Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 11, No. 3, 2025

Construction and Implementation of Coal Mine Safety Knowledge Graph Based on Historical Accidents

Minmin Li *, Yanyu Diao, Xin Wang

School of Big Data and Artificial Intelligence, Anhui Xinhua University, Hefei Anhui, 230088, China

*** Corresponding author:** Minmin Li (Email: 464130929@qq.com)

Abstract: With the rapid development of intelligent coal mine, the field of coal mine also ushered in a large number of heterogeneous data information. Aiming at the data information existing in the field of coal mine, a coal mine safety knowledge atlas centered on coal mine accident cases is constructed. The construction of knowledge graph mainly uses crawler technology to obtain text data of coal mine, Albert-BilSTM-CRF model to realize entity recognition, Albert-BilSTM-ATT to complete relation extraction, and obtain triplet form in the field of coal mine safety. Finally, the obtained triplet form is imported into Neo4J graph database. Form a structured knowledge graph.

Keywords: Coal Mine Accident; Entity Recognition; Relation Extraction.

1. Introduction

Coal mining industry, as an important economic pillar industry in China, has always been vigorously developed by the state, and while the coal mining industry is rapidly transforming, its safety problems have also been attached great importance. Although in recent years in the country's continuous transformation of the coal mining industry, coal mine safety accidents have been reduced year by year, but in order to better prevent the occurrence of coal mine accidents, this paper through the coal mine safety network on the coal mine accident case text data, to build the knowledge graph in the field of coal mine safety, through the knowledge graph will be associated between the accident, so as to analyze it, to find out the similarities between the accidents, and to avoid similar accidents from reoccurring.

In this paper, the accident cases in the coal mine safety production network are used as the data source for the construction of coal mine safety knowledge graph, and the text mainly analyzes the common seven kinds of coal mine accident cases, including: gas accidents, roof accidents, transportation accidents, electromechanical accidents, water penetration accidents, fire accidents, and coal dust explosion accidents. Data crawling is realized by using java language, combined with deep learning algorithm ALBERT-BiLSTM-CRF model, ALBERT-BiLSTM-ATT model realizes entity recognition as well as relationship extraction of coal mine text data, and the ternary data obtained by relationship extraction is saved in CSV format, then imported into Neo4j database, and through Neo4j graph database to visualize the coal mine safety knowledge graph.

Knowledge graph is a knowledge base in the form of a graph that represents the relationships between entities in the real world [1]. Through data mining technology and then analyze, organize, and discover the association between entities and entities, which constitutes a semantic network graph. With the gradual maturity of the development of knowledge graph technology, knowledge graph has been gradually applied to various fields. Peng Yunong[2] and others applied knowledge graph in the field of corn agriculture planting, using knowledge graph technology to build the completed corn production process, knowledge graph can enhance the efficiency of users to obtain knowledge of corn planting, and improve the management level of corn planting. Zeng Ziling[3] and others stored massive medical knowledge in the form of mapping, which reduces the loss of information, preserves the original meaning of knowledge, and contributes to the in-depth application of knowledge in the Chinese medicine industry. Pu Tianjiao[4] and others fully mining the data information in the power Internet of Things, to achieve the construction of knowledge map in the field of power, the concepts, entities and events that exist in the power system through the map of the relationship between the structured form of display, the proposed method makes the power industry chain has a more effective cross-media big data organization that is management.

As the coal mining industry data information is messy and heterogeneous, the content is not only extensive but also complex. Therefore, the application of knowledge graph to the coal mining industry is less and incomplete. Li Zhe[5] and others mainly focus on the class of coal mine mechanical and electrical equipment accidents, through the analysis of coal mine mechanical and electrical historical accident cases, analyze the historical data of coal mine mechanical and electrical equipment accident cases, and correlate the accidents obtained, the equipment involved in the accidents, the causes, and the treatment measures in the face of the accidents, to achieve the diagnosis of coal mine mechanical and electrical equipment accidents and risk management. Liu Peng[6] and others made a preliminary exploration of intelligent query for coal mine safety knowledge graph construction by introducing knowledge graph technology into the field of coal mine safety, categorizing and modeling relevant knowledge concepts, and proposing a language-based query method based on the constructed knowledge graph. Based on the above research, this paper proposes to construct a knowledge graph in the field of coal mine safety based on seven types of historical coal mine accident cases, to correlate the same accidents and analyze the implied information between the accidents, so as to better prevent the occurrence of coal mine accidents.

