

Research Article

A Knowledge Graph System for the Maintenance of Coal Mine Equipment

Guozhen Zhang ^{1,2}, Xiangang Cao,^{1,2} and Mengyuan Zhang^{1,2}

¹School of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an 710054, Shaanxi, China

²Key Laboratory of Intelligent Testing of Mine Mechanical and Electrical Equipment, Xi'an 710054, Shaanxi, China

Correspondence should be addressed to Guozhen Zhang; 1318649644@qq.com

Received 5 August 2021; Revised 11 October 2021; Accepted 27 October 2021; Published 10 November 2021

Academic Editor: Rahib Abiyev

Copyright © 2021 Guozhen Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of coal mine intelligent technology, the complexity of coal mine equipment has been continuously improved and the equipment maintenance resources have been continuously enriched. The traditional coal mine equipment maintenance knowledge management technology can no longer meet the current needs of equipment maintenance knowledge management, and the problems of low utilization rate, poor interoperability, and serious loss of knowledge have gradually emerged. It is urgent to study new knowledge system construction and knowledge management application technology for large-scale coal mine equipment maintenance resources. Knowledge graph is a technical method to describe the relationship between things in the objective world by using a graph model. It can effectively solve the problem of knowledge dynamic mining and management under large-scale data. Therefore, this paper focuses on the establishment of a coal mine equipment maintenance knowledge graph system by using knowledge graph technology. The main research contents are as follows: Firstly, based on the current situation that there is no unified basic knowledge system in the field of coal mine equipment maintenance, this paper establishes the coal mine equipment maintenance ontology (CMEMO) to effectively solve the problem that there are no unified representation, integration, and sharing of coal mine equipment maintenance knowledge in this field and provide support for the construction of coal mine equipment maintenance knowledge graph. Then, aiming at the problem that the traditional named-entity recognition method has a poor recognition effect and relies too much on artificial feature design, this paper proposes a named-entity recognition model for coal mine equipment maintenance based on neural network (BERT-BiLSTM-CRF) and applies the model to the coal mine equipment maintenance data set for verification. The experimental results show that, under the same data set, the entity recognition effect of this model is more leading than that of other models. Finally, through demand analysis and architecture design, combined with the constructed ontology model of coal mine equipment maintenance field, the entity identification of coal mine equipment maintenance is completed based on the BERT-BiLSTM-CRF model and the Django application framework is used to build the coal mine equipment maintenance knowledge graph system to realize the functions of each module of the knowledge graph system.

1. Introduction

In recent years, traditional coal mines (CMs) have been developing gradually toward informatization, automation, and intelligence [1]. The essence of CM intellectualization is that, from development design to production management, the main CM systems should have the basic abilities of self-perception, self-learning, self-decision-making, and self-execution [2]. The concept of intelligent CMs is aligned with

the technological revolution currently taking place in the coal industry, and it provides core technical support for high-quality development of that industry [3, 4]. In the process of intelligent CM development, as CM equipment (CME) becomes more complex, so CM enterprises have higher requirements for the reliability and supportability of CME maintenance (CMEM) and have accumulated many knowledge resources related to CMEM. There are many problems with using CMEM knowledge resources [5–7],

including the following: (1) *Knowledge Representation*. Currently, there are no scientific, effective, and unified means of knowledge expression and structure composition in CMEM. Structured knowledge is stored mostly in tables or forms, semistructured knowledge exists mostly in web pages related to CMEM, and unstructured knowledge is often retained as text in maintenance manuals, case bases, and standard books. These types of knowledge are not closely related and often exist independently; therefore, it is difficult to obtain knowledge with a unified form of expression. (2) *Establishing a Knowledge Management System*. Because the coal industry currently does not attach great importance to CMEM knowledge and does not provide professional personnel for knowledge management, there is currently no complete knowledge management system that can deal quickly and efficiently with knowledge from a wide range of sources with diverse structures. (3) *Knowledge Sharing*. Because each system is designed independently, CMEM personnel often rely on personal experience, maintenance manuals, and expert connections in equipment maintenance, so the scope of knowledge retrieval is small and the efficiency is low, thereby prolonging the downtime of equipment.

To solve the above problems effectively, at present, some foreign experts and scholars have analyzed and studied the big data of CMEM. Edyta Brzychczy applied the concept of PM for the first time to study the normal working process of roof bolters working in underground mines so as to improve the operation efficiency of assets [8]. Kinga Stecula used the intelligent informatics module for registration and assessment of causes of breaks in selected mining machines [9]. Joel Löw formulated the notion of Mining 4.0 (Industry 4.0 in the mining industry), where we try to create an image of how the future might look from a miner's perspective and how mining companies may navigate their way to a future that works for all miners [10]. Jaroslaw Brodny and Magdalena Tutak put forward the total productive maintenance (TPM) strategy to improve the use efficiency of the equipment and machines owned by the company [11]. In the management of CMEM knowledge, although many domestic CMs have moved gradually from traditional human inspection and management to information management based on computer technology, many CM enterprises still use manual records to manage equipment maintenance and repair, and because of having to read many maintenance manuals, consult experts, and use other methods to repair and maintain equipment, the query content is limited and the efficiency is low. There is an urgent need for a knowledge management method that can integrate, comb, and store CMEM resources. Therefore, given the above problems in CMEM knowledge management and from analyzing the applications of knowledge graphs (KGs) in other fields, combined with the methods proposed by some experts and scholars, KG technology is applied herein to CMEM knowledge management to solve the above problems effectively.

The concept of KG comes from the Semantic Web, which was originally used to study the transformation from the World Wide Web to the Semantic Web. It uses nodes to

express structured semantic information, constructs network diagrams according to the intersection relationship between nodes, and realizes visual display. Because of its potential value in the application of knowledge graph, it has become a new technology and method for research and application in various disciplines. In 2012, Google launched a search engine service with Freebase as the back end of KG, which triggered an upsurge of exploration of knowledge graph in various university disciplines [12]. Knowledge Atlas is very different from knowledge engineering in the traditional expert system stage. The traditional expert system mainly relies on the manual extraction of knowledge by experts, and the modern knowledge graph has the remarkable characteristics of large scale, so the traditional manual method is no longer suitable for constructing the modern knowledge graph. As shown in Figure 1, Douglas Lenat created the common sense knowledge base CYC in 1984. As an early knowledge base, it contains only 7 million fact descriptions [13]. WordNet is a dictionary based on cognitive linguistics, which is jointly designed by experts in various fields at Princeton University. It defines the semantic relationship between parts of speech, and there are nearly 200000 semantic relationships. At the end of the 20th century, a common sense knowledge base called ConceptNet appeared. The creation of this knowledge base integrates the knowledge of the Internet, domain experts, and games and was established by Marvin Minsky. So far, although ConceptNet 5.0 has experienced multiple rounds of updates, it only contains less than 30 million RDF triplet relationship descriptions. However, with the development of knowledge graph technology and methods, the triple volume contained in the current knowledge graph has reached tens of millions [14], such as the industry knowledge graph covering the company's core business commodity information data created by large enterprises such as Google, Baidu, and Ali. In 2013, Sogou also announced its own search engine "Knowledge Cube" [15], and Microsoft also released "Satori" [16] in the same year.

