

Article

A Semi-Automatic Ontology Development Framework for Knowledge Transformation of Construction Safety Requirements

Zhijiang Wu ^{1,*} , Mengyao Liu ^{2,*} and Guofeng Ma ³¹ College of Architectural Science and Engineering, Yangzhou University, Yangzhou 225009, China² School of Civil Engineering and Architecture, East China Jiaotong University, Nanchang 330013, China³ Department of Construction Management and Real Estate, School of Economics and Management, Tongji University, Shanghai 200092, China; guofengma@tongji.edu.cn

* Correspondence: zjwu@yzu.edu.cn (Z.W.); mengyaoliujd@126.com (M.L.)

Abstract: Construction safety requirements (SRs), which serve as critical information encapsulating a wide range of safety-related issues, constitute a fundamental basis for effective construction safety management. The constraints of the complex information characteristics and uncertainty of knowledge migration, however, lead to the failure to transform most of the requirement information into effective knowledge. This study proposes a multi-stage knowledge transformation framework for realizing the transformation of SRs from abstract information to canonical knowledge, and it accurately completes the knowledge transformation through document matching, knowledge extraction, and knowledge representation. Meanwhile, a semi-automated model was introduced into this study to develop a domain ontology knowledge base for SRs and to represent each type of knowledge through class definitions. The proposed framework was validated by testing project documents collected from two types of building projects, and the results show that the RD-based association rules can accurately match documents associated with SRs and adapt to match different types of sentiment attribute documents. Moreover, the improved TF-IDF algorithm improved by 20% in precision and recall, showing that the algorithm can extract tacit knowledge by combining knowledge points. Further, the domain ontology knowledge base facilitates normative documentation and representation for each type of knowledge in SRs.



Academic Editors: David J. Edwards and Osama Abudayyeh

Received: 19 November 2024

Revised: 24 December 2024

Accepted: 21 January 2025

Published: 13 February 2025

Citation: Wu, Z.; Liu, M.; Ma, G. A Semi-Automatic Ontology Development Framework for Knowledge Transformation of Construction Safety Requirements. *Buildings* **2025**, *15*, 569. <https://doi.org/10.3390/buildings15040569>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: construction safety requirements; knowledge transformation; ontology

1. Introduction

The extensive requirement information generated from construction safety management is highly applicable and can support managers in decision making. This information is part of the SRs and records the details of entities, actions, and objects [1]. Moreover, managers can discover the tacit knowledge associated with SRs through knowledge transfer and thus assess the safety status of the site [2,3]. In this context, transforming the various types of requirements information into knowledge that can be directly applied to site safety management is crucial for enhancing the performance of safety management.

SRs have complex information characteristics, which contain three types of elements of requirement information, and there is also strict correlation logic between the elements [4–6]. Considering the constraints of the information characteristics, it is necessary to analyze the content and element relationships before acquiring the core knowledge communicated through this information [7]. According to the impact mechanism of SRs, there is a series of consequences that may be triggered when the requirements are violated [1,8].

Specifically, the unfulfilled requirements may induce potential safety risks and lead to corresponding safety incidents [9]. Accordingly, the subsequent information emerges when the requirements are not met. Thus, this information can be categorized as tacit knowledge and is equally valuable.

Previous studies have attempted to retrieve SRs in conjunction with content structure and categorize them according to content attributes [2,4,10]. Considering the information characteristics of SRs, some scholars have refined the information content based on entity elements [11,12]. To be able to transform redundant information into valuable knowledge, structured representations have also been used for knowledge transformation of SRs [13]. Meanwhile, ontology modeling also benefits from the ability to specify the classes and relationships commonly used to manage the information of SRs [14,15]. Some scholars have decoded the information of SRs to build a knowledge structure with knowledge elements, classes, and subclasses as the main structure [1,16]. Further, ontology knowledge has also been used for safety incident profiling to extract knowledge from SRs [16–18]. Although there have been some attempts in existing studies to transform the knowledge of SRs, they failed to extract the complete knowledge of SRs (especially tacit knowledge) and lacked normative documentation and representation for each type of knowledge. In this context, the value of SRs is inadequately demonstrated in safety-related decisions, thereby diminishing their role in guiding management practices on construction sites.