2. Coal Mine Safety Knowledge Graph Construction Process and Techniques

Knowledge graph construction generally adopts both top-down and bottom-up approaches [7]. Top-down construction of the knowledge graph needs to first define the ontology and data schema, and add new entities to the existing knowledge base, i.e., top-down construction can continuously add new entities to the existing knowledge base, and construct the knowledge graph for the existing knowledge base. Bottom-up construction first obtains data from various web pages, and after processing, screens out entities and relationships that meet the requirements and adds them to the knowledge base. In this paper, the bottom-up construction method is chosen for the construction of knowledge graph in the field of coal mine safety. The construction process is shown in Figure 1.

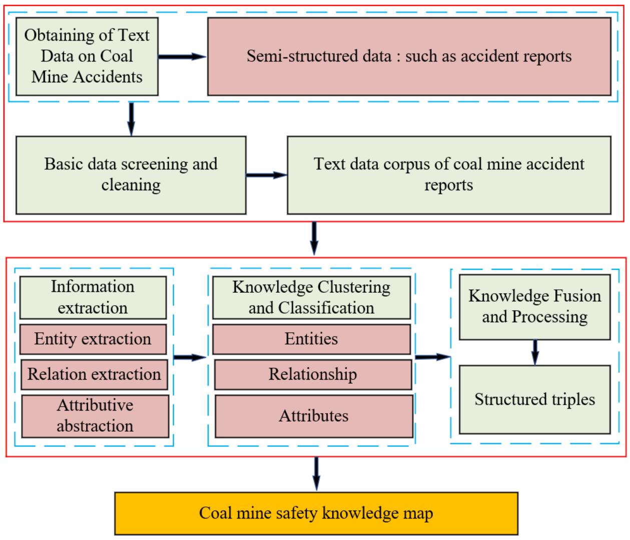

Figure 1. Construction process of knowledge map of coal mine safety

2.1. Coal Mine Accident Text Data Acquisition and Processing

Acquisition of coal mine accident text data is the first step of constructing knowledge graph, the acquisition of data in this paper is mainly through the crawler technology to the coal mine accident cases in the coal mine safety network, the use of java language to realize the crawler technology, through the connection of the webpage URL, the webpage is identified cyclically until the termination condition is reached, the specific process is shown in Figure 2.

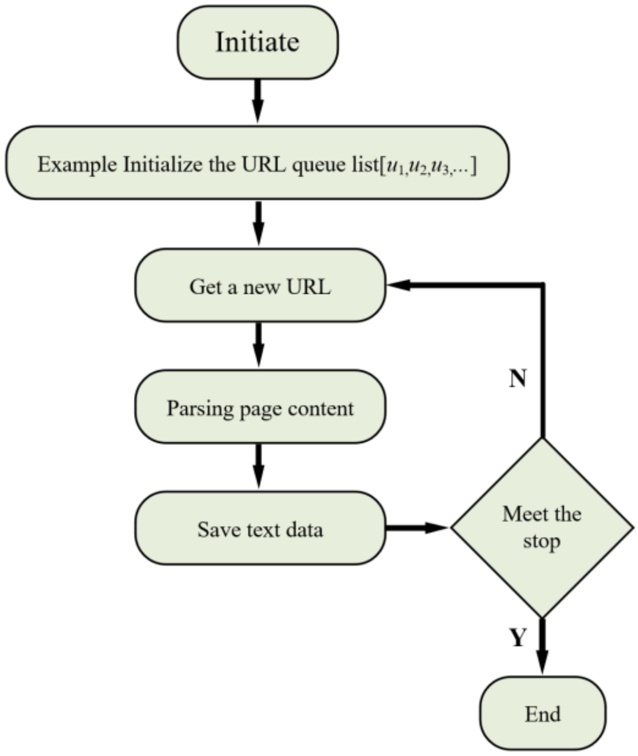

Figure 2. Data acquisition flow chart

Each web page page is parsed as shown in Figure 3. Get every incident link on every page with "<div class="imgr">...</div>" and then visit each incident link to get the specific text content.