With the continuous improvement of KG technology, the application of KGs in academic research and industry has developed [17]. To promote intelligent CM development, proposed here is building a KG of CMEM; this would integrate the scattered CMEM data and information effectively and upgrade them to a CMEM knowledge base with system logic and a precise structure, thereby reducing the "knowledge islands" phenomenon of CMEM and realizing higher-level sharing of CMEM knowledge.

KGs did not emerge by accident; their historical origin can be traced back to semantic networks in the 1960s [18, 19]. The related technologies include ontology, knowledge representation, and natural language processing (NLP), and the specific development process is shown in Figure 2.

The KG concept originated from transforming the World Wide Web into the Semantic Web. This involves using nodes to express structured semantic information, constructing network graphs according to the intersection relationships among nodes, and realizing a visual display [20]. Because of their potential application value, KGs now constitute an emerging technology and method in research

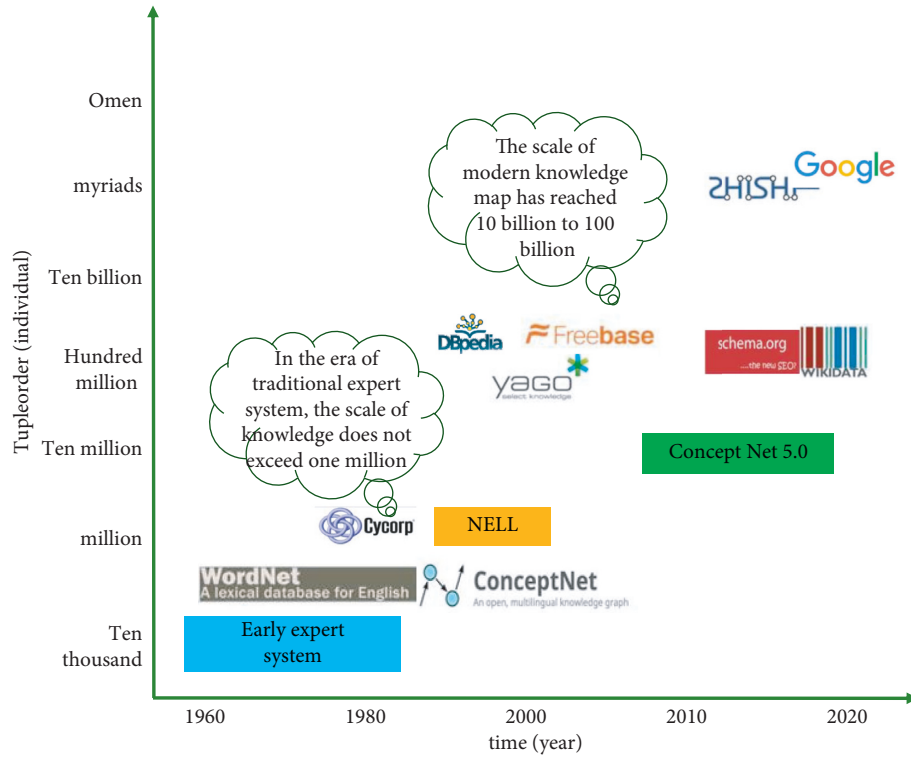


FIGURE 1: Large-scale development of the modern knowledge graph.

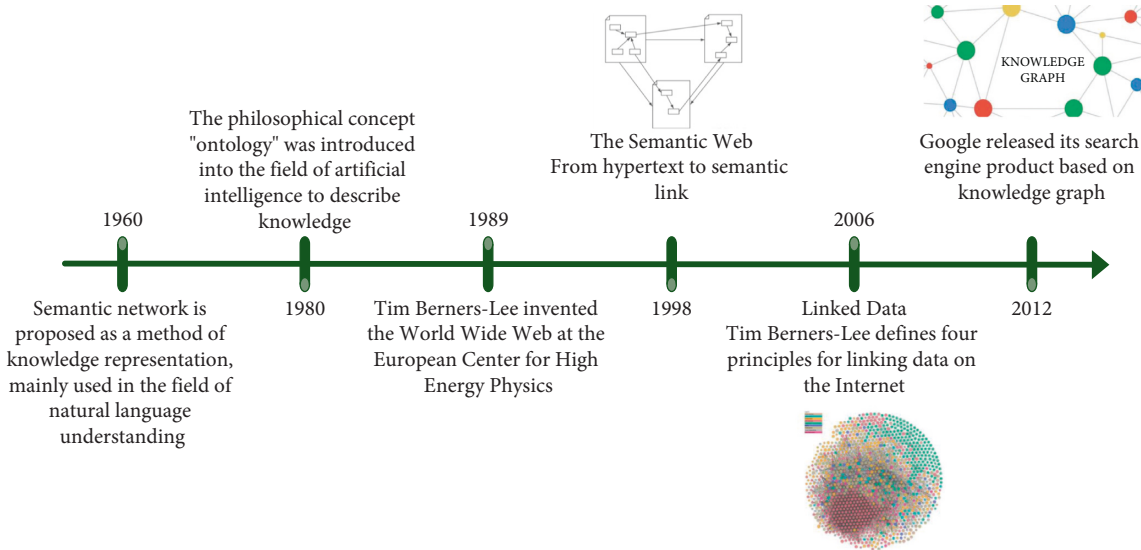


FIGURE 2: Development history of the knowledge graph.

and applications in various disciplines. In 2012, Google launched a search engine service with Freebase as the KG back end, thereby triggering an upsurge in KG exploration in various university disciplines [21, 22].

2. Construction of the Knowledge Graph System for CMEM

Aimed at the current low utilization rate, poor interoperability, and serious loss of knowledge in CMEM knowledge management, we propose building a KG of CMEM and

using the related advanced technology and theory of KGs to transform CMEM resources into valuable knowledge, thereby laying the foundation for mining CMEM knowledge. Based on the KG, we can mine the potential relationship between knowledge; using the KG of CMEM, we can realize knowledge query, knowledge management, and knowledge display so as to achieve the goal of effective organization and management of CMEM knowledge and then provide support for CMEM decision-making. The architecture of the KG system for CMEM is shown in Figure 3 and has five main parts: (i) data source acquisition