To address this limitation, this study aims to accurately capture the complete knowledge embedded in SRs and develop a domain ontology knowledge base in conjunction with a semi-automated model for normative documentation and representation of each type of knowledge. Specifically, this paper proposes a multi-stage knowledge transformation framework for transforming SRs from redundant information to normative knowledge. The framework incorporates the dichotomous influence mechanism of SRs and analyses the corresponding explicit and tacit knowledge. Meanwhile, this study also condenses the knowledge transformation process of SRs and divides it into three stages: document matching, knowledge extraction, and knowledge representation. Further, a semi-automated model is introduced into the framework to develop a domain ontology knowledge base to manage the transformed knowledge in a standardized manner.

The content of this paper is organized as follows: In the next section, we review the research background of SRs in the areas of knowledge extraction, knowledge transfer, and knowledge representation. Next, a multi-stage knowledge transformation framework is constructed and combined with a semi-automated model to develop a domain ontology knowledge base. The testing and knowledge transformation of the framework are recorded in the result analysis of this study, and detailed knowledge extraction and transfer are provided. Finally, the discussion and conclusions of this study are also elaborated in the subsequent sections.

2. Literature Review

2.1. Knowledge Extract of SRs

Many scholars have executed a series of works on the information mining, analysis, as well as processing of SRs [4,19,20]. Some scholars have proposed a hybrid granularity specification knowledge base-building method, which uses ontology modeling software to build a construction safety specification knowledge base and presents the requirement information in the form of accident portraits [16,21]. By evaluating the safety-related automated system of a nuclear power plant, the critical matter analysis method has also been introduced into the study of extracting requirements-related knowledge [22]. After acquiring the requirements knowledge, the specification of the description of this knowledge has also received attention. For instance, some scholars have used domain knowledge to

describe the requirement information and developed a knowledge structure of requirement information with the main structure of the knowledge elements, subjects, and subcategories of subjects [23]. Further, ontology has been introduced to describe the knowledge structure of SRs and to automate the identification of these requirements data in Revit [24].

In addition, the integration of multiple techniques has been used to uncover and manage knowledge derived from SRs. For example, ontology modeling and document modeling have been integrated to semi-automatically identify SRs from construction safety standards and obtain knowledge through information representation [4,8,25]. Moreover, the combination of natural language processing (NLP) and text mining (TM) has been used to extract SRs from project documents and obtain effective knowledge through structured representation [1,2,8]. Considering the uncertain safety risks brought by the COVID-19 pandemic to infrastructure construction, some scholars have integrated knowledge management and BIM for safety risk identification and explicitly presented the requirement information under the corresponding risks through the extraction of topological relationships and visualization [21,26].

These studies show that scholars have transitioned from focusing on information retrieval of SRs to knowledge acquisition, and the modes of acquisition are thus gradually being enriched. Moreover, previous studies have also confirmed the rich value of the multifaceted knowledge embedded in SRs that can contribute to safety strategies in project management.

2.2. Knowledge Transfer for SRs

Due to the rich information content of SRs, some scholars have begun to pay attention to the value transfer of requirement information in knowledge management. For instance, some scholars combine the “2–4” phase model of behavioral safety to describe unsafe behaviors and violations at the individual and organizational levels as safety knowledge, applying it to safety corrective actions, safety standard determination, and safety project inspection [27]. Accordingly, some requirement information can directly intervene in safety decision making, including overtime information that can be used to send physical fitness warnings through wearable devices [7,28], danger alerts for large heart rate fluctuations [29], and siren alerts for risky behaviors [30,31].

In addition, SRs are used as a criterion for safety management implementation, and potential risks in the current safety management model are assessed through requirements consistency checks [7,31]. In particular, SRs were also used as safety target settings for similar projects, and quantitative values from this information were extracted as key indicators [32]. Also, previous studies have started to use normalization to extract knowledge from SR and to assess the value of the knowledge. To extract knowledge comprehensively, a new approach with modeling and reasoning capabilities has been used by scholars for the semi-automatic identification of SRs from building safety standards and knowledge transformation of safety-related information [1,4]. A system capable of providing safety information following acts and regulations was developed to extract knowledge from SRs and judge the value of the knowledge against defined norms [33]. Since knowledge assessment of SRs is labor-intensive, automated extraction incorporating information technology has begun to gain attention. For instance, information extraction methods embedded with deep learning can be used to automatically extract named entities describing fall protection requirements from building safety regulations and to resolve referential ambiguity [2,8]. Further, the combination of knowledge mining with machine learning and reinforcement learning can improve SR knowledge extraction accuracy and canonical categorization of knowledge [9,12,34].