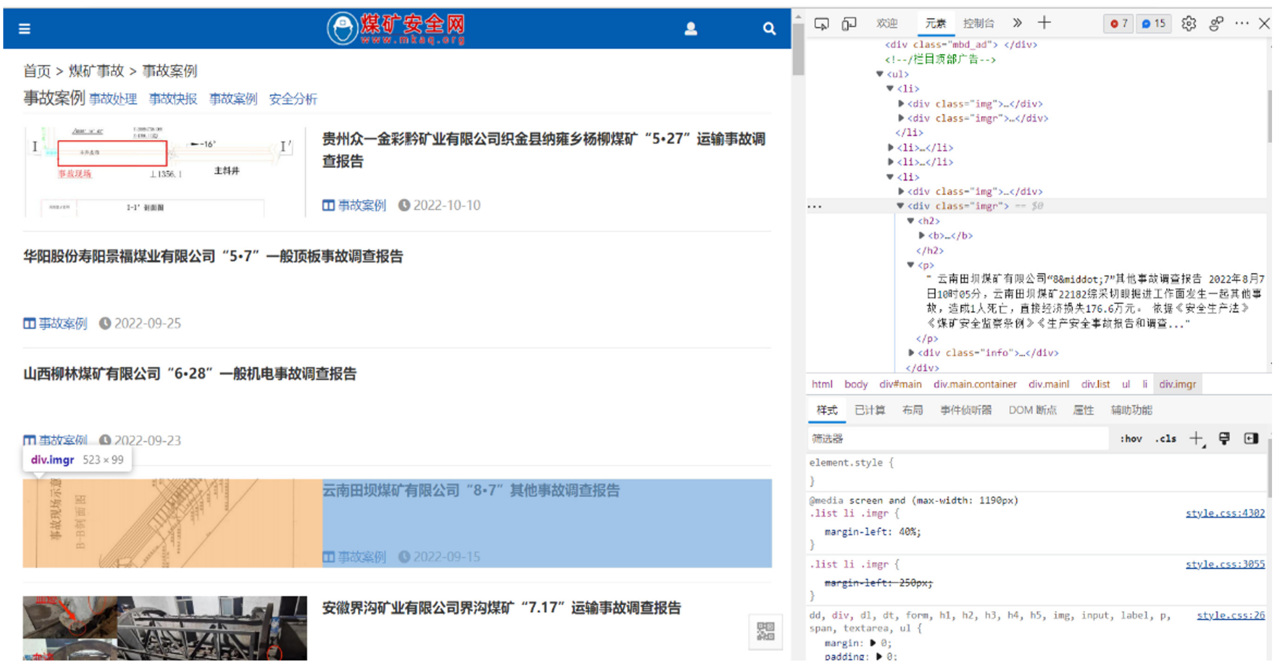

Figure 3. Coal mine accident case webpage source code

The text data obtained through the web crawler are saved in txt text, because the data obtained by the crawler will have some repetitive text, as well as the text contains a lot of useless punctuation, in order to improve the quality of the subsequent entity recognition as well as relationship extraction, so it is necessary to clean the text content.

2.2. Coal Mine Accident Data Entity Recognition

Text entity recognition of coal mine accidents is a key step in the construction of coal mine safety knowledge graph. As a research hotspot in natural language processing, entity recognition can be roughly divided into three categories : named entity recognition based on rules and dictionaries, named entity recognition based on statistical models and rules, and named entity recognition based on deep learning. In recent years, with the development of natural language technology, deep learning has been widely used in the field of named entity recognition, and has achieved good results. Therefore, this paper uses the named entity recognition method based on deep learning to obtain the entities in the field of coal mine safety. According to the characteristics of coal mine accident text, eight entities are set up, namely : time, place, mine name, number of casualties, fault phenomenon, fault location, direct cause and indirect cause. And the BIO format is used to label the acquired coal mine data, where B represents the beginning of the entity, I represents the middle or end of the entity, and O represents the non-entity. The ALBERT-BiLSTM-CRF model is used to extract entities from data texts in the field of coal mine safety. The model is shown in Figure 4.