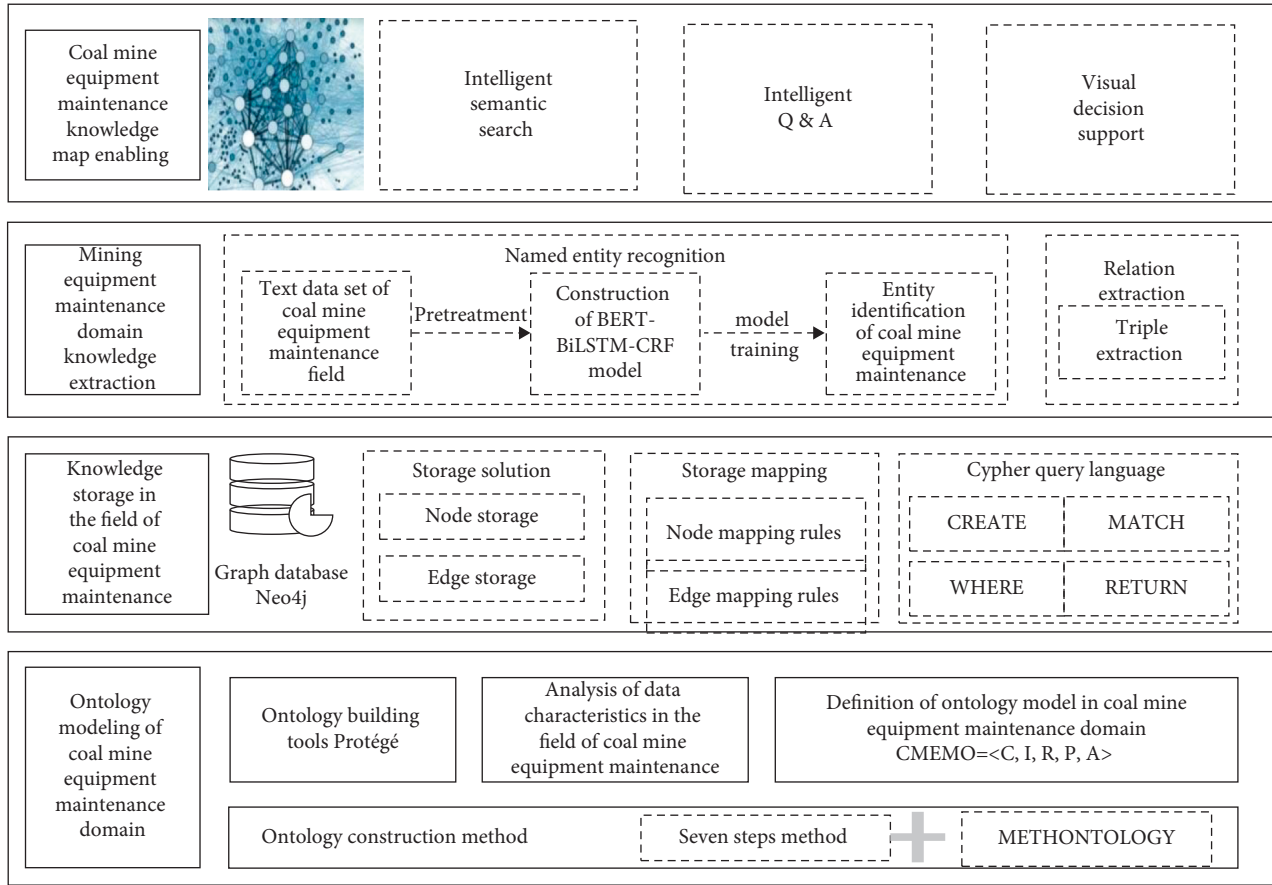


FIGURE 3: Architecture of the knowledge graph system for coal mine equipment maintenance (CMEM).

and processing, (ii) ontology modeling, (iii) knowledge storage, (iv) knowledge extraction, and (v) KG enabling.

In the layer for data source acquisition and processing, through the collection and data preprocessing of structured, semistructured, and unstructured knowledge in CMEM, the CMEM corpus is obtained, cleaned, segmented, and annotated to provide a reliable corpus for constructing the CMEM KG.

In the layer for ontology construction, the technical terms and data characteristics in CMEM are analyzed, and then the ontology model of the field is defined. Finally, according to the requirements of ontology construction in the field, the task of ontology construction in CMEM is completed by combining the seven-step method with the METHONTOLOGY method.

In the layer for knowledge storage, Neo4j—the most popular graph database—is selected to store the knowledge of CMEM. According to the storage requirements of the CMEM KG, the storage scheme and graphing rules are formulated to realize the storage of CMEM knowledge.

The layer for knowledge extraction is divided into two parts: entity recognition and relation extraction. CMEM

named-entity recognition is based on the BERT-BiLSTM-CRF model.

Based on the realization of subtasks in the above levels, the KG enabling layer constructs the KG of CMEM and enables it to realize knowledge query, management, and display.

3. Ontological Modeling of the CMEM Domain

3.1. Ontological Definition of the CMEM Domain. By collecting and analyzing CMEM data, summarized here are four characteristics of such data: (i) multisource heterogeneity, (ii) complex hierarchy, (iii) strong correlation, and (iv) rapid change. Based on these characteristics and combined with ontology modeling metalanguage, the ontological model of CMEM (CMEMO) is established. The definition of CMEMO is as follows: the CMEM domain ontology is a five-tuple CMEMO = <C, I, R, P, A>, where C, I, R, P, and A are as follows:

- (1) C is the class set of CMEM knowledge. A class set is a set of key concepts in the process of CMEM that have

nothing to do with the specific content, for example, CME and fault maintenance.

- (2) I is the instance set of the ontology; for example, build a specific equipment MG200/500-AWD shearer in the coal mine equipment class.
- (3) R is the set of relations on C and refers to the relationships among CMEM ontology classes, such as the utilization relationship between the equipment-status class and the fault maintenance class. By defining the relationships among all the classes in the CMEM ontology, the complex semantics in CMEM can be described, thereby providing a logical basis for CMEM reasoning.
- (4) P is the attribute set of the CMEM domain ontology.
- (5) A is the set of axioms; for example, “roadheader belongs to fully mechanized mining equipment” is an axiom.

3.2. Construction Criteria for the CMEM Ontological Model.

In constructing the CMEM ontology, to use CMEM knowledge efficiently and accurately and minimize the complexity of the ontological knowledge, the following principles should be followed in the modeling process as shown in Figure 4:

- (1) *Clarity*. In the process of building the model, it is required to ensure the clarity of the ontology model in the division of conceptual level so that users can understand and use the model better. In constructing the CMEM ontological model, it is necessary to clarify the key tasks, elaborate on the main elements, and describe the relationships among the elements. At the same time, it is necessary to select the appropriate ontological description language according to the requirements of ontological construction, and the description language must be clear and easy to understand.
- (2) *Consistency*. To realize the reasoning of CMEM knowledge in the later stage and ensure the quality of the CMEM ontology, the key concepts must be determined correctly and the use of terms must conform to the definition of domain terms so as to ensure effectively the correctness of the knowledge reasoning.
- (3) *Scalability*. CMEM knowledge is constantly expanding in its depth and breadth, and new maintenance methods and fault maintenance schemes must be updated in time. Therefore, it is necessary to ensure that every update of the ontology in the field of CMEM reduces the impact on other concepts as much as possible.
- (4) *Reusability*. In the domain ontology of CMEM, the reusability of ontology can be improved by using multilevel key concept classes and the inheritance among concept classes. That is to say, in CMEM, ontological information has operability and interoperability at the conceptual level, which is

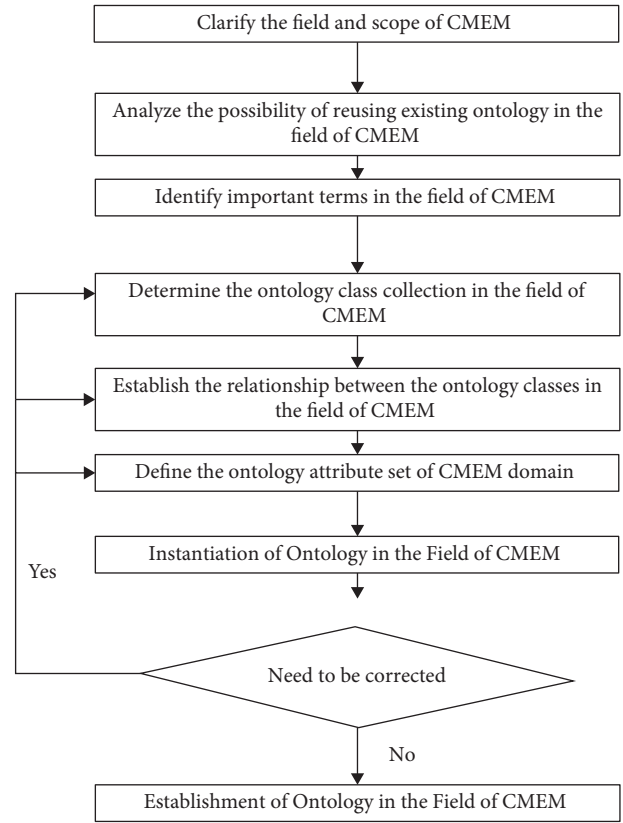


FIGURE 4: Process of constructing the CMEM ontological model.

convenient for using CMEM knowledge efficiently in the later stage.

