buildings

Article

Relation Extraction of Domain Knowledge Entities for Safety Risk Management in Metro Construction Projects

Na Xu [1], Hong Chang [2,*], Bai Xiao [3], Bo Zhang [4], Jie Li [5] and Tiantian Gu [1]

[1] School of Mechanics and Civil Engineering, China University of Mining and Technology, Xuzhou 221116, China
[2] Shenzhen Urban Public Safety and Technology Institute, Shenzhen 518000, China
[3] Tianjin Jingang Construction Co., Ltd., Tianjin 300456, China
[4] School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China
[5] School of Civil Engineering, Neijiang Normal Universities, Neijiang 641100, China
* Correspondence: changhong@szsti.org

Abstract: Gathering experience and organizing knowledge from a large number of engineering construction projects is conducive to more effective and efficient safety risk management in construction projects. Metro construction practitioners often find it difficult to determine what professional knowledge is needed to establish better management. By constructing the knowledge structure of safety risk management, which is composed of domain knowledge entities (DKEs) and their hierarchical relations, practitioners can systematically master the knowledge of safety management, enhance safety management levels, and reduce the occurrence of accidents. Traditionally, domain knowledge structure was determined by experts, the mistakes occur due to the limitations of individual knowledge, and high time costs are unavoidable due to the massive amount of data. Therefore, in this study, we used a rule-based Chinese-language natural language processing (C-NLP) method to automatically extract the hierarchical relations between DKEs from a large dataset of unstructured text documents; we aimed to clarify the affiliation relationship and parallel relationship between DKEs. First, 68,817 sources of literature written in Chinese were collected. Next, the specific syntactic structures of relations of the DKEs were analyzed. Hierarchical extraction rules, including 16 hyponymic indicators and 8 appositive indicators, were revealed based on the linguistic characteristics. Then, the relations were extracted from test dataset. The precision and recall values were used to verify the model. Finally, the hierarchical relations of all the DKEs were extracted, and the knowledge structure was formed. The proposed method of hierarchical relation extraction contributes to the quick automatic construction of knowledge structures and minimizes expert bias. The knowledge structures can be used to guide safety training and can assist practitioners in safety risk management.

Keywords: metro construction; safety risk management; relation extraction; natural language processing; rule-based

Citation: Xu, N.; Chang, H.; Xiao, B.; Zhang, B.; Li, J.; GU, T. Relation Extraction of Domain Knowledge Entities for Safety Risk Management in Metro Construction Projects. *Buildings* **2022**, *12*, 1633. https://doi.org/10.3390/buildings12101633

Academic Editors: Jun Wang, Shuyuan Xu, Yong Liu, Feng Yu

Received: 16 August 2022
Accepted: 28 September 2022
Published: 8 October 2022

1. Introduction

Construction is one of the most dangerous industries worldwide [1]. As a typical type of knowledge-intensive work, metro engineering has many risks that cannot be ignored in the construction stage due to the complex and unpredictable characteristics of the underground working environment, which leads to the occurrence of safety-related accidents [2]. Therefore, managers need to continuously expand and improve their domain knowledge structure, have systematic and comprehensive knowledge and awareness of safety risks, make reasonable risk-related decisions, and, eventually, improve the level of safety risk management.

Domain knowledge structures are composed of domain knowledge entities (DKEs) and the relation between them. A DKE, an elementary fragment of domain knowledge [3], is the most basic unit of the domain knowledge structure, which represents a thinking unit with complete logic. It can be a concept, a procedure, a feature, a regulation, or an axiom [4]. The DKEs of metro construction safety risk are composed of a collection of professional knowledge, experts' experience, and work skills in order to effectively complete the relevant management tasks and achieve the management objectives. There are various connections among DKEs, which influence and coordinate with each other to promote the final accomplishment of the project objectives. The types of relations between DKEs include but are not limited to causal relations, hierarchical relations, co-referential relations, subject-object relations, and part-whole relations. Among these, the hierarchical relations can be divided into affiliation relations and parallel relations. The affiliation relation represents a semantic relation between generic terms and specific terms. The generic term is called a hypernym and the specific term is called a hyponym [5]. The affiliation relation can be represented by the pattern of "X is a type/category of Y", where Y is the hypernym of X and correspondingly, X is the hyponym of Y, for example, "Object Strike" is a type of "Accident type". The parallel relation represents the relationship between words at the same level in a knowledge structure, for example, "Object Strike" and "Fall from height" are both types of "Accident type". A simple example of domain DKEs and their hierarchical relations is shown in Figure 1. In addition, the descriptions of specific domain terms are displayed in Appendix C. The clarification of hierarchical relationships is a significant basis for the storage and positioning of professional knowledge [6], which is an important step in building the domain knowledge structure. With the help of domain knowledge structure, project managers and construction workers can obtain more accurate and timely professional knowledge experience.

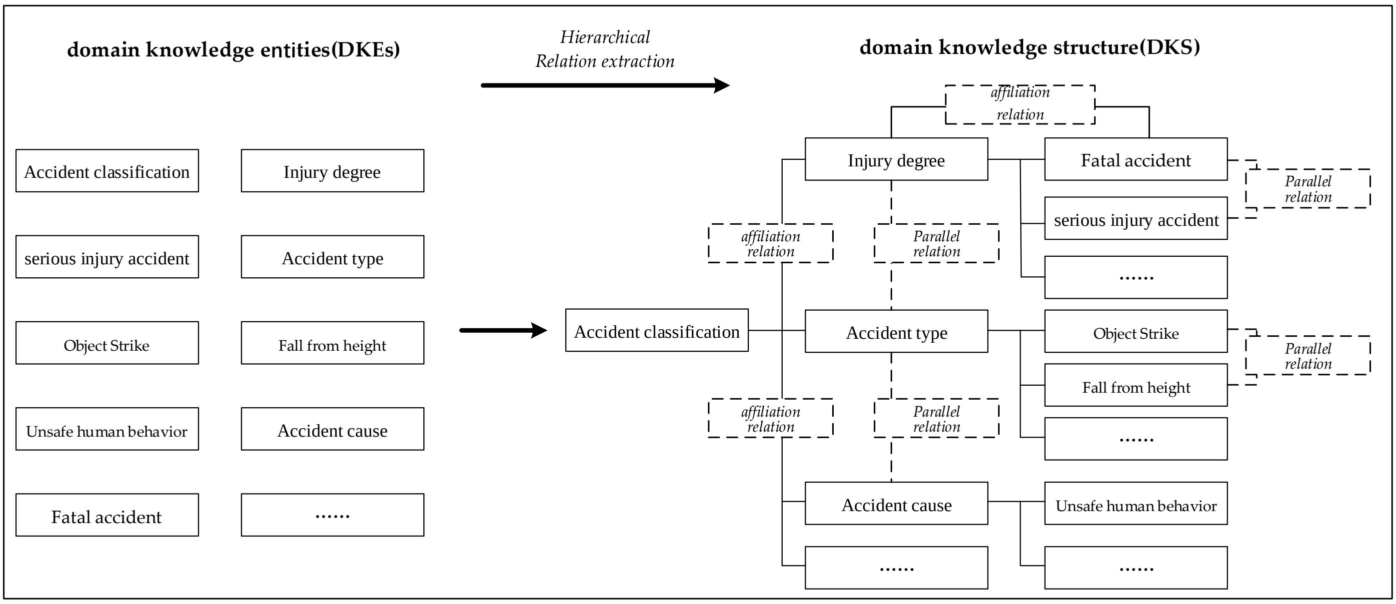

Figure 1. A simple example of domain DKEs and their hierarchical relations.