These studies show that SRs both realize the value extraction of single information and acquire the knowledge associated with the requirement information through knowledge migration. However, previous studies have failed to achieve a canonical representation of multivariate knowledge and have lacked the extraction of tacit knowledge.

2.3. Knowledge Representation for SRs

The canonical representation of the knowledge to which SRs belong is a prerequisite for their ability to be of value in project management and has also gained attention in previous studies. Traditional studies tend to break down the information in SRs in terms of semantic structure and divide the information structure according to lexical properties and word categories [35,36]. To address the multiple types of emotions involved in the information in SRs, knowledge representation methods that incorporate semantic emotions are also used for the processing of fragmented safety knowledge [7,37]. Moreover, knowledge graphs, as a commonly used knowledge specification representation vector, are also used by some scholars to represent the knowledge of SRs. For instance, the relationship between hypergraphs and knowledge graphs is established to improve the ability of knowledge graphs to represent complex safety knowledge as well as knowledge aggregation [38]. To mine the knowledge of multimodal information in SRs, previous studies have also defined the energy function of ternary groups as textual features with correlation relationships and multimodal features of entities to obtain semantically explicit knowledge [2]. Also, methods such as metatheory, rule logic, and domain ontology have been gradually applied to knowledge graphs to develop safety knowledge representation systems that are both personalized and functional [16,39,40].

Compared to the knowledge representation method that employs manual decomposition and processing of information semantics, intelligent representation models that incorporate new information technologies have been gradually applied to the research. Combining knowledge representation with deep learning enables the extraction of SRs from unstructured text and enables an automated description of the relationship between safety incidents and SRs [2,38]. Considering the uncertainty of the construction site, previous research has automatically extracted information from texts and tables based on safety scenarios and combined knowledge graphs with deep learning to achieve a canonical representation of safety-related knowledge [23,41,42].

Previous studies have also represented the knowledge of SRs in the context of information processing and have analyzed the semantic structure of this knowledge. However, these studies lacked a structured representation of the knowledge associated with SRs; in particular, semantic categories and structural features were not integrated into the knowledge representation.

3. Methodology

3.1. Knowledge Transformation Framework for SRs

Considering the complexity of knowledge types in SRs, the corresponding knowledge transformation needs to consider the characteristics of knowledge content. To realize the transformation of SRs from abstract information to canonical knowledge, this paper analyzes the knowledge composition of ontology in the field of SRs and constructs a multi-stage knowledge transformation framework by combining the mechanism of ontology knowledge transformation (Figure 1). The mechanism defines both the scope of construction safety management and the criteria for judging deviations in implementation.

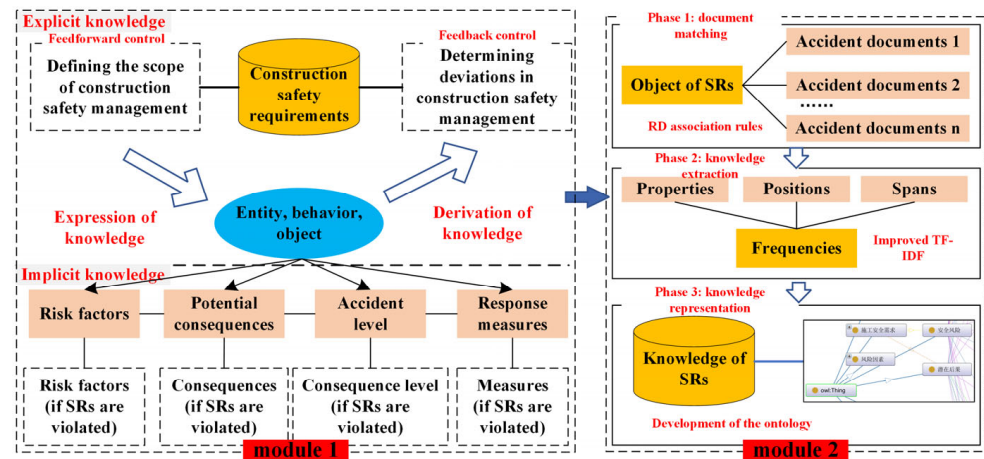


Figure 1. A knowledge transformation framework for SRs.

To design a rational knowledge transformation process for SRs, the whole framework is divided into two modules (i.e., knowledge dissections and knowledge transformation). Module 1 dissects the knowledge composition of SRs, including explicit and tacit knowledge. The concept of explicit knowledge suggests that the knowledge can be expressed in written text, diagrams, and numerical formulas. Accordingly, the requirement information is structurally expressed, which is consistent with the characteristics of explicit knowledge and therefore is explicit knowledge. Moreover, the three elements of SRs (entity, behavior, and object) are directly expressed ontological knowledge [1]. Further, this explicit knowledge can also be used to determine implementation deviations in construction safety management through knowledge transfer.