3.3. Process of Constructing the CMEM Ontological Model.

In the process of building ontology with the seven-step method, the updating of ontological content cannot be responded to in time, but CMEM data have the characteristic of fast updating. In the process of building domain ontology, there are often changes in knowledge or a need for updating, so the seven-step method is no longer applicable. By contrast, the METHONTOLOGY method modifies and revises the ontology after every operation in the process of ontology construction, thereby compensating for the above shortcomings to a certain extent. Therefore, proposed herein is an ontology construction method based on the seven-step method by introducing the feedback mechanism of the METHONTOLOGY method to compensate for the lack of information feedback in the seven-step method. The specific development process is shown in Figure 3.

The first step for CMEM is to determine the scope of the field to ensure the correctness of the top-level mode construction of the domain knowledge graph. Second, the reusability of existing ontology such as shearer fault diagnosis is analyzed to reduce the cost of ontology construction. Then, the CMEM classes, their attributes, and the relationships among classes are determined, and the ontological system for CMEM is established. Finally, the evaluation index in the quality system standard is used to evaluate the

quality of the constructed ontology in the field of CMEM, evaluate the comprehensive ability of the data source in the field of CMEM according to the ontology, and finally complete the construction of the ontology in the field of CMEM.

4. CMEM Knowledge Storage

4.1. Storage Scheme Based on the Neo4j Graph Database. In KGs, relational database management systems, resource description framework (RDF) triples, and graph database systems are usually used to store knowledge [23]. When building the CMEM knowledge graph, because of the characteristics and application requirements of CMEM knowledge, the stored entities and the relationships among entities often must be queried and updated. However, a relational database will add new tables and fields in each processing, thereby increasing the difficulty and cost of development. With its special data storage scheme and query processing mechanism, a graph database can realize efficient data storage and query by inserting nodes and edges [24]. Therefore, this study uses Neo4j—the most popular native graph database—to realize the query and storage of the knowledge graph. In a graph database, node attributes define entity attributes, and the node label represents the entity class to which it belongs. Because the label is not unique, some entities belonging to multiple classes can be defined using different labels.

In CMEM, the complex graph structure of the ontology and the storage process of the Neo4j graph database involve three main aspects: (i) analyzing the OWL (Web Ontology Language) files of the CMEM domain ontology, (ii) realizing the relational mapping between the RDF graph and the Neo4j graph database, and (iii) data loading. The specific process is shown in Figure 5, and Table 1 gives the knowledge storage scheme based on Neo4j, showing how to store entities and their relationships in the CMEM knowledge graph.

The ontology of the CMEM domain is mapped to the Neo4j graph database by using a knowledge storage mapping algorithm, and the ontological model and knowledge storage of the CMEM domain are realized. The specific implementation steps include reading the parsed RDF file, traversing all triples in the file, and realizing knowledge storage according to storage mapping rules. The related knowledge storage mapping algorithm is given in Algorithm 1.

5. Named-Entity Recognition of CMEM Based on the BERT-BiLSTM-CRF Model

Based on the constructed ontology model of the coal mine equipment maintenance field, this chapter uses the BERT-BiLSTM-CRF model to complete the named-entity recognition of coal mine equipment maintenance according to the text data retained by each subsidiary of A coal mine group, such as coal mine equipment maintenance manual.

5.1. CMEM Named-Entity Recognition Process. Aimed at the problems of the huge amount of CMEM entities, the high frequency of alternation, and the complex coal semantics in CMEM entity recognition, there is a need to select an appropriate named-entity recognition method. Because the statistical learning method based on supervision relies on a large annotated corpus for model training in the process of entity recognition, it is unsuitable for CMEM without a large-scale professional corpus and is prone to inaccurate entity recognition. Therefore, herein an improved neural network model is used to realize CMEM entity recognition. Combined with the bidirectional neural network model of conditional random fields (BiLSTM-CRF), based on the BiLSTM-CRF neural network model, the BERT pretraining language model is introduced as the named-entity recognition model of CMEM (abbreviated as BERT-BiLSTM-CRF). First, the preprocessed data are divided into a training set and a test set: the training set transforms the input text sequence into a 768-dimensional word vector with rich context semantics through the Bidirectional Encoder Representations from Transformers (BERT) model, the output vector of the BERT model is used as the input of the BiLSTM model to extract the eigenvalues of the context, the CRF model is used to label the entity of the input annotation sequence, and finally, the recognition result is obtained. The specific entity identification process is shown in Figure 6.

5.2. Named-Entity Recognition (NER) Model of CMEM. Herein, BERT is introduced into the bidirectional long short memory network (BiLSTM) combined with a conditional random field (CRF), and a named-entity recognition model based on BERT-BiLSTM-CRF is constructed. The BERT-BiLSTM-CRF model transforms the entity recognition task directly into a sequence annotation problem; that is, by constructing the BERT-BiLSTM-CRF sequence annotation model, the model structure is as shown in Figure 7. The sequence annotation model has five layers: (i) the input layer, (ii) the BERT layer, (iii) the BiLSTM layer, (iv) the CRF layer (the input is the output result of the BiLSTM layer), and (v) the output layer. The input layer represents the input sequence to be labeled, and the output layer represents the labeled sequence.

5.2.1. BERT Layer. The difference between the BERT model and other models is that it adjusts the content of all layers in a joint way and uses a bidirectional language model for pretraining, which is the main difference from other language representation models. This model uses a bidirectional transformer as an encoder to pretrain depth bidirectional representation, which is different from other word vector pretraining models such as Elmo and GPT. Many studies have shown that the text feature extraction ability of the BERT model is much better than that of recurrent and convolutional neural network models. The BERT model also involves two new unsupervised prediction tasks, namely, “masked language model” and

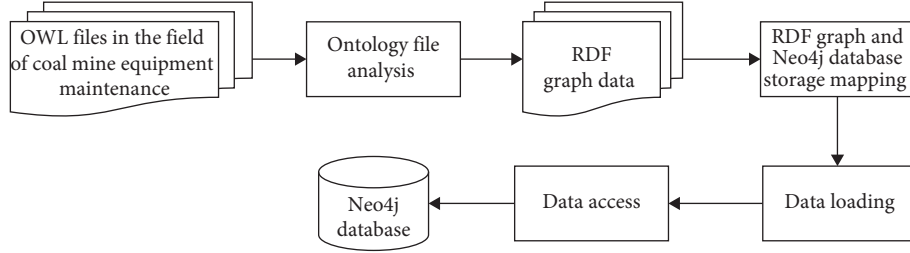
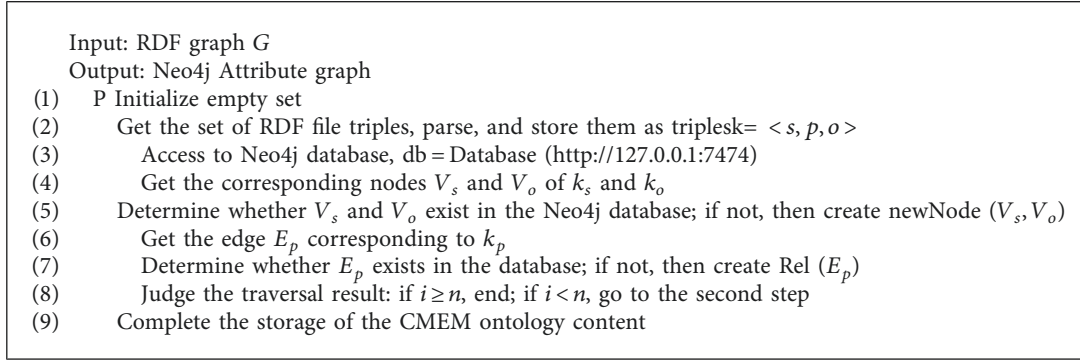


FIGURE 5: Storage process based on the Neo4j diagram database.