Traditionally, researchers have used methods to acquire and accumulate DKEs and their relations, such as expert interviews, questionnaire surveys [7], and case analyses [8]. Researchers have organized knowledge structures using techniques such as interpretation structure models [9], fault tree analyses, and event tree analyses [10]. The advantages of the traditional methods are as follows: (1) domain experts have rich experience, coupled with the skillful methods of questionnaire surveys and case analyses, which is conductive to the rapid and accurate identification of domain knowledge and their relations. (2) The knowledge structure constructed has a high degree of credibility because of its actual case-based engineering and clear analytical logic [11]. However, it is difficult to avoid the following deficiencies of the traditional manual methods: (1) human cognitive bias and subjectivity might affect the collection of domain knowledge, because experts with extensive safety experience cannot judge or predict all the safety conditions in complex and changeable construction sites [12]. (2) With the increasing number of cases, the cost of manpower and time increases geometrically. (3) It is difficult to integrate scattered knowledge into knowledge structure [13], traditional knowledge acquisition focuses on a single accident case or text document, and lacks horizontal analysis and integration among multiple similar cases, which makes this method insensitive to various relations and interactions between DKEs.

Natural language processing (NLP) has played a prominent role in the field of text mining, especially in knowledge mining and relation extraction. However, the Chinese-language text documents have the characteristics of a large vocabulary, fuzzy boundaries, flexible sentence patterns, and frequent omissions, resulting in few studies and applications of Chinese RE in metro construction. The study focused on safety risk management in Chinese metro construction using a text mining method to extract the hierarchical relations between DKEs. This method contributes to the automatic construction of knowledge structure.

The main contents are as follows:

1. The syntactic structure of the Chinese language was analyzed, which contains hierarchical relations (affiliation relations and parallel relations) in unstructured domain text documents;
2. The hierarchical relationship demonstrative words and syntactic rules were proposed;
3. The hierarchical relations of DKEs were automatically extracted in a big dataset of metro construction text documents.

The main contributions of this work are as follows:

1. Theoretically, this research provides a rule-based Chinese natural language processing (C-NLP) approach to automatic extract hierarchical relations from unstructured metro construction professional text documents. The proposed approach provides a technical support for the subsequent construction of domain knowledge structure and its expansion and innovation.
2. Practically, the clarification of the hierarchical relations between DKEs are beneficial to locate the professional knowledge and content for project managers and construction workers in safety risk management. The constructed domain knowledge structure can be used to consult the relevant knowledge, guide safety training, and construct domain knowledge graphs.

This paper is organized as follows. The current research status of knowledge-based safety in the construction industry and relation extraction in NLP are reviewed in Section 2. The method and model of domain knowledge hierarchical relation extraction are proposed in Section 3. In Section 4, the experiment is described step-by-step and the results are presented. The analysis of the results and the research limitations are discussed in Section 5. Finally, conclusions are drawn, informing future works.

2. Literature Review

2.1. Knowledge-Based Safety in the Construction Industry

Safety accidents occur frequently during the construction and operation of construction projects. Metro construction was taken as an example in this study; a total of 298 safety accidents occurred between 2001 and 2018, causing a large number of casualties and economic losses according to the statistics from the Ministry of Housing and Urban-Rural Development of China [14,15]. Through knowledge management and the accumulation of experience to reduce injuries, incidents, accidents, and illness rates, safety can be effectively improved, thereby increasing the efficiency, competitiveness, productivity, and profitability of enterprises [16–18]. However, knowledge management is frequently neglected in the establishment of an engineering safety culture [19]. It is an arduous task to determine the knowledge needed in engineering practice, because such knowledge is essentially based on experience, which is often intangible and elusive [20], and is sometimes forgotten as the project ends [21].

With the development of computer technologies and the increasing application of data-driven methods, researchers have used deep learning, NLP, and other methods to extract and share relevant knowledge from a large number of accident cases, in order to improve the performance of construction safety risk management. Bekhti proposed a risk knowledge management system to store professional knowledge and achieve knowledge sharing through the transmission of risk knowledge [22]. Kanapeckiene et al. developed an integrated model of knowledge management for the long-term accumulation and reuse of knowledge [23]. Ding et al. proposed a subway engineering safety risk identification system (SRIS) based on construction drawings for risk identification and risk assessment, so as to improve safety before construction [24]. Tixier et al. applied NLP technologies to extract meaningful structured attributes and data from unstructured building safety damage reports to improve safety management [25]. Su et al. established a case-based reasoning model to guide the pre-control and decision-making of safety accidents in the construction industry [26]. The researchers in the engineering field have striven to build an objective and comprehensive knowledge structure (system), drawing experience and knowledge from "historical lessons". Through knowledge management and knowledge accumulation, the decision-making guidance and risk pre-control of engineering projects can be used to improve the safety level of construction projects that are underway or planned to begin.

Domain knowledge structure refers to a knowledge system in which the hierarchical structure is formed by the knowledge entities and their interrelations [27]. In recent years, the authors have focused on knowledge mining and knowledge discovery in the domain of urban rail transit construction safety risk management. First, [28] developed a rule-based NLP approach for extracting DKEs and revealed the Chinese linguistic patterns and linguistic features from domain text documents. Then, the co-word co-occurrence network (CCN) and the association rule mining (ARM) was used to find the connected knowledge elements and expand domain knowledge elements (DKEs). Now, the determination of hierarchical relationships is an important object of this study. There are many difficulties in the whole process of domain knowledge structure construction: (1) the research has largely focused on the professional texts in the English language; there are few studies regarding the construction of domain knowledge structure based on Chinese materials or other non-English language texts. (2) The determination of the relationships requires a huge database. However, there are few publicly available ontology databases established from the perspective of design and engineering [29]. It is essential to determine the relationships between the domain DKEs from the Chinese corpus, either to establish and expand the knowledge structure or to recommend security precautions based on the knowledge structure.

2.2. Relation Extraction: Rule-Based Natural Language Processing

Natural language processing (NLP) is a field that uses artificial intelligence (AI) to enable computers to process natural language text in a manner similar to humans [30], which involves multiple fields including lexical, syntactic, semantic, and pragmatic analysis; text classification; sentiment analysis; automatic summarization; machine translation; and social computing [31].

Studies have shown that the relations between words are very important, both in domain model construction and in the application of NLP. NLP researchers have long had a common interest in building domain structures or semantic networks to characterize text structure and to find related terms [32–37]. Relation extraction (RE) between words is a sub-field of information extraction, whose purpose is to automatically extract the semantic relations between entities. In Chinese relation extraction, researchers have made a simple summary of relational demonstrative words (conjunctive phrases). Zheng proposed that parallel relations are manifested by the use of conjunctions between words such as “and”, “with”, “as”, and “as well as”, and commas are used as punctuation marks in general [38]. Tang put forward that “is a kind/category/a” is a typical pattern of an affiliation relation, in which the lower concept comes before “is a kind/category/a” and the upper concept comes afterwards [39]. RE tasks involve named entity recognition (determining DKEs), trigger word recognition (determining relation indicators), and relation extraction [40].

The relation extraction can be traced back to 2002 in construction engineering. Abuzir extracted terms and relations from HTML documents and constructed a thesaurus of civil engineering [41]. Clariana proposed an RE method that relies on a list of predefined domain concepts provided by experts [42], but he did not propose possible connectives. Al Qady identified conceptual relations and extracted semantic knowledge in construction contracts using NLP, the aim of which was to improve electronic document management (e.g., document classification and retrieval) [43]. After more than 20 years of conference development, the theories and methods of RE have become increasingly rich [44], such as those exemplified by the MUC (message understanding conference), ACE (automatic content extraction), TAC (text analysis conference) and SemEval (semantic evaluation).

Relation extraction methods can be divided into rule-based methods, machine learning-based methods, and the combination of the two methods. In the rule-based method, firstly, experts summarize the features of domain texts in data structure and grammatical structure, then they manually construct the corresponding grammatical or semantic rules, and finally extract target instances from the texts through automatic matching rules by a computer. In the machine learning method, various statistical algorithms (e.g., SVM and CRF) are used to transform the RE into classification problems, and a classification model based on feature learning is obtained. The relationships between the corresponding entities and entity types is established through the model [45]. Compared with machine learning-based extraction, rule-based approaches follow a mostly declarative pattern, leading to highly transparent and expressive models that generally achieve better precision [46].