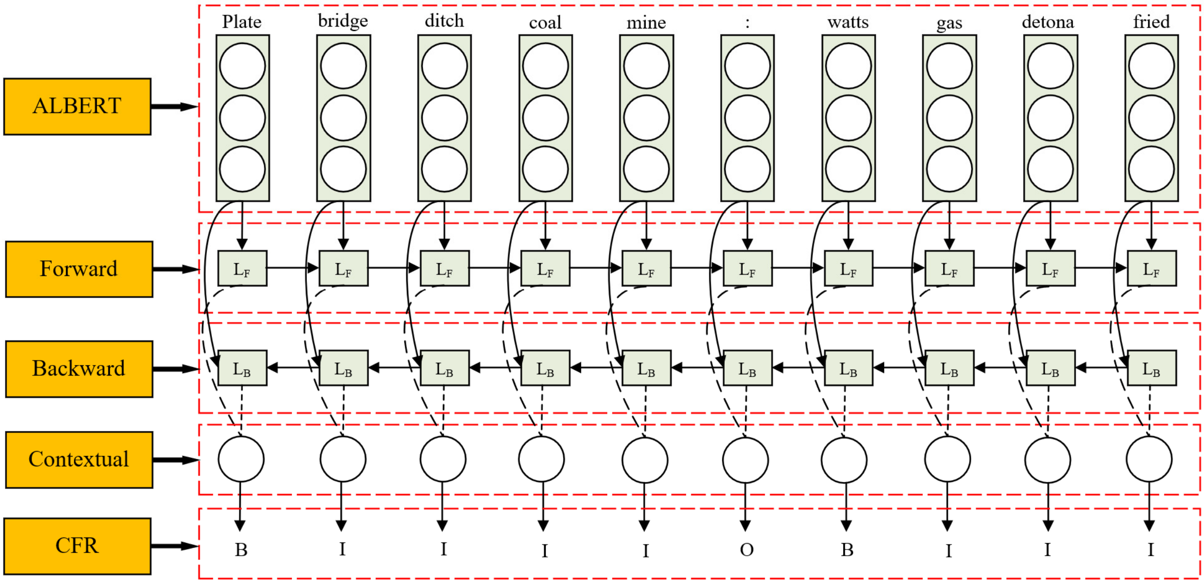

Figure 4. ALBERT-BiLSTM-CRF model

As shown in Figure 4, the whole model is divided into three parts: ALBERT layer, BiLSTM layer and CRF layer.

(1) The ALBERT layer is used to obtain the word vector ($x_1, x_2, x_3 \dots x_n$) of the sentence as the input of the next layer of the model.

(2) The BiLSTM layer is used as the coding layer of the model. By combining the output of the forward LSTM with the output of the backward LSTM, the context semantics of the input sentence is obtained, and the implicit state sequence of the input sentence is obtained, such as formula (1).

$$(h_1, h_2, \cdots h_n) \in R^{n*m} \tag{1}$$

(3) The CRF layer is used to receive the output of the BiLSTM layer. The output of the BiLSTM layer is input to the CRF layer to obtain the label sequence of the sentence $y = (y_1, y_2, \cdots y_n)$, where the label score of y is :

$$s(y|x) = \sum_{j=1}^{m}\sum_{i=1}^{n} \mu_i f_j(x, i, y_i, y_{i-1}) \tag{2}$$

In the formula, i represents the position of the word in the sentence, y_i is the label of the current word, y_{i-1} is the label of the previous word, m corresponds to the number of features, n represents the length of the input sentence.

The obtained score is converted into a probability value after normalization, and the maximum probability value in the output result is used as the final sequence label.

