 3980-3990.
35. Seo J, Lee S, Liu L, Choi W. TA-SBERT: Token attention sentence-BERT for improving sentence representation. *IEEE Access*. 2022;(10):39119-39128. doi: 10.1109/ACCESS.2022.3164769
36. Cheng HQ, Liu SQ, Sun WW, Sun Q. A neural topic modeling study integrating SBERT and data augmentation. *Appl Sci Basel*. 2023;13(7):4595. doi: 10.3390/app13074595

Appendix

Table A1. Part of the data sources for constructing the dam safety emergency knowledge graph

Data source	Internet link	Category
Association of state dam safety officials	https://damsafety.org/	Website
Hydropower station dam safety operation safety monitoring platform	https://v3.dam.com.cn/	Website (not open to the public)
China's river and lake dictionary	/	Book
Dam hazard assessment and hazard removal and reinforcement technology	/	Book
Risk management and control of water conservancy and hydropower projects	/	Book
Identification and treatment technology of safety hazards for reservoirs and dams	/	Book
Analysis of typical failure cases of equipment and facilities in pumped storage power stations	/	Book
Analysis of abnormal working behavior of hydraulic structures	/	Book
Examples of reinforcement of dangerous reservoirs	/	Book
Technical guidelines for handling safety hazards of earth-rock dams in small reservoirs	/	Standard
Design of hazard removal and reinforcement of the main dam of Fuxi Reservoir	/	Article
Leakage analysis and evaluation of Huidong hydropower station	/	Article
Analysis on risk elimination and reinforcement measures for Bazai reservoir project	/	Article
Study on the risk elimination and reinforcement project for chaganghe reservoir	/	Article
Research on key technologies for Danan reservoir hazard removal and reinforcement	/	Article
Analysis on the scheme for eliminating danger and reinforcing the Lingxi reservoir dam	/	Article