TABLE 1: Neo4j knowledge storage scheme.

| Category | Use | Object scope |
|----------|---|--|
| Label | Describe the concept of CMEM knowledge | CME category, maintenance plan category, fault-repair category, staff category, maintenance case category, and equipment-status category |
| Node | Describe knowledge entity of CMEM | Shearers, roadway boring machines, electrical maintenance workers, and so on |
| Edge | Describe the relationship among CMEM knowledge entities | Ownership relationship, father-son relationship, utilization relationship, causal relationship, parallel relationship, and so on |



ALGORITHM 1: Knowledge storage graph algorithm.

“next sense prediction.” The former is used to obtain word-level representation, and the latter is used to obtain sentence-level representation and combine the results.

5.2.2. BiLSTM Layer. The role of the BiLSTM model in the present named-entity recognition model is to capture the context features of a text sequence. The essence of this model is the combination of forward long short-term memory (LSTM) and backward LSTM. The comprehensive calculation process of the LSTM model can be summarized as follows: by forgetting and memory-updating information about the state of hidden layer cells, relevant information can be calculated and then transmitted, and irrelevant information is discarded. The specific process is shown in Figure 8. Here, x_{t-1} needs to discard part of the information before time iteration t , which can be realized by increasing the forgetting threshold. Then, some new information is added for x_{t-1} . At this time, x_t starts the next time iteration $t+1$ and transfers a copy of the long-term state information to the tanh function. Finally, we get the short-term state that has completed filtering at the output threshold.

Because the LSTM model can only capture past information when dealing with the sequence annotation problem, the reverse calculation process is added to capture future information; that is, the forward and backward LSTM models are combined to form the BiLSTM model. In the BiLSTM model, the process of coding the sentence “shearer rocker arm does not start” is shown in Figure 9. The text sequence gets a feature vector sequence $\vec{h}_t = (\vec{h}_1, \vec{h}_2, \vec{h}_3)$ through the forward LSTM coding process and another feature vector sequence $\overleftarrow{h}_t = (\overleftarrow{h}_1, \overleftarrow{h}_2, \overleftarrow{h}_3)$ through the backward LSTM coding process. \vec{h}_t and \overleftarrow{h}_t are spliced to get h_T ; that is, $h_T = [\vec{h}_t; \overleftarrow{h}_t]$, h_T is output as the hidden layer after BiLSTM superposition, and the probability of each word belonging to different labels is predicted. Then, in the softmax layer, the tag with the highest probability is obtained as the prediction value of the location. The output of the BiLSTM layer is used as the input of the CRF layer, which makes the whole model consider the correlation among class labels.

5.2.3. CRF Layer. Although BiLSTM can learn the context information of a word or phrase and can use the tag result with the highest probability when selecting the output tag, it

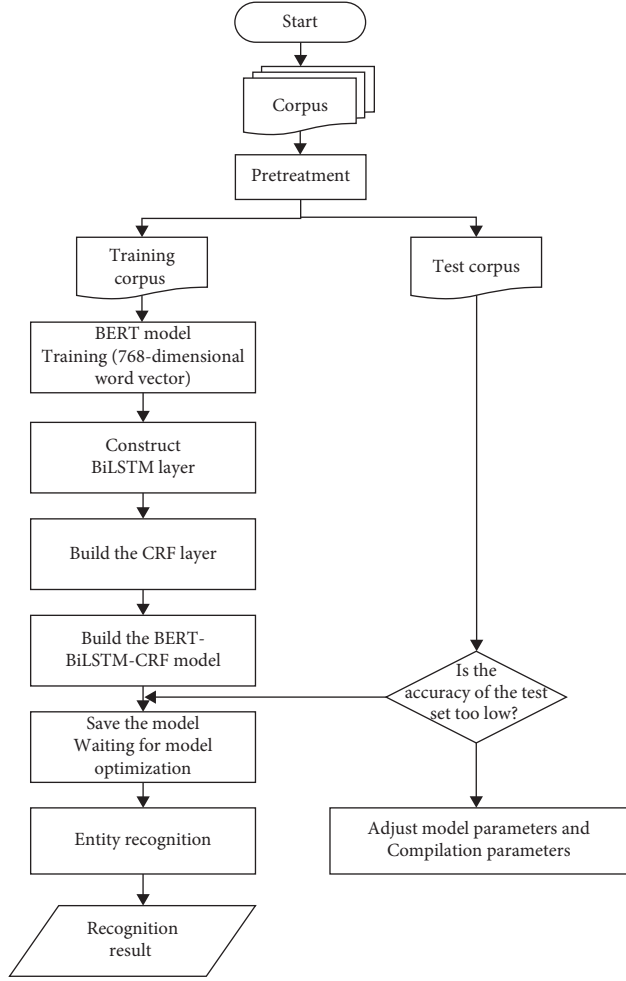


FIGURE 6: Entity recognition process based on BERT-BiLSTM-CRF.

does not take into account the relationships among different words, so the output label may be confused and lack logic. Therefore, the CRF model is introduced to solve the problem of the nonrecognition of word relationships in BiLSTM and to capture full-text information and predict the results. During CRF model training, to ensure the correctness of word labels, constraints are added when making prediction labels. According to the needs of this study, the purpose is to solve the problem of text sequence annotation, so we choose

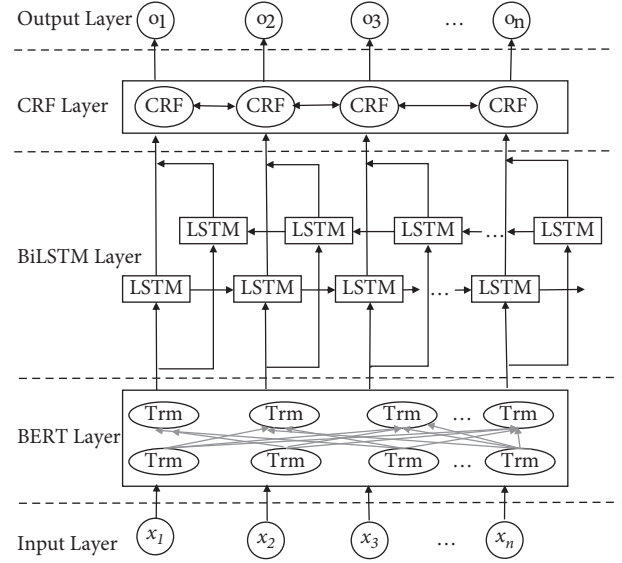


FIGURE 7: Named-entity recognition (NER) model based on BERT-BiLSTM-CRF.

to add the linear chain conditional random field (Linear-CRF is a kind of conditional random field) layer to solve the above problems effectively. The operation process of the CRF model is as follows: (i) the characteristic function f_α is defined in advance; (ii) the given data are used to train the model and determine the weight parameter λ_k ; (iii) the determined model is used to realize sequence annotation.