3. Methodology

3.1. Data Collection and Method Selection

Metro construction knowledge is described in many Chinese text materials, such as news reports, website announcements, design documents, construction documents, meeting records, accident investigation reports, and the relevant literature. Compared with non-technical texts (e.g., news articles and website information), the domain literature, such as professional technical texts, is more suitable for NLP with better interpretability and less semantic ambiguity. The reasons are as follows: (1) there are fewer homonym conflicts. For example, in news articles, the term “bridge” may refer to a structural bridge, the card game, a bridge of communication, or a dental bridge. (2) There are fewer coreference resolution problems. For example, construction regulation texts tend to explicitly mention the subject (e.g., project manager) for each provision rather than referring to the subject using pronouns (e.g., “he”) [47]. (3) The literature is abundant, the content is objective, the language is concise and accurate, the discussion is more comprehensive, in-depth, and cutting edge. Therefore, the domain literature was selected as the original data for relation extraction.

For small- and medium-sized samples such as metro construction projects, rule-based methods have demonstrated more promising capabilities. The main reasons are as follows: (1) when a sufficient number of positive training examples cannot be provided, the performance might be poor and the accuracy might be affected such as in traditional machine learning [48,49]. Tixier chose to develop an NLP system based on manual coding rules to avoid these problems [25]. (2) Rules based on manual coding can achieve higher accuracy, because the researchers can transfer their expertise, data knowledge and human intelligence into the system [50]. (3) Rule-based methods avoid the relatively opaque characteristics of machine learning [51]. In summary, this study used a rule-based Chinese NLP method to extract the relations between DKEs in Chinese-language domain texts.

3.2. Hierarchical Relation Extraction Framework

In this study, the model for rule-based Chinese NLP hierarchical relation extraction was designed as shown in Figure 2.

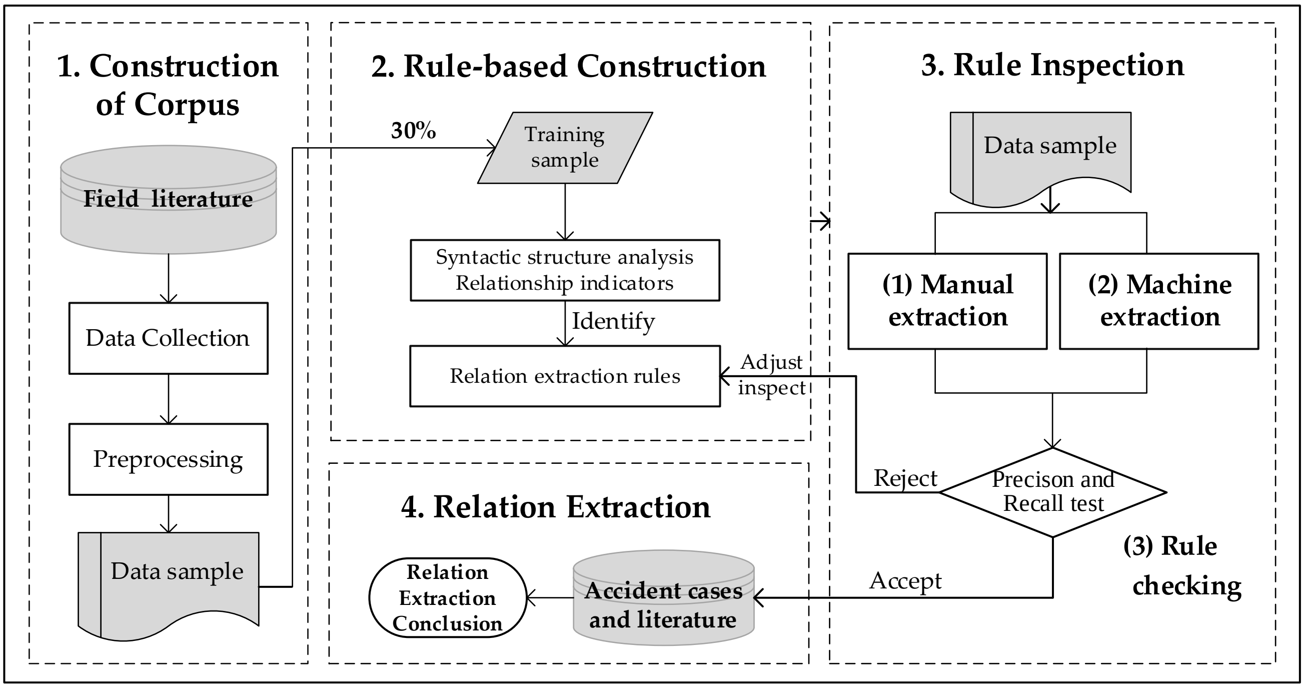

Figure 2. Rules-based Chinese NLP hierarchical relation extraction model.

Step 1. Construction of the corpus, including data collection and preprocessing. (1) Chinese texts were collected. (2) The texts were preprocessed into data samples arranged in sentence units. (3) A corpus of domain knowledge RE was formed.

Step 2. Rule-based construction: (1) A total of 30% of the sentences from the data sample were randomly selected at equidistant intervals, forming a training sample. (2) The dependency parsing of the training samples was analyzed to clarify the specific syntactic structure and relationship indicators. (3) The extraction rules of the hierarchical relations were determined.

Step 3. Rule inspection, including manual extraction and machine extraction. The two results of the extraction were compared and analyzed.

1. Manual extraction: The hierarchical relations between the DKEs were extracted and tested manually by two experts, including the affiliation and parallel relations. The two experts were a university professor who has rich theoretical knowledge and a project manager of construction enterprises who has more than ten years of practical experience in construction safety risk management.
2. Machine extraction: (1) Chinese NLP was used to analyze the dependency parsing of the training samples. The researchers recorded the linguistic features of the sentences. (2) Rules were constructed and hierarchical relations were extracted according to the linguistic features. The method path is expanded in Section 3.2.
3. Rule checking: The two results were compared, and the precision and recall were used to test the rules. (1) The inspection was qualified if the precision and recall met the requirements. The rules needed to be adjusted and improved if the values of

precision and recall were too low. (2) Steps 2 and 3 were cycled until the rules reached the acceptable range.

Step 4. Relation extraction: The rules were applied to the whole corpus for hierarchical relation extraction and the extraction results were taken.

3.3. Rule-Based Hierarchical Relation Extraction and Inspection

The text materials were written in natural Chinese language, and the form of the texts was composed of Chinese characters, phrases, sentences, paragraphs, and chapters. The domain knowledge entities (DKEs) and the relations between them were hidden in meaningful words and phrases. We first analyzed the language patterns of the corpus, and then formulated the rules of relation extraction according to the dependency parsing.

Chinese text language pattern analysis is realized by part-of-speech tagging (POS) and dependency parsing (DP) as follows: (1) POS: The sentence is divided into linguistic units (words), and the part of speech of the linguistic units is marked. (2) DP: The dependencies between the language units are analyzed to reveal the syntactic structure, including SBV (subject-verb relations), COO (coordinating relations), and others.

Taking the sentence A, "Rock lithology includes geological age, rock name, weathering degree, color, main minerals, structure, and rock quality" as an example, the POS and DP of this sentence are shown in Figure 3. For example, the word "color" is numbered 9, meaning that it is the ninth token in order and its POS tag is "noun" (n). The acronyms in the bottom line (COO, WP, etc.) indicate the syntactic dependencies of the linguistic units. In addition, the descriptions of the POS tagging and DP relationships are displayed in Appendixes A and B.