When training on the ALBERT-BiLSTM-CRF model, the standard used in the model is Precision, Recall, f_1-score, and the specific formula is :

$$P = \frac{TP}{TP + FP} \tag{3}$$

$$R = \frac{TP}{TP + FN} \tag{4}$$

$$f_1 = \frac{2 * P * R}{P + R} \tag{5}$$

Table 1. Experimental training results

Model	P	R	F1
ALBERT-BiLSTM-CRF	71.05	70.26	70.65
BERT-BiLSTM-CRF	62.43	61.73	62.08
BiLSTM-CRF	58.13	60.14	59.12

In order to prove the validity of the model proposed in this paper, two sets of comparative experiments are selected to compare with the model. Through the training of experimental data, the results are shown in Table 1.

It can be seen from Table 1 that the ALBERT-BiLSTM-CRF model proposed in this paper has an increase of 8.62 % in P value, 8.53 % in R value, and 8.57 % in F1 value compared with the BERT-BiLSTM-CRF model. Compared with the BiLSTM-CRF model, the ALBERT-BiLSTM-CRF model increased the P value by 12.93 %, the R value by 10.12 %, and the F1 value by 11.53 %. Therefore, ALBERT-BiLSTM-CRF has a better recognition effect on named entity recognition in the field of coal mine safety.

2.3. Entity Relationship Extraction

Relation extraction is an important part of knowledge graph construction. In this section, ALBERT-BiLSTM-Att model is used to extract text data of coal mine accidents, and the relationship between entities is stored in the form of triples. According to the entity category, eight relationships can be defined, including accident time (the time when the accident occurred), accident location (the place where the accident occurred), accident coal mine (the name of the mine where the accident occurred), result (the number of deaths caused by the accident), specific location (the working face where the accident occurred), accident phenomenon (the phenomenon of accident occurrence), primary cause (the direct cause of the accident), and secondary cause (the indirect cause of the accident).

The model algorithm mainly uses the ALBERT layer to complete the pre-training task, and obtains a low-dimensional dense vector containing word feature information. The BiLSTM layer is used to obtain the semantic context features, and the input word vector is bidirectionally encoded to obtain the output $h=[h_1,h_2,h_3,\cdots h_N]$. The Att layer is used as the last layer of the model, which is mainly used to assign the weight of the words in the sentence. The greater the contribution to the entire sentence, the higher the weight obtained. The specific calculation formula is :

$$q_i = W_q h_i + b_q \tag{6}$$

$$k_i = W_k h_i + b_k \tag{7}$$

$$v_i = W_v h_i + b_v \tag{8}$$

In the formula, q_i, k_i, v_i are query vector, key vector, value vector, W_q , W_k , W_v is the corresponding parameter matrix, b_q , b_k , b_v is the corresponding parameter matrix. The specific calculation process of ATT is shown in Figure 5.

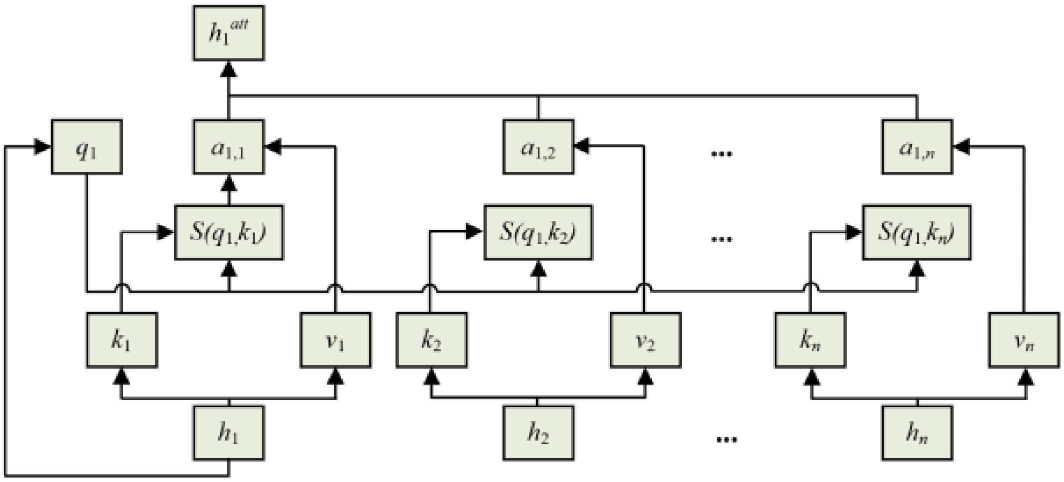

Figure 5. Self-attention mechanism calculation process

As shown in Figure 5, where q_i, k_i are used to calculate the correlation between the two vectors, the formula is as follows :