In comparison to explicit knowledge, tacit knowledge is widely acknowledged as knowledge that remains unarticulated and has not been effectively codified or transformed [43]. Specifically, tacit knowledge needs to be extracted gradually as the SRs are violated, including risk factors, potential consequences, accident levels, and response measures [4,16]. Accordingly, this knowledge is directly related to the SRs and can only be converted from “not present” to “present” when the requirements are violated.

Module 2 depicts the knowledge transformation process of SRs, including document matching, knowledge extraction, and knowledge representation. Document matching involves retrieving project documents based on the content of SRs and establishing requirement–document (RD) association rules to obtain valuable project documents. Conversion to knowledge extraction involves extracting the tacit knowledge in the acquired project documents and labeling the knowledge in response to different characteristics. Corresponding to knowledge representation, the extracted tacit knowledge is structurally expressed through a ternary model, and the domain ontology knowledge base is constructed.

3.2. The Association Between SRs and Documents

Since SRs are affected by semantic context, it may result in the matched documents violating the actual requirements due to semantic sentiment deviation. In this regard, it is also impossible to obtain each type of tacit knowledge. Thus, this study first identified the semantic tendency of the document and then established RD association rules for matching documents based on different semantic sentiments. Also, details of the individual variables in the equation are recorded in Table A1 in the Appendix A.

3.2.1. Semantic Tendency Judgement for Project Documents

This paper introduces the semantic tendency degree of viewpoint words to make semantic sentiment judgments on requirement information [44]. The semantic tendency

degree indicates the degree of semantic preference for behavioral elements, and the words corresponding to behavioral elements are viewpoint words. For example, the viewpoint words in “qualified cement” and “inferior formwork” represent positive and negative semantics in SRs, respectively.

The semantic tendency of viewpoint words is categorized into positive and negative semantic tendencies based on semantic sentiment. Meanwhile, this study also used the seed viewpoint words proposed by Rogers et al. [45] as a benchmark to calculate the semantic tendency degree of SRs. Accordingly, w_{opi} denotes the behavioral elements, and w_{seeds} denotes the seed viewpoint words, and the corresponding positive and negative semantic tendency degrees (SO_o^+ and SO_o^-) are calculated as follows.

$$\begin{cases} SO_o^+ = \frac{1}{|S_P|} \sum_{w_{seeds} \in S_P} \cos(v_{w_{opi}}^i, v_{j \leftarrow seeds}^+) \\ SO_o^- = \frac{1}{|S_N|} \sum_{w_{seeds} \in S_N} \cos(v_{w_{opi}}^i, v_{j \leftarrow seeds}^-) \end{cases} \quad (1)$$

where $v_{w_{opi}}^i$ denotes the affinity vector of w_{opi} in row i , and $v_{j \leftarrow seeds}^+$ and $v_{j \leftarrow seeds}^-$ are the row j vectors corresponding to the positive and negative tendency seed viewpoint words in the affinity matrix, respectively [1]. Moreover, the computation of similarity between w_{opi} and w_{seeds} is taken from the positive and negative inclination viewpoint word sets S_p and S_n , respectively.

According to the rules for calculating the semantic tendency degree, this paper classifies the documents to be associated as single-emotion documents and double-emotion documents. Specifically, documents with only one semantic tendency (positive or negative) of viewpoint words are defined as single-emotion documents, while documents with two semantic tendencies (positive and negative) of viewpoint words are defined as double-emotion documents. Meanwhile, the topic words of the documents are selected as the associated words, while the object elements are selected as the representative words of the SRs. The vocabulary can represent the core information of SRs, and two other elements can also be discovered indirectly through the relationship between the vocabularies. As a result, the semantic sentiment of each project document is obtained, and document matching can be performed according to the corresponding semantic sentiment category.