The CRF model is the last layer based on the BERT-BiLSTM-CRF model, and its main responsibility is to capture the dependency between the front and back tags and constrain the tags. The model can be represented by $P(Y|X)$, where X is the input variable, representing the labeled observation sequence, and Y is the output sequence, representing the tag sequence corresponding to X . Given that the random variable x is the observation sequence (x_1, \dots, x_i) , the random variable y is the hidden-state sequence (y_1, \dots, y_i) , and each (x_i, y_i) pair is the largest clique in a linear chain and satisfies

$$P(y_i|x, y_1, y_2, \dots, y_n) = P(y_i|x, y_{i-1}, y_{i+1}). \quad (1)$$

Given a preset observation sequence x , the equation for solving the hidden-state sequence y by CRF is

$$P(y|x) = \frac{1}{Z(x)} \prod_i \exp \left(\sum_k \lambda_k f_k(y_{i-1}, y_i, x, i) \right) = \frac{1}{Z(x)} \exp \left(\sum_{i,k} \lambda_k f_k(y_{i-1}, y_i, x, i) \right), \quad (2)$$

where the subscript i represents the current location of the node (token), the subscript k represents the current characteristic function, and each characteristic function is given a

weight λ_k . Because there is a limited relationship between the state sequence and the two tokens, the continuous characteristic function is defined for the CRF model, that is, the

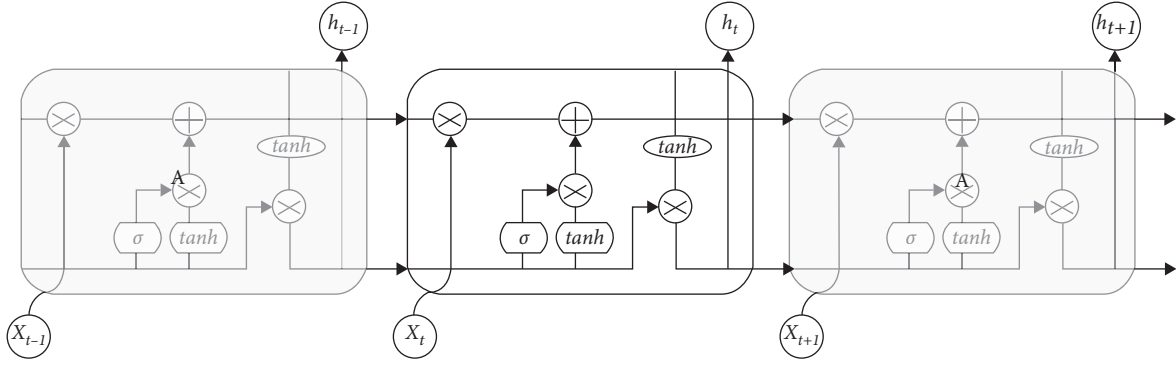


FIGURE 8: Comprehensive calculation process of the long short-term memory (LSTM) model.

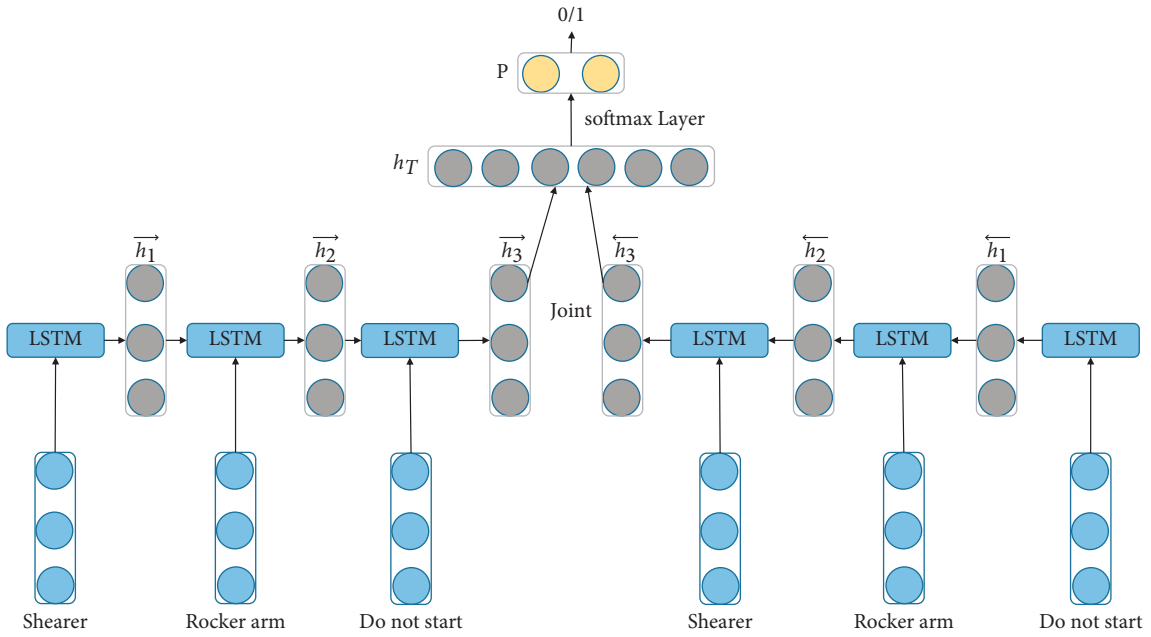


FIGURE 9: Encoding process of "shearer rocker arm does not start" by the BiLSTM model.

transfer characteristic function and the state characteristic function:

$$P(y|x) = \frac{1}{Z(x)} \exp \left(\sum_{i,j} \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_{i,j} \mu_l s_l(Y_i, x, i) \right), \quad (3)$$

$$Z(x) = \exp \left(\sum_y \exp \left(\sum_{i,j} \lambda_j t_j(y_{i-1}, y_i, x, i) \right) + \sum_{i,j} \mu_l s_l(y_i, x, i) \right),$$

where $Z(x)$ is used for normalization to form the probability value; t_j represents the transfer feature at i , and the corresponding weight is λ_j ; s_l is the state feature at i , and the

corresponding weight is μ_l ; and j and l denote the numbers of the characteristic functions.

When the value range of t_j is (1,0), the formula is

$$t_j(y_{i-1}, y_i, x, i) = \begin{cases} 1, & y_{i-1}, y_i, \text{ the value of } x \text{ satisfies the transition condition,} \\ 0, & \text{others.} \end{cases} \quad (4)$$

When the value range of s_l is (1,0), the formula is

$$s_1(y_{i-1}, y_i, x, i) = \begin{cases} 1, & y_{i-1}, y_i, \text{ the value of } x \text{ satisfies the transition condition,} \\ 0, & \text{others.} \end{cases} \quad (5)$$

Given that the weight of the state characteristic function s_l is u_l and the weight of the transfer characteristic function t_j is λ_j , the weight parameter $\lambda_k = (u_1, u_2, \dots, u_l; \lambda_1, \lambda_2, \dots, \lambda_j)$ of the CRF model is obtained. When training the CRF model,

$$\text{score} = \sum_{i,k} \lambda_k f_k(y_{i-1}, y_i, x, i), \quad (6)$$

is used to score each word, and the conditional probability model $P(Y|X)$ is calculated by maximum-likelihood estimation. In the actual prediction process, for a given observation sequence, the Viterbi algorithm is used to calculate the maximum label sequence.