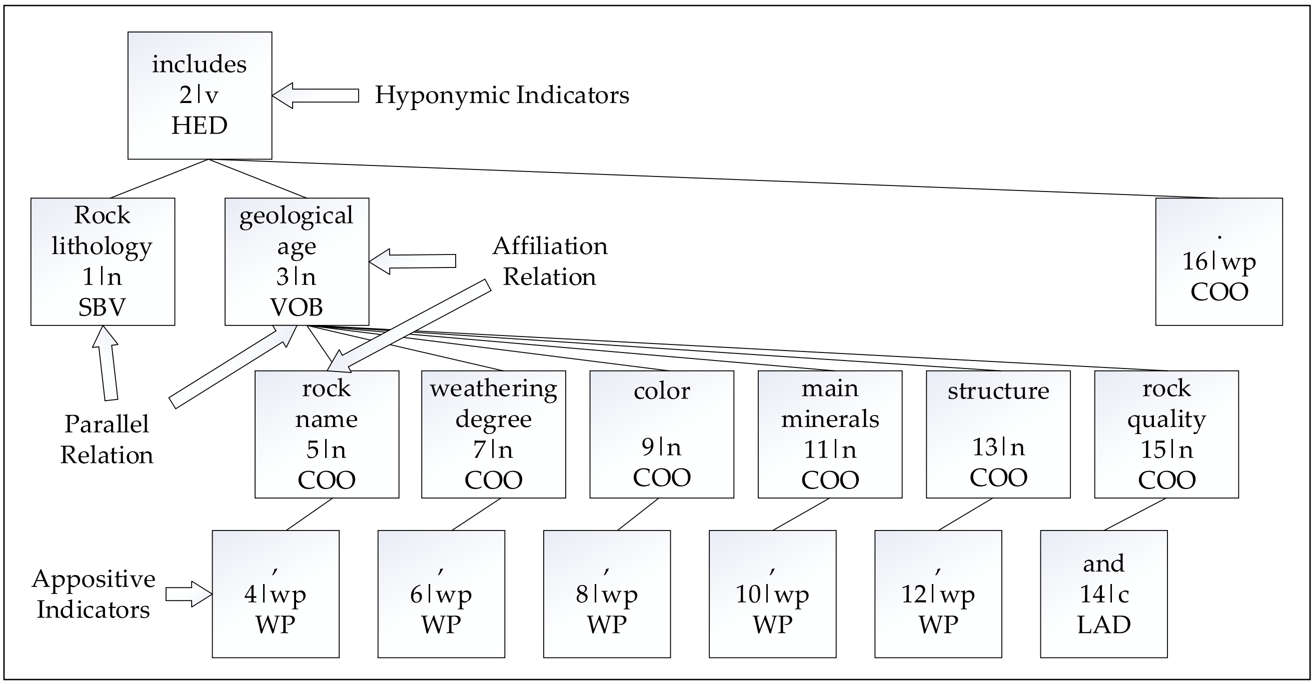

Figure 3. Example of Chinese-language text pattern analysis.

In sentence A, there are seven affiliation relations via the hyponymic demonstrative word "include" ("rock lithology" with "geological age", "rock name", etc.) and six parallel relations via the appositive demonstrative words "," with "and" ("geological age" and "rock name"). The dependency markers of the parallel relations are COO (coordinating relations). Chinese texts with hierarchical relations have commonalities in relational indicators and dependence relations; the relation extraction rules can be constructed according to the statistics of these common features. All the relation indicators and hierarchical extraction rules of DKEs can be clarified based on the NLP scheme of the LTP (language technology platform).

The rules need to be checked after being determined. Precision (P) and recall (R) are two metrics widely used in information retrieval and statistical classification to evaluate the quality of results. Precision measured the reliability of the hierarchical relations between DKEs, and recall (R) measured how many relations between DKEs were extracted from the test. The precision and recall were used to measure the two results (manual extraction and machine extraction), as shown in Formulas (1) and (2):

$$P = A/(A + B) \tag{1}$$

$$R = A/(A + C) \tag{2}$$

where A and B represent the correct and incorrect hierarchical relations extracted by the computer, respectively, and C represents the hierarchical relations identified by the experts but missed by the computer. The correct, incorrect, and missed relations were evaluated by manual extraction in Step 3 (1) (Figure 2).

4. Experiment and Results

4.1. Construction of the Corpus

The domain knowledge entities (DKEs) in metro construction, which were selected from the research results of our work [28,52], were used as keywords to search in the Chinese-language CNKI and Wanfang databases. The CNKI (China National Knowledge Infrastructure) database is the largest academic paper database and academic electronic resource integrator in China and contains more than 200 million papers, documents, and academic resources. The Wanfang database is a large network database developed by China Wanfang Data Corporation, covering journals, meeting minutes, papers, academic achievements, and academic conference papers. The abstracts of the Chinese-language papers were excerpted as the original texts. The original texts were divided into separate sentences by correcting spelling errors. A total of 550 sentences containing DKEs were randomly selected equidistantly by the computer as the sample data, and the corpus was constructed.

4.2. Rule-Based Construction

Esmaeili's research showed that it is reasonable to select 30% when manually analyzing text corpora and constructing rules [53]. Therefore, 165 sample sentences (30%) were randomly selected from the corpus equidistantly as training texts, which are shown in Table 1.

Table 1. Training texts with hierarchical relations.

Serial Number	Training Texts
1	Rock lithology including geological age, rock name, weathering degree, color, main minerals, structure, and rock quality.
…	…
78	It is necessary to strengthen the prevention and control work of landslides, collapses, mudslides, ground collapses, and ground subsidence.
…	…
165	The terrain (i.e., plain, hill, mountain, plateau, and basin) is also controlled in the model.

Through the manual statistics of 165 training texts, 523 hyponymic relations and 611 appositive relations were obtained.

In this study, the language technology platform (LTP) was used to analyze the selected 165 training texts, and all the hierarchical demonstrative words were counted in the process of dependency parsing, which is shown in Table 2. The LTP system is an open Chinese-language NLP system developed by the Harbin Institute of Technology. Compared to other NLP libraries, the LTP integrates the function of text segmentation, POS markup, and syntactic parsing. Its graph-based parsing method is beneficial to the visualization of syntactic structure features [54].

Table 2. Statistical table of hierarchical relation demonstrative words.

	Hierarchical Demonstrative Words			
Hyponymic relation demonstrative words (HRDW)	like	such as	that is	include
	divide into	can be divided into	contain	mainly consist of
	consist of	for example	mainly have	mainly refer to
	in turn	involve	()	:
Appositive relation demonstrative words (ARDW)	as well as	along with	and	as
	with	or	and others	、

The hierarchical relation extraction rules of the DKEs were determined by the dependency parsing results and the relational demonstrative words based on the LTP, as shown in Table 3.

Table 3. The hierarchical relation extraction rules.

Serial Number	**Hierarchical Relation**	**Specific Syntactic Structure**
Rule1	parallel relation	DKE1--COO--DKE2--ARDW
Rule2	parallel relation	DKE*--COO--DKE1--ARDW DKE*--COO--DKE2--ARDW
Rule3	affiliation relation	DKE1--HRDW--DKE2
Rule4	affiliation relation	DKE1—COO--DKE*--HRDW--DKE2

Parallel relation rules:

Rule 1: Two DKEs are directly connected by a coordinating relation (COO), and one of the DKEs is connected with the ARDW.

Rule 2: Two DKEs are subordinate to one DKE*, and are connected to the ARDW.

Affiliation relation rules:

Rule 3: Two DKEs are connected by HRDW.

Rule 4: A DKE is linked to a DKE* through a coordinating relation (COO), and the DKE* is connected to another DKE by HRDW.

4.3. Relation Extraction and Inspection

The hierarchical relation extraction rules need to be tested before machine extraction. The method was as follows: the relation extraction rules were reapplied to the randomly selected domain literature to compare the different results of the manual extraction and machine extraction. The effectiveness of the rules was tested using the precision rate (P) and the recall rate (R). The results are shown in Table 4.

Table 4. Rule inspection results.