$$s(q_i,k_j) = (q_i)^T \cdot k_j \tag{9}$$

Normalize $s(q_i,k_j)$ and calculate $a_{i,j}$.

$$a_{i,j} = \frac{\exp(s(q_i,k_j))}{\sum_{t=1}^{N}\exp(s(q_i,k_t))} \tag{10}$$

Through the formula calculation in the above steps, the final output vector is:

$$h_i^{att} = \sum_{j=1}^{N} a_{i,j} y_j \tag{11}$$

The model algorithm flow chart is shown in Figure 6.

The experimental data are trained by ALBERT-BiLSTM-Att model, and the evaluation indexes of the experimental results are still taken to measure P, R, and F1 in 2.2, and the experimental results of 8 relationship categories are obtained, as shown in Table 2.

Table 2. Extraction results of relation types

Relation type	P	R	F1
Occurrence time	78.63	76.12	77.38
Accident location	80.00	81.02	80.51
Accident mine	84.15	82.13	83.13
Result in	72.16	70.84	71.50
Specific location	63.42	61.05	62.21
Accident phenomenon	60.72	58.26	59.46
Primary cause	54.71	56.87	55.77
Secondary cause	51.67	53.29	52.47

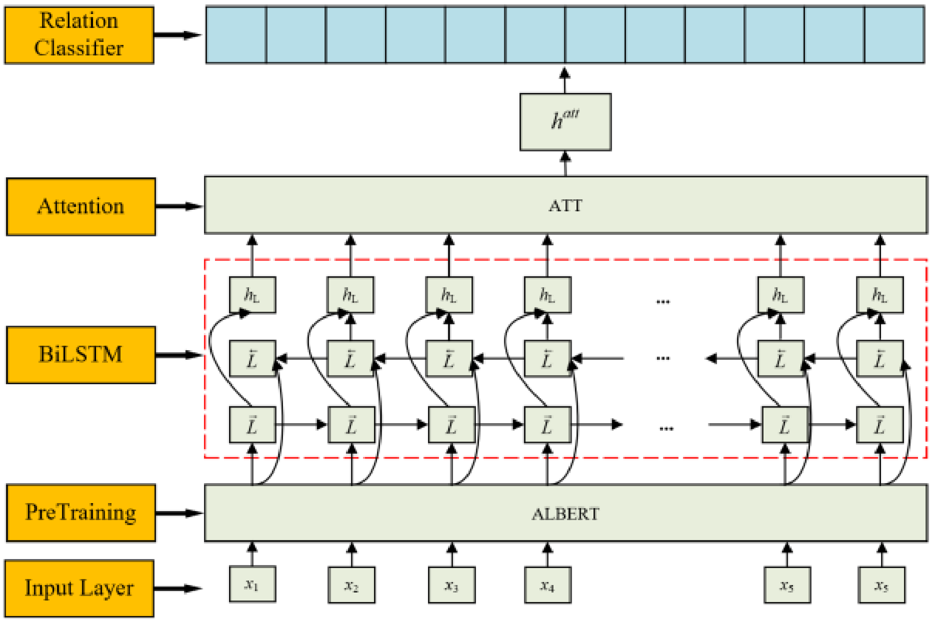

Figure 6. Relational extraction algorithm model

As can be seen from Table 2, the P, R, and F1 of the primary and secondary causes are lower, mainly due to the fact that the primary and secondary cause sentences are longer and more complex, so the P, R, and F1 obtained when performing the relational extraction are lower. The experimental results of the ALBERT-BiLSTM-Att model are shown in Table 3.