3.2.2. Design of RD Association Rules

This paper introduces the concept of relatedness proposed by Al Qady and Kandil [46] to the matching of vocabulary and documents and proposes the RD association rule for requirements and documents, as in Equation (2). In particular, this rule applies to single-emotion documents since it only considers the document to contain a single emotion attribute.

$$\begin{cases} \delta_i = \frac{w_{1st}^{d_i} + w_{2nd}^{d_i}}{w^{d_i}} \quad i = 1, 2, \dots, N \\ \text{Rel}(w^*, d_i) = [\lambda_1 \cdot Sso(w^*, w_{1st}) + \lambda_2 \cdot Sso(w^*, w_{2nd})] \times \delta_i \end{cases} \quad (2)$$

To address the association deviation of documents due to word count differences, a weighting factor δ_i is introduced into the RD association rule. This coefficient represents the proportion of the first two topic words to the total number of words in the document. $w_{1st}^{d_i}$ and $w_{2nd}^{d_i}$ are the word counts of the two topic words, w^{d_i} is the total number of words in the document, and N is the number of project documents to be matched. $Sso(w^*, w_{1st})$ and $Sso(w^*, w_{2nd})$ denote the semantic similarity between the object element w^* and the two topic words (denoted as w_{1st} and w_{2nd}) in the project document to be evaluated, and

the computation rule is as in Equation (3). λ_1 and λ_2 are moderators that take the value of the relative importance of semantic similarity, typically taking values of 0.7 and 0.3.

$$Sso(w^*, w_{1st}) = 1 - \left[\frac{Dis(w^*, w_{1st})}{2 \times Dis_{max}} \right]^{1/u} \quad (3)$$

where $Dis(w_i, w_j)$ is the corresponding semantic distance, the size of which is generally determined according to the semantic network. Furthermore, Dis_{max} is the maximum distance from the vocabulary to the root node, while u denotes the moderation factor (usually taken as 2). In a single-emotion document, both positive and negative sentiments can be matched as long as they are associated with SRs.

For double-emotion documents, the RD association rules need to be adjusted according to their sentiment features [47]. Since the document contains both positive and negative semantics, the degree of association is calculated by combining the two semantic emotions. Therefore, this paper introduces the emotional attribute coefficient (EAC) based on the RD association rule for single-emotion documents, which are positive and negative emotional attributes, as in Equation (4).

$$\begin{aligned} Rel(w, d_i) &= \{PEa \cdot [\lambda_1 \cdot Sso(w^*, w_{1st}) + \lambda_2 \cdot Sso(w^*, w_{2nd})] \\ &+ NEa \cdot [\lambda_1 \cdot Sso(w^*, w_{1st}) + \lambda_2 \cdot Sso(w^*, w_{2nd})]\} \times \delta_i \\ &= (PEa + NEa) \times [\lambda_1 \cdot Sso(w^*, w_{1st}) + \lambda_2 \cdot Sso(w^*, w_{2nd})] \times \delta_i \end{aligned} \quad (4)$$

In Equation (4), PEa and NEa represent the proportion of positive and negative semantics of viewpoint words, respectively, calculated as in Equation (5); $w_p^{d_i}$ and $w_N^{d_i}$ are the number of positive and negative viewpoint words, respectively; $w_{P+N}^{d_i}$ is the total number of viewpoint words of the document. This amendment rule considers both positive and negative semantics for the degree of satisfaction of the object elements so that it can be more accurately associated with the relevant documents.

$$PEa = \frac{w_p^{d_i}}{w_{P+N}^{d_i}}, \quad NEa = \frac{w_N^{d_i}}{w_{P+N}^{d_i}} \quad (5)$$

3.3. Knowledge Extraction of Documents

To accurately extract the knowledge in the associated documents, a keyword retrieval method was introduced into this study. Taking into account the current research trends, it can be seen that the TF-IDF algorithm is a popular method in keyword analysis due to the simplicity of the procedure [48–50]. The algorithm also considers the frequency and distribution of vocabulary in the document, which can reflect the importance of vocabulary [51]. Moreover, project documents are often long documents, and the rapid computational ability of the algorithm is suitable for the processing of large-scale text data. Therefore, the TF-IDF algorithm was screened for knowledge extraction in this study. For a certain term, TF-IDF is expressed as Equation (6).

$$TF\text{-}IDF(w_i) = tf \times idf = \frac{w_i^d}{w_{Total}^d} \times \log\left(\frac{N}{n_w} + 0.01\right) \quad (6)$$

where w_i^d is the number of words (w_i) in the document d , and w_{Total}^d is the total number of words in document d . N and n_w denote the total number of documents in the document set and the number of documents containing vocabulary w_i , respectively. However, the traditional TF-IDF algorithm primarily considers vocabulary weights based on word frequency, neglecting lexical features and the distribution of vocabulary within the document.