Relation Extraction Type	Affiliation Relations	Parallel Relations	Hierarchical Relations
The number of manually extracted relations	523	611	1134
The correct relations according to the rule-based extraction (A)	348	603	951
The incorrect relations according to the rule-based extraction (B)	22	21	43
The relations identified by experts but missed by the rules (C)	175	8	183
The precision rate of extraction (P)	94.05%	96.63%	95.67%
The recall rate of extraction (R)	66.53%	98.69%	83.37%

The precision rate and recall rate of the hierarchical relation extracted by the rules were good, and the recall rate of the affiliation relations was slightly low, as shown in Table 4. The reasons were as follows: the specific syntactic structures summarized by the rules cannot cover all the sentences. Taking the sentence "The measurement of hydrogeological parameters mainly involves the measurement of groundwater level, groundwater permeability coefficient, and pour coefficient." as an example, the hypernymic DKE "hydrogeological parameters" and the hyponymic DKE "groundwater level" were not directly connected through "involves"; the following content was added: "the measurement of". The affiliation relations could not be extracted by the rules in the above case. In order to make up for the lower recall rate of affiliation, sentences can be screened out according to the DKEs in advance. An affiliation relation was determined if the sentence conformed to a specific syntactic structure, and the sentence was analyzed and judged manually if the syntactic structure was sparse. The rule-based method needs to be continuously optimized and enriched in the future.

The results showed that the precision rate of hierarchical relation extraction reached 95.67%, which indicates that the conjunctions, punctuations, syntactic structures, and dependencies showed prominent commonalities in Chinese-language professional knowledge texts. The 16 hyponymic relation demonstrative words and the 8 appositive relation demonstrative words summarized in the study were able to accurately reveal the hierarchical relations in Chinese-language professional texts. The higher precision rate might be caused by the limited data of the corpus, but this also proved the effectiveness and robustness of the rules and the specific relational demonstrative words in the process of relation extraction: the method of using rule-based NLP can effectively extract hierarchical relations. The characteristics of high precision in Chinese-language texts provide an effective guarantee for subsequent professional text mining and ontology construction.

4.4. Examples of the Results

Based on 550 pieces of data in the corpus, we extracted more than 1000 sets of affiliation relations and more than 2000 sets of parallel relations. Examples of the hierarchical relation extraction results are shown in Table 5.

Table 5. Examples of the hierarchical relation extraction results.

Sample Sentences	Specific Syntactic Structure	Satisfied Rules	Extraction Results	Literature Sources
Geological hazard prevention indicators [1] mainly include collapse [2], landslide [3], debris flow [4], ground collapse [5], ground subsidence [6], and ground fissure [7].	DKE [1]—HRDW—DKE [2]; DKE[2]—ARDW—DKE [3]--…--DKE [7];	Rule.1 Rule.3	Figure 3	Comprehensive assessment of major natural disasters
Soil parameters [1] mainly include soil types [2] (cohesive soil [8], non-cohesive soil [9]), relative density [3] or shear strength[4], soil internal friction angle [5], friction coefficient [6], soil specific gravity [7], etc.	DKE [1]—HRDW—DKE [2]; DKE [2]—ARDW—DKE [3]--…--DKE [7]; DKE [2]—HRDW—KE [8]—ARDW—DKE [9].	Rule.1 Rule.3 Rule.4	Figure 3	Application of structural support design in the treatment of submarine pipeline suspension span
The terrain of China is high in the west and low in the east; the western terrain [4] is dominated by mountains [1], plateaus [2], and basins [3], and the eastern terrain [4] is dominated by hills [5] and plains [6].	DKE [1]--ARDW—DKE [2]--ARDW—DKE [3]; DKE [1]—DKE [4]; DKE [5]--ARDW—DKE [6]—DKE [4];	Rule.1 Rule.2	Parallel Relation among DKE[1,2,3,5,6]	Reshaping China's Economic Geography in an Open Environment: Rediscovery of "First Nature" and Recreation of "Second Nature"

* The superscript refers to the serial number of domain knowledge entities.

The domain knowledge relation graph was generated based on the extraction result of the sentences, as shown in Figure 4. The hierarchical relations between the DKEs are clearly shown in the relational diagram. Project workers can accurately locate domain knowledge, improve the knowledge structure, and guide specific construction. This work also lays a foundation for the subsequent development of domain knowledge retrieval and discovery, the application of ontology, intelligent question answering, and other systems. In the future, it will be necessary to continuously integrate knowledge structure diagrams and form a comprehensive and systematic domain knowledge structures. In addition, we need to continue to explore the automatic learning and updating of domain knowledge structure based on ontology and unsupervised machine learning.

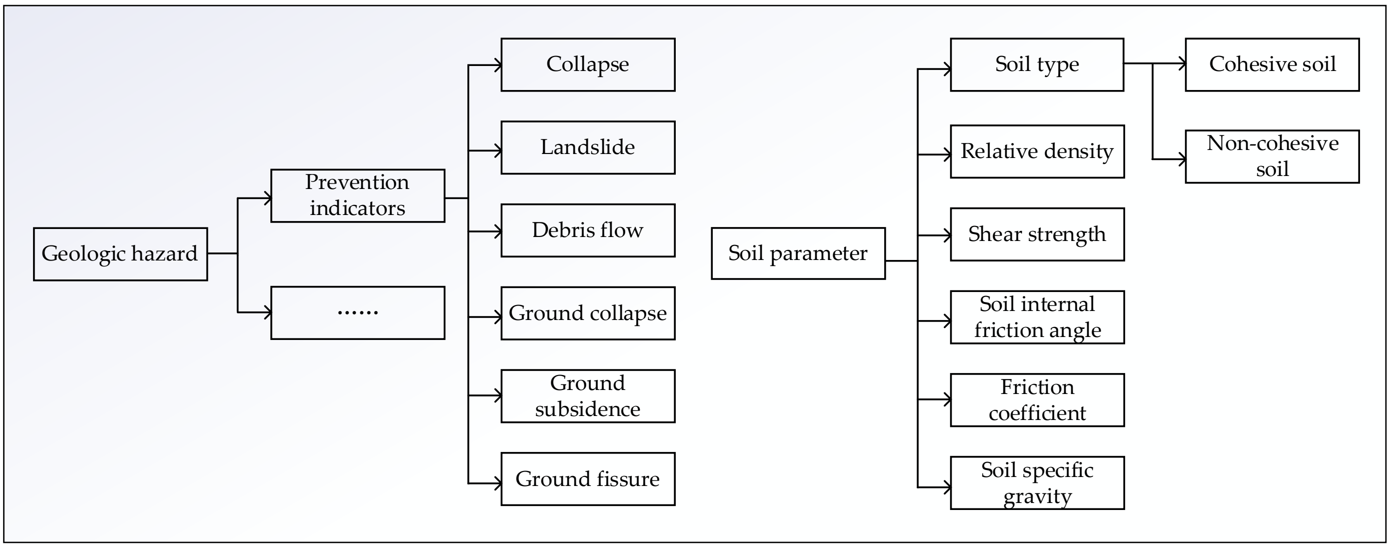

Figure 4. Schematic diagram of the knowledge structure generated by the sample sentences.

5. Discussion

The experiment showed that Chinese-language technical documents (e.g., accident investigation reports and the relevant literature) contain a large number of proper nouns and grammatical structures involving hierarchical relations; this has a potential advantage in the structuring of textual data. The rule-based Chinese NLP method can efficiently and accurately extract the hierarchical relations in domain technical literature. The extraction results are easily understood and applied. Compared with statistical-based methods such as machine learning, the proposed method can be applied to professional texts with a small training sample, effectively avoiding a large amount of text labeling work.