6. Results and Discussion

6.1. Experiment Results. This paper uses the collected coal mine equipment maintenance data, including maintenance manual, maintenance case base, equipment maintenance plan, and equipment integrity standard, and crawls the data related to coal mine equipment maintenance from the web page through the data crawler to form the corpus set of coal mine equipment maintenance named-entity recognition. Firstly, the data is preprocessed; that is, the corpus is cleaned and the useless information is eliminated. Then Jieba word segmentation toolkit is used to segment the corpus. Finally, the tagging of the corpus set is completed according to the tagging strategy. During the experiment, the corpus is randomly divided into training set, test set, and evaluation set. The training set contains 2254 sentences and 48972 words; the test set includes 1636 sentences and 26548 words; the evaluation set contains 899 sentences and 9837 words.

In this paper, four test indexes defined by SIGHAN, precision (P), recall (R), F1 value (F1-measure, F1), and accuracy (A), are used as the evaluation indexes of named-entity recognition in this paper. They are usually used to evaluate the experimental results of retrieval tasks, classification tasks, and recognition tasks. The specific definition formula is as follows:

$$\begin{aligned} P &= \frac{TP}{TP + FP} \times 100\%, \\ R &= \frac{TP}{TP + FN} \times 100\%, \\ A &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\%, \\ F_1 &= \frac{2PR}{P + R} \times 100\%. \end{aligned} \quad (7)$$

The experiments reported herein were all completed in the same configuration environment using the same corpus to show the superiority of the CMEM entity recognition based on the proposed BERT-BiLSTM-CRF model:

- (1) *Word2ve-LSTM Model.* This model, like the Glove model, is one of the methods for word embedding. It is a neural network model used to generate word vectors. The output word vectors of this model are used as the input of the LSTM model to extract context features.
- (2) *BiLSTM-CRF Model.* This model uses the BiLSTM model to obtain the features of the text and the maximum probability of the output tag, and it takes the output result as the input of CRF to realize the constraint of the relationship between tags. It is the mainstream model for solving the problem of sequence annotation.
- (3) *BERT-CRF Model.* In this model, the context information before and after each character is learned through the BERT layer to predict the entity annotation results corresponding to each character. CRF is used to adjust the results that violate the annotation rules to reduce the error rate.

Based on the evaluation system, we introduce precision, recall, F1 (F-measure), and accuracy to analyze the entity recognition results of each model. The experimental results are given in Table 2 and show that, for the recognition effect of CMEM entities in the same data set, the recognition of the BERT-BiLSTM-CRF model is superior to that of the Word2ve-LSTM model in terms of accuracy, recall, and F1 value. Compared with the basic Word2ve-LSTM model, the accuracy, recall, and F1 value of BiLSTM-CRF are improved greatly by 5.85%, 8.49%, and 7.21%, respectively, indicating that the bidirectional neural network model has the advantage of high accuracy in named-entity recognition. Compared with BiLSTM-CRF, the F1 value of BERT-CRF has a slight increase. Compared with the BiLSTM-CRF model, the recall rate and F1 value of the recognition are increased by 2.94% and 2.86%, respectively. Compared with the BERT-CRF model, the recall rate and F1 value of the recognition are increased by 2.93% and 2.71%, respectively. This shows that the combination of BERT and BiLSTM is better for CMEM named-entity recognition.

To show the recognition of each model more intuitively, Figure 10 compares the F1 value of each model with different

TABLE 2: Comparison of experimental results.

| Model | P (%) | R (%) | A (%) | F1 (%) |
|-----------------|-------|-------|-------|--------|
| Word2ve-LSTM | 79.92 | 76.48 | 78.84 | 78.05 |
| BiLSTM-CRF | 83.77 | 82.97 | 83.53 | 83.38 |
| BERT-CRF | 85.91 | 84.98 | 85.75 | 85.44 |
| BERT-BiLSTM-CRF | 88.56 | 87.91 | 88.42 | 88.23 |

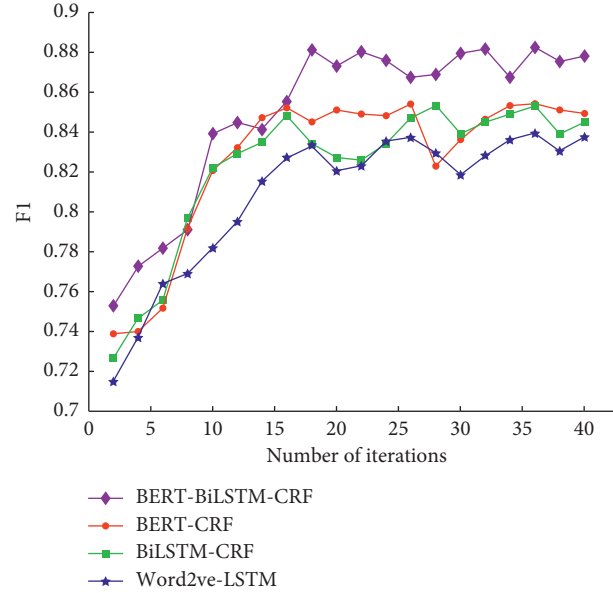


FIGURE 10: Analysis of F1 value under different training iterations.

training iterations. As can be seen, the F1 value of the BERT-BiLSTM-CRF model is higher than those of the other three models, especially after the number of training iterations is increased to 15, after which the leading of the F1 value of the BERT-BiLSTM-CRF model tends to be stable.

Based on Figure 10 and Table 2, the improved models based on Word2ve-LSTM and BiLSTM-CRF have certain limitations in CMEM named-entity recognition, mainly because these two models are limited to the feature extraction of words. The BERT-CRF model and the BERT-BiLSTM-CRF model can solve this problem well. The core of the BERT model is to build a two-way transformer to pretrain with context. Therefore, the BERT model realizes the learning of not only phrase-level information representation but also rich semantics. Upon comparing the above models, the BERT-BiLSTM-CRF model performs best in all indicators, thereby showing that the BERT-BiLSTM-CRF model performs best at entity recognition.

6.2. Discussion. Through the knowledge graph, applications such as intelligent semantic search, intelligent question and answer (Q&A), and visual decision support for CMEM can be realized. The knowledge graph can also help users efficiently query and learn CMEM concepts, maintenance schemes, and other types of knowledge, infer potential knowledge, strengthen knowledge sharing, and improve the level of CMEM:

- (1) *Intelligent Semantic Search.* Semantic search solves the problem of keyword semantic disambiguation in traditional search by using the entity links of the knowledge graph. With the help of semantic retrieval, CMEM personnel can obtain highly accurate and relevant query results, thereby helping to improve work efficiency and ensure the normal operation of CME.
- (2) *Intelligent Q&A.* A question-answering system is an advanced information service mode in which a computer can reply automatically to user questions [20]. Unlike existing search engines, a question-answering system replaces the documents with high keyword relevance with accurate natural language search answers and returns them to users. Intelligent question-answering can understand the natural language input by CMEM personnel, give the answers to the questions from the knowledge graph or target data, and evaluate the returned answers to determine the priority order. It transforms CMEM personnel from carrying out maintenance work based on experience to doing so based on data and is conducive to realizing efficient searches.
- (3) *Visual Decision Support.* This is a unified graphical interface combined with visualization, reasoning, retrieval, and so on to provide users with access to information. For example, decision support can

interpret information about shearer fault components and fault-component operation parameters in the CMEM knowledge graph through the visualization technology of Atlas, thereby helping CMEM personnel to make the best decisions and improve maintenance efficiency and ability.