For documents containing tacit knowledge, keywords associated with different types of knowledge often exhibit distinct characteristics in terms of lexical properties, positional information, and distribution patterns within the document [52]. Consequently, relying solely on word frequency can lead to retrieval biases. For this reason, this paper proposes an improved TF-IDF algorithm by considering the weights of word properties, word positions, and word spans.

- (1) **Word properties:** According to the construction safety management terminology proposed by Malekitabar et al. [53], requirement keywords are usually represented by nouns or noun phrases, followed by verbs or quantifiers. In other words, a word can be deemed to have a higher level of significance when it frequently appears within the same lexical category; otherwise, its importance is considered lower. Thus, the given lexical weighting factor (W_cla) is as follows, where $w_{i_cla}^d$ denotes the total number of similar words in the document d :

$$W_cla = \frac{w_i^d}{w_{i_cla}^d} \quad (7)$$

- (2) **Word positions:** Research by scholars has shown that documents usually have keywords indicating the subject of the article in the article title or subsection headings [54]. Project documents, as standardized records, retain the aforementioned characteristics. Thus, the extraction of implicit knowledge keywords must account for the specificity of word positions. Moreover, this study used three levels to calculate word weight coefficients (W_loc), and $w_{i_loc}^{d*}$ and $w_{i_loc}^d$ are the total number of words in article titles and subsection headings, respectively:

$$W_loc = \begin{cases} 1 + \frac{w_i^d}{w_{i_loc}^{d*}}, & \text{if location of } w_i \in (\text{title or subsection headings}) \\ 1 + \frac{w_i^d}{w_{i_loc}^d}, & \text{if location of } w_i \in (\text{first or last paragraph}) \\ 1, & \text{otherwise} \end{cases} \quad (8)$$

- (3) **Word spans.:** There may be keyword recognition bias in the project document caused by the frequent occurrence of word w_i locally in the document, so it is necessary to consider whether the keywords are distributed across the document. In other words, an important characteristic of global keywords is the number of spanning paragraphs. Equation (9) is used in this paper to calculate the spanning weight of the words, and l_w_i and L are the number of paragraphs in which the word is located and the total number of document paragraphs, respectively:

$$W_spa = 1 + \frac{l_w_i}{L} \quad (9)$$

As a result, this paper combines the weight coefficients that affect the degree of importance of the keywords of tacit knowledge and proposes an improved TF-IDF calculation rule, as shown in Equation (10).

$$\text{TF-IDF}(w_i) = \frac{w_i^d}{w_{Total}^d} \times \log\left(\frac{N}{n_w} + 0.01\right) \times W_cla \times W_loc \times W_spa \quad (10)$$

Thus, the importance degree of each candidate keyword can be obtained by using this rule, and the keywords corresponding to the tacit knowledge can be extracted based on the priority ranking. As a result, four types of tacit knowledge can be obtained by analyzing the relevant information of the keywords.

3.4. Representation of Ontological Knowledge

3.4.1. Defining Ontology Classes and Relationships

Based on the explicit and tacit knowledge corresponding to SRs, the scope of domain ontology knowledge includes SRs, risk factors, potential consequences, accident levels, and response measures. Consequently, this knowledge is categorized into distinct classes, each comprising multiple subclasses. Simultaneously, there are instances of each class, and the content and association of these elements form the knowledge structure of the entire ontology.

The object elements were used as information for SRs to be expressed in the ontology due to their low probability of repetition in the same project. Further, this study classified risk factors into five categories including human factors based on the 4M1E approach [55,56]. The potential consequences were then identified as casualties, economic losses, and construction disruptions. According to the hazard level of safety accidents, they were divided into four classes including particularly significant accidents. Finally, the response measures include safety risk prevention and safety accident disposal.

3.4.2. Attribute Definition and Creation of Ontologies

This paper represents the ontological knowledge of SRs in Protégé 5.5.0 and establishes the object attributes and data attributes between each class. The attribute definition and creation process of the ontology are depicted in Figure 2.