In addition, the analyzed and summarized relational demonstrative words, the constructed rules, and the proposed framework of hierarchical RE covered most of the features of the syntactic structures in the corpus. Compared with other rule-based (pattern-based) hierarchical relation extraction tasks [5], the method used in this study differed from the previous common-sense hierarchical relation extraction based on a large corpus. We focused on small-scale professional texts in the metro construction field, automatically extracted the hierarchical relations between DKEs, and a better precision rate and recall rate were obtained. A comparative analysis is shown in Table 6. In future research, we will continue to expand and optimize the rules on the basis of this research, and explore text mining research in other professional engineering fields within the construction industry.

Table 6. Comparative analysis of hierarchical relation extraction.

Author	Methods	Corpus	Result
Rydin [55]	A hierarchical structure consisting of hyponym-hypernym pairs was created using five different lexical patterns	293,692 Swedish daily news articles	1000 pairs in the generated hierarchical structure were selected with a 67.4–76.6% accuracy
Ando [56]	Seven hypernymy patterns	32 years newspapers of Japanese	130 target hypernyms with 49–87% precision
Snow [57]	Noun-noun type initial pairs were used to extract hypernymy dependency paths. These patterns were used to classify pairs	Corpus of 6 million words	Compared to Hearst's patterns [33], the F measurement score achieved a relative success rate of 132%
Yildiz [58]	Four lexico-syntactic patterns were used The corpus frequency-based and context word similarity-based eliminations methods were used to eliminate wrong pairs	Corpus of 500 million Turkish words	An average of 83% precision was achieved for four different target hypernym concepts
Sahin [59]	Nine different lexico-syntactic patterns were used The total pattern frequency, different pattern frequency, and word2vec vector similarity methods were used to evaluate correctness of extracted new pairs	Turkish news-based corpus of 500 million words	81–83% average precision was obtained for 15 target hypernym concepts
This paper	Rule-based Chinese NLP	Documents regarding metro construction	95.67% precision and 85.67% recall was achieved for hierarchical relation pairs

The clarification of the hierarchical relations can effectively connect the key theories, technologies, methods, materials, resource information, and other DKEs in the various stages of metro construction safety risk management. The extraction of hierarchical relationship connects knowledge elements, which is more conducive to the transfer and reuse of knowledge. Connecting knowledge elements is more conducive to the transfer and reuse. Throughout the whole process of metro construction, project managers can determine or make up for key nodes or construction techniques in the construction process that may be missed based on the domain knowledge from the early stage of construction, and can further consult the relevant standards and specifications to clarify the corresponding construction content and operations. On this basis, project managers can continue to supplement risk knowledge and risk response methods, and formulate a systematic and effective safety risk management system suitable for the project's characteristics. In addition, project managers can also formulate corresponding accident management tasks in advance based on specific accident types to improve the ability to prevent construction accidents, improve the level of emergency response after accidents, and effectively improve the level of safety risk management in metro construction.

Effective safety risk management requires a lot of theories, professional knowledge, and rich experience. The improvement of management ability requires knowledge-oriented training. However, under the pressure of time, many construction workers lack effective safety training [60]. For example, some workers have construction experience but lack safety knowledge related to specific operations. The determination of the hierarchical relations is convenient for them to accurately locate the target knowledge, fill their knowledge gap, improve safety awareness, and avoid the occurrence of safety-related accidents.

There are still some limitations in the study, involving the following:

1. Domain thesauruses for the construction industry are rare, which increases the difficulties of word segmentation, syntactic analysis, and relation extraction [44]. Due to the diversity of Chinese expressions and the complexities of the engineering field, the rules cannot cover all of the relevant linguistic phenomena [61]. It is difficult to extract relations from new patterns that are completely different from existing patterns [50]. With the increase in the size of the corpus, it is necessary to continuously expand and optimize the rules and improve the accuracy and robustness of relation extraction.
2. In this study, only public documents were extracted. Some potential and sudden safety-related accidents and construction problems were ignored and hidden. It is difficult to cover all the accidents and accident types in the corpus, which may lead to some differences between the domain knowledge structure and the knowledge needs of managers. In future research, it is necessary to continuously expand the knowledge corpus for subway construction and extensively extract the tacit knowledge of experienced experts, project managers and construction personnel.

6. Conclusions and Future Works

In the study, the rule-based C-NLP method was used to extract the hierarchical relations of DKEs in metro construction safety management. This study provides methods and solutions by which to reveal the hierarchical relations of unstructured professional texts. Our research provides knowledge support for the construction and improvement of domain knowledge structure, knowledge retrieval and discovery, the development of ontology systems and intelligent question answering. The main conclusions are as follows:

The hierarchical relations of the DKEs can be divided into affiliation relations and parallel relations. The construction of domain hierarchical relations strengthens the connection between DKEs, which helps project managers and construction workers to quickly and accurately locate knowledge blind spots and fill in the knowledge gaps. Based on the knowledge structure, project managers and construction workers can enrich safety management knowledge, improve decision-making capabilities, and improve the levels of knowledge-based safety risk construction management.

The specific syntactic structures of the hierarchical relation extraction were proposed. A total of 16 hyponymic demonstrative words and 8 appositive demonstrative words were revealed. For small-scale professional texts, the rule-based C-NLP technology proved to be suitable for knowledge mining and relation extraction. The results of relation extraction had high precision and recall rates. The relational demonstrative words, rules-constructed, and RE framework can be applied to text mining of other construction engineering fields.

For future research, it is necessary to continuously enrich and expand the rules to improve the coverage and accuracy of relation extraction. We should continuously expand DKEs and the relations among them, build a comprehensive and systematic knowledge structure for subway construction or the construction industry, explore the automatic expansion and improvement of domain knowledge structures by deep learning methods, and train a comprehensive model combining rules and statistics. It is necessary to "learn from history" to improve the safety risk management level of construction projects.

Author Contributions: The manuscript was compiled from the contributions of all authors. Conceptualization, N.X. and H.C.; methodology, N.X., H.C., and J.L.; software, B.Z.; validation, B.X. and T.G.; investigation, H.C.; resources, H.C. and J.L.; data curation, B.X.; writing—original draft preparation, H.C.; writing—review and editing, N.X.; visualization, T.G.; supervision, H.C.; project administration, N.X.; funding acquisition, N.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 71901206, and the Social Science Fund of Jiangsu Province, grant number 22GLB023.

Data Availability Statement: Some or all of the data, models, and code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: This research was supported by the National Natural Science Foundation of China (Grant No. 71901206) and the Social Science Fund of Jiangsu Province (Grant No. 22GLB023).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Descriptions of POS Tagging

The key symbols for the parts of speech (POS) and dependency parsing (DP) in Chinese-language NLP used in the paper are provided below. More descriptions of the POS and DP can be found at Language Technology Platform Cloud [62].

The following POS tags are used in this paper.

Tag	Description	Example
n	general noun	structure
v	verb	include
c	conjunction	and
wp	punctuation	,

Appendix B. Descriptions of DP

The following DP are used in this paper.

Tag	Description	Example
SBV	subject-verb relationship	Rock lithology includes geological age… ("Rock lithology" is the subject of the verb "includes".)
HED	head word	"includes" in "Rock lithology includes geological age…" (the verb is often the core of the whole sentence in Chinese.)

VOB	verb-object relationship	Rock lithology includes geological age… ("includes" is the verb governing the object "geological age.")
COO	coordinate relationship	structure and rock quality ("structure" and "rock quality" are coordinate related.)

Appendix C. A Glossary of Terms

The specific domain terms used in the article are as follows to help readers understand.