7. Conclusion

With the increasing maturity of knowledge graph technology and its successful application in various fields, knowledge graph provides a new research idea for knowledge management in various industries. Based on this background, the main research work completed in this paper is summarized as follows: (1) The knowledge graph system in the field of coal mine equipment maintenance is established, and the knowledge graph technology is innovatively applied to the research in the field of coal mine equipment maintenance, so as to provide theoretical support for the research on the key technologies of knowledge graph construction in the following paper. (2) Based on the analysis of the knowledge system and characteristics of coal mine equipment maintenance, ontology is introduced into the construction of the coal mine equipment maintenance knowledge model. The coal mine equipment maintenance knowledge ontology model CMEMO is constructed. (3) The entity recognition model of BERT-BiLSTM-CRF coal mine equipment maintenance is proposed. The latest BERT model released by Google is introduced to make the accuracy and F1 value of named-entity recognition of coal mine equipment maintenance reach 88.56% and 88.23%, respectively. The experimental results show that the model has a better named-entity recognition effect. (4) The knowledge Atlas system of coal mine equipment maintenance is designed. According to the system requirements, the system architecture and functional module division are formulated, and the construction of the coal mine equipment maintenance knowledge graph is completed.

The following problems need to be further studied in the future: the coal mine equipment maintenance knowledge graph in this paper is constructed based on the collected data such as coal mine equipment maintenance standards, maintenance case base, and integrity standards of coal mine electromechanical equipment, which will inevitably be missing and wrong. At present, the knowledge Atlas of coal mine equipment maintenance has not fully covered all the knowledge in this field. In the next step, the knowledge system of coal mine equipment maintenance will be gradually supplemented and improved.

Data Availability

The experiment data used to support the findings of this study were supplied by Zhang Mengyuan under license and so cannot be made freely available. Requests for access to these data should be made to Zhang Mengyuan, 582060393@qq.com.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] G. Wang, H. Ren, and Y. Pang, "Research and engineering progress of coal mine intelligent technology system," *Coal Science and Technology*, vol. 48, no. 7, pp. 1–27, 2020.
- [2] G. Wang, F. Liu, and Y. Pang, "Intelligent coal mine—the core technology support for high quality development of coal industry," *Journal of Coal Industry*, vol. 44, no. 2, pp. 349–357, 2019.
- [3] H. Ren, W. guofa, and G. Zhao, "Intelligent coal mine information logic model and mining system decision control method," *Acta Coalae Sinica*, vol. 44, no. 9, pp. 2923–2935, 2019.
- [4] M. Li, S. Yang, and Z. Sun, "Research on framework and development prospect of smart mine," *Coal Science and Technology*, vol. 45, no. 1, pp. 121–128, 2017.
- [5] X. Han, *Research on Key Issues of Intelligent Mine Information Standardization System*, China University of Mining and Technology, Beijing, China, 2016.
- [6] S. Deng, *Design and Implementation of Large Equipment Maintenance Information Management Platform in Open Pit Coal Mine*, Inner Mongolia University, Hohhot, China, 2019.
- [7] E. J. Ding, Y. B. Liao, and L. Zhang, "Review and prospect of coal mine informatization construction," *Industrial and Mining Automation*, vol. 46, no. 7, pp. 5–11, 2020.
- [8] E. Brzychczy, P. Gackowiec, and M. Liebetrau, "Data analytic approaches for mining process improvement-machinery utilization use case," *Resources*, vol. 9, no. 2, p. 17, 2020.
- [9] K. Stecula, J. Brodny, and M. Tutak, "Use of intelligent Informatics module for registration and assessment of Causes of Breaks in selected mining machines," in *Advances in Intelligent Systems and Computing*, A. Burduk and D. Mazurkiewicz, Eds., Springer, Cham, Switzerland, 2018.
- [10] J. Löow, L. Abrahamsson, and J. Johansson, "Mining 4.0—the impact of new technology from a work place perspective," *Mining, Metallurgy & Exploration*, vol. 36, pp. 701–707, 2019.
- [11] Jaroslaw Brodny and Magdalena Tutak, "A fast, versatile and cost-effective proceedings publication service," *IOP Conference Series: Earth and Environmental Science*, vol. 95, Article ID 042019, 2017.
- [12] S. Amit, *Introducing the Knowledge Graph*, America: Official Blog of Google, 2012.
- [13] R. Speer and C. Havasi, "Representing general relational knowledge in conceptnet 5," in *Proceedings of the LREC Conference*, pp. 3679–3686, Cambridge, UK, 2012.
- [14] S. Gallagher, "How google and microsoft taught search to understand the web," 2012, <http://arstechnica.com/information-technology/2012/06/inside-the-architecture-of-googles-knowledge-graph-and-microsofts-satori/>.
- [15] Li Jing, *Research on Construction Method and Application of Domain Ontology*, Chinese Academy of Agricultural Sciences, Beijing, China, 2009.
- [16] H. L. Nguyen, T. Dang, and J. J. Jung, "Knowledge graph fusion for smart systems: a Survey," *Information Fusion*, vol. 61, 2020.
- [17] Y. Liu, X. Cao, and T. Li, "Influence of accessibility on land use and landscape pattern based on mapping knowledge domains: review and implications," *Journal of Advanced Transportation*, vol. 2020, no. 6, 12 pages, Article ID 7985719, 2020.
- [18] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American*, vol. 284, no. 5, pp. 34–43, 2001.
- [19] N. Shadbolt, T. Berners-Lee, and W. Hall, "The semantic web revisited," *IEEE Intelligent Systems*, vol. 21, no. 3, pp. 96–101, 2006.

- [20] O. Etzioni, M. Banko, S. Soderland, and D. S. Weld, "Open information extraction from the web," *Communications of the ACM*, vol. 51, no. 12, pp. 68–74, 2008.
- [21] A. Carlson, J. Betteridge, and B. Kisiel, "Toward an architecture for never-ending language learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 24, no. 1, pp. 17–24, 2010.
- [22] M. Pasca, D. Lin, and J. Bigham, "Organizing and searching the world wide web of facts—step one: the one-million fact extraction challenge," *AAAI*, vol. 6, pp. 1400–1405, 2006.
- [23] A. Duque-Ramos, M. Boeker, L. Jansen, S. Schulz, M. Iniesta, and J. Tomás Fernández-Breis, "Evaluating the good ontology, design guideline (GoodOD) with the ontology quality requirements and evaluation method and metrics (OQuaRE)," *PLoS One*, vol. 9, pp. 1–14, Article ID e104463, 2014.
- [24] R. Gu, W. Hu, and Y. Huang, "Rainbow: a distributed and hierarchical RDF triple store with dynamic scalability," in *Proceedings of the 2014 IEEE International Conference on Big Data (Big Data)*, pp. 561–566, Washington, DC, USA, October 2014.