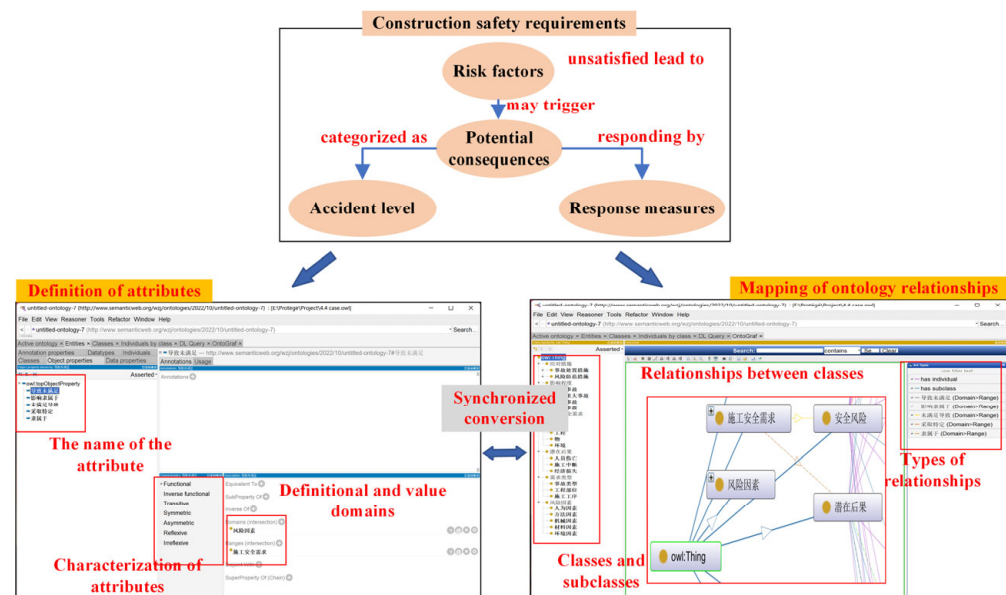


Figure 2. Attribute definition and creation process of ontology.

Object attributes refer to the properties of a class associated with another class, i.e., the edges of a directed graph in the semantic web. In this regard, this study identified several object attributes based on the relationships between classes. From the knowledge derivation process of SRs, it is known that SRs “unsatisfied leads to” risk factors, indicating that “unsatisfied leads to” is the object attribute that connects SRs to risk factors. Correspondingly, the object attribute between risk factors and potential consequences is “may trigger”, and the potential consequences belong to the degree of accidents and are connected by the object attribute “categorized as”. Furthermore, the object attribute between potential consequences and countermeasures is defined as “responding by”, indicating that different potential consequences require appropriate measures.

Data attributes characterize the properties of the class itself. Furthermore, data attributes are configured in relation to specific instances, with their definition domains (corresponding to different classes) and value domains (representing data ranges) explicitly delineated. Diverse data attributes exist in various subclasses corresponding to SRs (including worker, machine, and material), and therefore are set up adaptively. In contrast, the contents of other classes or subclasses are more confusing and need to be defined according to the characteristics of the class. For instance, the potential consequences were subdivided into the subcategories of “casualties”, “economic losses”, and “construction interruption”, and the corresponding data attributes were determined as the number of people, amount, and time. Also, specific symbols were introduced to represent the data attributes, including “string”, “int”, and “float”.

3.4.3. Instance Creation of Ontologies

For the domain ontology of SRs, the essence of creating instances expresses the classes through specific content. After clarifying the relevant concepts, object attributes, and data attributes of the domain ontology of SRs, appropriate instances were added for each class. By adding instances, the objects of each class can be clarified, which helps the practical application of ontology knowledge and logical reasoning. In this study, we added instances for different levels of classes in the individual interface of Protégé 5.5.0 and edited the related contents.

Also, the addition of instances must be combined with the definition of classes and the setting of object properties and data attributes. Accordingly, the content of instances in each class was obtained from real projects to build a reliable ontology knowledge base. By mining the ontological knowledge expressed by instances, the contents of classes, relationships between classes, attributes, and many other contents can be obtained, thus realizing the transformation of domain ontological knowledge in SRs.

4. Result Analysis

4.1. Data Collection

To maximize the generalizability of the research findings, this study selected two types of construction projects as case studies, based on project scale, construction techniques, and management models. Given that the majority of construction projects are either civil or industrial buildings, these two categories were chosen as examples to identify appropriate projects for the case studies. The civil building is an 18-story monolithic building with a total floor area of 11,214 m². During the construction of this project, a steel mold support collapse accident occurred and caused one death and nine injuries. The industrial building was a production plant of a pharmaceutical factory, with a total construction area of 31,206.24 m², and the main body is a steel and frame structure. There was a heavy fall accident caused by a falling steel beam, which resulted in one death and two injuries.