Terms	Description	Example
affiliation relation	a semantic relation between generic terms and specific terms.	"Object Strike" is a type of "Accident type"
hypernym	The generic term in affiliation relation	"Accident type" in the above example
hyponym	the specific term in affiliation relation	"Object Strike" in the above example
hyponymic indicators	hyponymic relation demonstrative words, which aims to find the affiliation relations between the DKEs	"Include", "contain", "divide into" …
parallel relation	the relationship between words at the same level in a knowledge structure	"Object Strike" and "Fall from height" (They are both types of "Accident type")
appositive indicators	Appositive relation demonstrative words, which aims to find the parallel relations between the DKEs	"as well as", "and", "or" …

References

1. Sacks, R.; Rozenfeld, O.; Rosenfeld, Y. Spatial and Temporal Exposure to Safety Hazards in Construction. *J. Constr. Eng. Manag.* **2009**, *135*, 726–736. https://doi.org/10.1061/(asce)0733-9364(2009)135:8(726).
2. Liu, H.; Xie, Y.; Liu, Y.; Nie, R.; Li, X. Mapping the Knowledge Structure and Research Evolution of Urban Rail Transit Safety Studies. *IEEE Access* **2019**, *7*, 186437–186455. https://doi.org/10.1109/access.2019.2961434.
3. Durlach, P.; Lesgold, A. (Eds.). *Adaptive Technologies for Training and Education*; Cambridge University Press: Cambridge, UK, 2012. https://doi.org/10.1017/cbo9781139049580.
4. Gao, G.W.; Wang, Y.G.; Li, Y.X. Review of the Research of Domestic Knowledge Element. *Inf. Sci.* **2016**, *34*, 161–165. https://doi.org/10.13833/j.cnki.is.2016.02.032.
5. Sahin, G. Extraction of Hyponymy, Meronymy and Antonymy Relation Pairs: A Brief Survey. *Int. J. Nat. Lang. Comput.* **2017**, *6*, 1–18. https://doi.org/10.5121/ijnlc.2017.6201.
6. Wang, Y. The Application of Natural Language Processing Technology in Building POE. *South Archit.* **2019**, *01*, 82–87. https://doi.org/10.3969/j.issn.1000-0232.2019.01.082.
7. Sun, Y.; Fang, D.; Wang, S.; Dai, M.; Lv, X. Safety Risk Identification and Assessment for Beijing Olympic Venues Construction. *J. Manag. Eng.* **2008**, *24*, 40–47. https://doi.org/10.1061/(asce)0742-597x(2008)24:1(40).
8. Chi, S.; Han, S.; Kim, D.Y.; Shin, Y. Accident risk identification and its impact analyses for strategic construction safety management. *J. Civ. Eng. Manag.* **2015**, *21*, 524–538. https://doi.org/10.3846/13923730.2014.890662.
9. Liu, P.; Li, Q.; Bian, J.; Song, L.; Xiahou, X. Using Interpretative Structural Modeling to Identify Critical Success Factors for Safety Management in Subway Construction: A China Study. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1359. https://doi.org/10.3390/ijerph15071359.
10. Šejnoha, J.; Jarušková, D.; Špačková, O.; Novotná, E. Risk Quantification for Tunnel Excavation Process. *World Acad. Sci. Eng. Technol. Int. J. Mech. Aerosp. Ind. Mechatron. Manuf. Eng.* **2009**, *3*, 1200–1208.
11. Jiang, X.; Wang, S.; Wang, J.; Lyu, S.; Skitmore, M. A Decision Method for Construction Safety Risk Management Based on Ontology and Improved CBR: Example of a Subway Project. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3928. https://doi.org/10.3390/ijerph17113928.
12. Tixier, A.J.-P.; Hallowell, M.R.; Rajagopalan, B.; Bowman, D. Application of machine learning to construction injury prediction. *Autom. Constr.* **2016**, *69*, 102–114. https://doi.org/10.1016/j.autcon.2016.05.016.

13. Tao, P.; Hou, Y. Integration and Reconstruction of Engineering Knowledge. *China Metall. Educ.* **2019**, *03*, 32–33+36. https://doi.org/10.16312/j.cnki.cn11-3775/g4.2019.03.012.
14. Zhao, X.; Gu, B.N. Statistical Analysis of Urban Rail Transit Lines in 2018 China. *Urban Mass Transit* **2019**, *22*, 1–7. https://doi.org/10.16037/j.1007-869x.2019.01.001.
15. Yue, Y.; Xiahou, X.; Li, Q. Critical Factors of Promoting Design for Safety in China's Subway Engineering Industry. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3373. https://doi.org/10.3390/ijerph17103373.
16. Chan, A.P.C.; Wong, F.K.W.; Chan, D.W.M.; Yam, M.C.H.; Kwok, A.W.K.; Lam, W.M.; Cheung, E. Work at Height Fatalities in the Repair, Maintenance, Alteration, and Addition Works. *J. Constr. Eng. Manag.* **2008**, *134*, 527–535. https://doi.org/10.1061/(asce)0733-9364(2008)134:7(527).
17. Hon, C.K.H.; Chan, A.P.C.; Yam, M.C.H. Determining Safety Climate Factors in the Repair, Maintenance, Minor Alteration, and Addition Sector of Hong Kong. *J. Constr. Eng. Manag.* **2013**, *139*, 519–528. https://doi.org/10.1061/(asce)co.1943-7862.0000588.
18. Hon, C.K.; Chan, A.P.; Yam, M.C. Relationships between safety climate and safety performance of building repair, maintenance, minor alteration, and addition (RMAA) works. *Saf. Sci.* **2014**, *65*, 10–19. https://doi.org/10.1016/j.ssci.2013.12.012.
19. Deepak, M.D.; Mahesh, G. Developing a knowledge-based safety culture instrument for construction industry: Reliability and validity assessment in Indian context. *Eng. Constr. Arch. Manag.* **2019**, *26*, 2597–2613. https://doi.org/10.1108/ecam-09-2018-0383.
20. Kivrak, S.; Arslan, G.; Dikmen, I.; Birgonul, M.T. Capturing Knowledge in Construction Projects: Knowledge Platform for Contractors. *J. Manag. Eng.* **2008**, *24*, 87–95. https://doi.org/10.1061/(asce)0742-597x(2008)24:2(87).
21. Esmi, R.; Ennals, R. Knowledge management in construction companies in the UK. *AI Soc.* **2009**, *24*, 197–203. https://doi.org/10.1007/s00146-009-0202-9.
22. Bekhti, S.; Matta, N.; Djaiz, C. Knowledge representation for an efficient re-use of project memory. *Appl. Comput. Inform.* **2011**, *9*, 119–135. https://doi.org/10.1016/j.aci.2011.05.004.
23. Kanapeckiene, L.; Kaklauskas, A.; Zavadskas, E.; Seniut, M. Integrated knowledge management model and system for construction projects. *Eng. Appl. Artif. Intell.* **2010**, *23*, 1200–1215. https://doi.org/10.1016/j.engappai.2010.01.030.
24. Ding, L.; Yu, H.; Li, H.; Zhou, C.; Wu, X.; Yu, M. Safety risk identification system for metro construction on the basis of construction drawings. *Autom. Constr.* **2012**, *27*, 120–137. https://doi.org/10.1016/j.autcon.2012.05.010.
25. Tixier, A.J.-P.; Hallowell, M.R.; Rajagopalan, B.; Bowman, D. Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports. *Autom. Constr.* **2016**, *62*, 45–56. https://doi.org/10.1016/j.autcon.2015.11.001.
26. Su, Y.; Yang, S.; Liu, K.; Hua, K.; Yao, Q. Developing A Case-Based Reasoning Model for Safety Accident Pre-Control and Decision Making in the Construction Industry. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1511. https://doi.org/10.3390/ijerph16091511.
27. Liu, P.; Wu, Q. Detecting Disciplinary Knowledge Structure Based on Formal Concept Analysis: An Empirical Investigation on Library and Information Science. *Libr. Inf. Serv.* **2014**, *58*, 50–65. https://doi.org/10.13266/j.issn.0252-3116.2014.18.009.
28. Xu, N.; Ma, L.; Wang, L.; Deng, Y.; Ni, G. Extracting Domain Knowledge Elements of Construction Safety Management: Rule-Based Approach Using Chinese Natural Language Processing. *J. Manag. Eng.* **2021**, *37*, 04021001. https://doi.org/10.1061/(asce)me.1943-5479.0000870.
29. Shi, F.; Chen, L.; Han, J.; Childs, P. A Data-Driven Text Mining and Semantic Network Analysis for Design Information Retrieval. *J. Mech. Des.* **2017**, *139*, 111402. https://doi.org/10.1115/1.4037649.
30. Cherpas, C. Natural language processing, pragmatics, and verbal behavior. *Anal. Verbal Behav.* **1992**, *10*, 135–147. https://doi.org/10.1007/bf03392880.
31. Zhao, J.S.; Song, M.X.; Gao, X. Development and Application of Natural Language Processing. *Inf. Technol. Informatiz.* **2019**, *07*, 142–145.
32. Lancashire, I. Review of Aspects of text structure: An investigation of the lexical organization of text by Martin Phillips. North-Holland 1985. *Comput. Linguist.* **1987**, *13*, 347–350.
33. Hearst, M.A. Automatic acquisition of hyponyms from large text. *COLING-1992*, **1992**, pp. 539–545.
34. Pantel, P.; Lin, D. Discovering word senses from text. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Canada, 23–26 July 2002; pp. 613–619.
35. Widdows, D.; Dorow, B. A graph model for unsupervised lexical acquisition. In Proceedings of the 19th International Conference on Computational Linguistics, Taipei, China, 24 August 2002; pp. 1093–1099.
36. Kozareva, Z.; Hovy, E. A Semi-Supervised Method to Learn and Construct Taxonomies Using the Web. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, MIT, Cambridge, Massachusetts, USA, 9 October 2010; pp. 1110–1118.
37. Clark, M.; Kim, Y.; Kruschwitz, U.; Song, D.; Albakour, D.; Dignum, S.; Beresi, U.C.; Fasli, M.; De Roeck, A. Automatically structuring domain knowledge from text: An overview of current research. *Inf. Process. Manag.* **2012**, *48*, 552–568. https://doi.org/10.1016/j.ipm.2011.07.002.
38. Zheng, L.; Lü, X.; Liu, K.; Lin, J. Automatic Identification of Chinese Coordination Relations. *Acta Sci. Nat. Univ. Pekin.* **2013**, *49*, 20–24. https://doi.org/10.13209/j.0479-8023.2013.004.
39. Tang, Q.; Lü, X.; Li, Z. Research on Domain Ontology Concept Hyponymy Relation Extraction. *Microelectron. Comput.* **2014**, *31*, 68–71. https://doi.org/10.19304/j.cnki.issn1000-7180.2014.06.017.