Table 3. Results of model experiments

	P	R	F1
BERT-BiLSTM-Att	78.63	76.12	77.38
ALBERT-BiLSTM-CRF	54.71	56.87	55.77
ALBERT-BiLSTM-Att	51.67	53.29	52.47

In order to prove the effectiveness of the proposed model, three groups of comparison experiments are set up, which can be seen in Table 3 that the ALBERT-BiLSTM-Att model improves P, R, and F1 by 7.05%, 0.74%, and 3.86%, respectively, compared with the ALBERT-BiLSTM-CRF model, which shows that the introduction of the self-attention mechanism model in the relational extraction and the enhancement of the keyword's weight in the sentences can be effectively improve the performance of relation extraction. Comparing the BERT-BiLSTM-Att model with the ALBERT-BiLSTM-Att model, the latter has a slightly higher relational extraction performance in the field of coal mine safety, and it can be seen that the ALBERT model has a better model performance in relational extraction compared to the BERT model.

3. Coal Mine Safety Knowledge Graph Visualization

In this paper, coal mine accident text case data is used to get coal mine ternary data through two techniques of entity recognition and relation extraction, and Neo4j database is used to construct coal mine safety knowledge graph, which associates all the same accident cases as well as displays them in the form of graphs. The constructed graph contains 8 kinds of entities and corresponding 8 kinds of relationships, using Neo4j to store the ternary, and querying the trend of accidents in a certain region through Cypher.

In the coal mine safety knowledge graph, all the same accidents are associated with each other, so that practitioners can view all the same accidents at the same time and analyze them to get the hidden information in the accidents, so as to better prevent accidents from occurring.

4. Conclusion

This paper takes seven kinds of coal mine accident text data as the research object, processes the unstructured coal mine text data through entity recognition technology and relationship extraction technology, gets the structured ternary data imported into the Neo4j database, constructs the coal mine safety knowledge mapping, and visualizes the coal mine knowledge in the form of mapping. By visualizing the map, we can analyze the correlation between the same accidents and prevent the recurrence of accidents. Due to the preliminary construction of the coal mine safety knowledge map, there are deficiencies in the paper, the coal mine knowledge is not comprehensive enough, and we will continue to improve and expand the coal mine knowledge base in the future to build a more complete coal mine safety knowledge map.

Acknowledgments

The authors gratefully acknowledge the financial support from Anhui Provincial Education Department's Higher Education Quality Project 'Artificial Intelligence New Professional Quality Improvement Project' (2023xjzlts088) and Anhui Xinhua University's school-level scientific research project (2024zr016).

References

[1] Ma Rui-Xin, Li Ze-Yang, Chen Zhi-Kui, et al. Review of reasoning on knowledge graph[J]. Computer Science, 2022, 49 (S1): 74-85.

[2] Peng Yunong, Liu Pingzeng, Zhang Yan. Construction of knowledge mapping of maize production processes based on deep learning [J]. Journal of Chinese agricultural Mechanization, 2025, 46(2): 245-252.

[3] Zeng Ziling, Zhang Huamin, Yu Tong, et al. Review on the Research and Application of Knowledge Graph and Its Key Technologies in the Field of Traditional Chinese Medicine[J]. World Science and Technology -Modernization of Traditional Chinese Medicine, 2022, 24(02): 780-788.

[4] Pu Tianjiao, Han Xiao. Key Technologies of Artificial Intelligence Application in New Power Systems [J]. Electric Power Information and Communication Technology, 2024, 22(01): 1-13.

[5] LI Zhe, ZHOU Bin, LI Wenhui, et al. Construction and application of mine electromechanical equipment accident knowledge graph[J]. Industry and Mine Automation, 2022, 48(1): 109-112.

[6] LIU Peng, YE Shuai, SHU Ya, et al. Coalmine safety: knowledge graph construction and its QA approach [J]. Journal of Chinese Information Processing, 2020, 34(11): 49-59.

[7] MAOR B, ZHUJ, LI A W, et al. Construction of knowledge graph of industry chain based on natural language processing [J]. Journal of the China Society for Scientific and Technical Information, 2022, 41(3): 287-299.