40. Zhou, D.; Zhong, D.; He, Y. Biomedical Relation Extraction: From Binary to Complex. *Comput. Math. Methods Med.* **2014**, *2014*, 298473. https://doi.org/10.1155/2014/298473.
41. Abuzir, Y.; Abuzir, M.D.O. Constructing the Civil Engineering Thesaurus (CET) Using ThesWB. In Proceedings of the Computing in Civil Engineering, Washington, DC, USA, 2–3 November 2002; pp. 400–412.
42. Cañas, A.J.; Novak, J.D.; González, F.M.; Clariana, R.B.; Koul, R. A Computer-based Approach for Translating Text into Concept Map-like Representations. In Proceedings of the First International Conference on Concept Mapping, 2004, Pamplona, Spain, 14–17 September 2004; pp. 131–134.
43. Al Qady, M.; Kandil, A. Concept Relation Extraction from Construction Documents Using Natural Language Processing. *J. Constr. Eng. Manag.* **2010**, *136*, 294–302. https://doi.org/10.1061/(asce)co.1943-7862.0000131.
44. Li, D.M.; Zhang, Y.; Li, D.Y.; Lin, D.Q. Review of Entity Relation Extraction Methods. *J. Comput. Res. Dev.* **2020**, *57*, 1424–1448.
45. Wang, L. Research on Intelligent Knowledge Support for Urban Rail Transit Construction Safety Management. Ph.D. Dissertation, China University of Mining and Technology, Xuzhou, China, 2019.
46. Waltl, B.; Bonczek, G.; Matthes, F. Rule-Based Information Extraction: Advantages, Limitations, and Perspectives, **2018**. Available online: https://wwwmatthes.in.tum.de/pages/1w12fy78ghug5 (accessed on 3 April 2018).
47. Zhang, J.; El-Gohary, N.M. Semantic NLP-Based Information Extraction from Construction Regulatory Documents for Automated Compliance Checking. *J. Comput. Civ. Eng.* **2016**, *30*, 04015014. https://doi.org/10.1061/(asce)cp.1943-5487.0000346.
48. Prabowo, R.; Thelwall, M. Sentiment analysis: A combined approach. *J. Inf.* **2009**, *3*, 143–157. https://doi.org/10.1016/j.joi.2009.01.003.
49. Lee, J.; Yi, J.-S.; Son, J. Development of Automatic-Extraction Model of Poisonous Clauses in International Construction Contracts Using Rule-Based NLP. *J. Comput. Civ. Eng.* **2019**, *33*, 04019003. https://doi.org/10.1061/(asce)cp.1943-5487.0000807.
50. Sagae, K.; Lavie, A. Combining rule-based and data-driven techniques for grammatical relation extraction in spoken language. In Proceedings of the Eighth International Conference on Parsing Technologies, 2003, Nancy, France, 23–25 April 2003.
51. Barbella, D.; Benzaid, S.; Christensen, J.M.; Jackson, B.; Qin, X.V.; Musicant, D.R. Understanding Support Vector Machine Classifications via a Recommender System-Like Approach. In Proceedings of DMIN, 2009, Las Vegas, Nevada, USA, 13–16 July 2009; pp. 305–311.
52. Xu, N.; Ma, L.; Liu, Q.; Wang, L.; Deng, Y. An improved text mining approach to extract safety risk factors from construction accident reports. *Saf. Sci.* **2021**, *138*, 105216. https://doi.org/10.1016/j.ssci.2021.105216.
53. Esmaeili, B.; Hallowell, M.R.; Rajagopalan, B. Attribute-Based Safety Risk Assessment. I: Analysis at the Fundamental Level. *J. Constr. Eng. Manag.* **2015**, *141*, 04015021. https://doi.org/10.1061/(asce)co.1943-7862.0000980.
54. Sun, S.; Luo, C.; Chen, J. A review of natural language processing techniques for opinion mining systems. *Inf. Fusion* **2017**, *36*, 10–25. https://doi.org/10.1016/j.inffus.2016.10.004.
55. Rydin, S. Building a Hyponymy Lexicon with Hierarchical Structure. In Proceedings of the ACL-02 Workshop on Unsupervised Lexical Acquisition, Philadelphia, Pennsylvania, USA, 12 July 2002; pp. 26–33.
56. Ando, M.; Sekine, S.; Ishizaki, S. Automatic extraction of hyponyms from Japanese newspapers using lexico-syntactic patterns. In Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, Lisbon, Portugal, 26–28 May 2004.
57. Snow, R.; Jurafsky, D.; Ng, A.Y. Learning syntactic patterns for automatic hypernym discovery. In Proceedings of the 17th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 1 December 2004; pp. 1297–1304.
58. Yıldırım, S.; Yıldız, T. Corpus-Driven Hyponym Acquisition for Turkish Language. In Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics, New Delhi, India, 11–17 March 2012.
59. Sahin, G.; Diri, B.; Yildiz, T. Extraction of Turkish semantic relation pairs using corpus analysis tool. *Int. J. Comput. Inf. Technol.* **2016**, *05*, 491–499.
60. Pandey, S. Current status for safety knowledge and training for workers involved in tunnel construction: A case study. In Proceedings of the 1st KEC Conference, Lalitpur, Nepal, 27 September 2018; pp. 103–107.
61. Tan, Y.; Liu, S.; Lv, X. CNN and BiLSTM Based Chinese Textual Entailment Recognition. *J. Chin. Inf. Process.* **2018**, *32*, 11–19.
62. Language Technology Platform. Available online: https://www.ltp-cloud.com/intro (accessed on 1 June 2021).