For the documents used for knowledge extraction, this study relied on the scraps of Python 3.9 to crawl from accident cases published officially. Moreover, the project types were identified as civil and industrial buildings, and the project sizes were similar to the two cases. Meanwhile, the accident types were specified as “collapse of an object” and “struck by falling object”. A total of 183 accident case files were collected for testing in 2021–2022. Further, by determining whether the cause of the accident, risk analysis, and response measures were included in the project documents, 27 documents with incomplete information were excluded, and 156 valid accident case documents were obtained. This includes 117 documents related to civil buildings and 39 documents related to industrial buildings. All accident case information was systematically organized, stored in document form, and categorized with appropriate labels.

4.2. Results of SRs and Document Matching

Before document matching, the SRs of these two types of buildings need to be clarified. Thus, this study analyzed the key project documents of the two types of buildings, which are Special Construction Plan for Steel Mould Construction and Minutes of the Safety Management Meeting for Steel Structures Hoisting, respectively. Meanwhile, text mining was used to analyze these two project documents to obtain information about SRs. Finally, the object elements of civil construction were identified as “steel template”, “quality”, “check”, “precision”, and “perpendicularity”, while the object elements for industrial buildings were identified as “lifting”, “crane”, “wire rope”, “conductor”, and “signal”.

Concomitantly, the RD association rule was used to calculate the match between the object elements of SRs and accident case documents, and accident cases with high association degrees were selected as data sources for knowledge extraction. In this study, three accident case documents with the strongest importance (top three associations) were selected as target documents for knowledge extraction in conjunction with the calculation results, and the results are depicted in Figure 3. Figure 3a shows that the largest match for civil buildings is the accident case document corresponding to “quality” (Rel = 0.771), which also corresponds to accidents mainly due to defective quality of formwork, which is highly compatible with the requirement information recorded in the document Special Construction Program for Steel Formwork Construction.

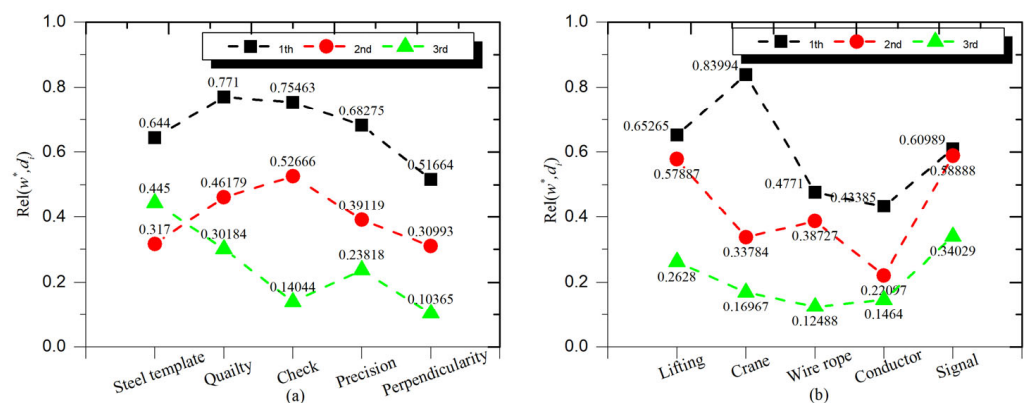


Figure 3. Associated results between object elements and accident case documents (top 3).

In Figure 3b, the correlation degree of the documents in the first two levels is relatively close, and there are deviations only in the documents corresponding to the keywords “crane” and “conductor”. Moreover, the document with the largest correlation (Rel = 0.840) records the accident caused by the driver’s faulty operation, which also matches the requirement information recorded in the document Minutes of Safety Management Meeting on Steel Structure Lifting. The results show that the documents matched according to the RD association rules are consistent with the actual SRs, which indicates that the method has a better effect on document matching.

Moreover, this study calculated the association results of single-emotion documents and double-emotion documents separately, and the results are depicted in Figure 4. Among them, the horizontal coordinate is the weight coefficient δ_i of the matching rule, which indicates the variation of the association degree with δ_i , while the vertical coordinate is the association degree. Meanwhile, “steel template” and “lifting” were selected as the object elements for document matching in civil and industrial buildings, respectively, and the top 10 accident case documents were finally associated as the analysis objects. Figure 4a shows the retrieval status of a single-emotion document, where industrial buildings in the weight coefficient δ_i are larger when the association degree is strong ($\delta_i > 0.1$), while civil buildings are relatively weak. Meanwhile, as δ_i increases, the corresponding document

matching degree shows an increasing trend but fails to show an incremental increase. This result reflects that the RD association rule is not much affected by δ_i in single-emotion documents and is mainly affected by semantic similarity.

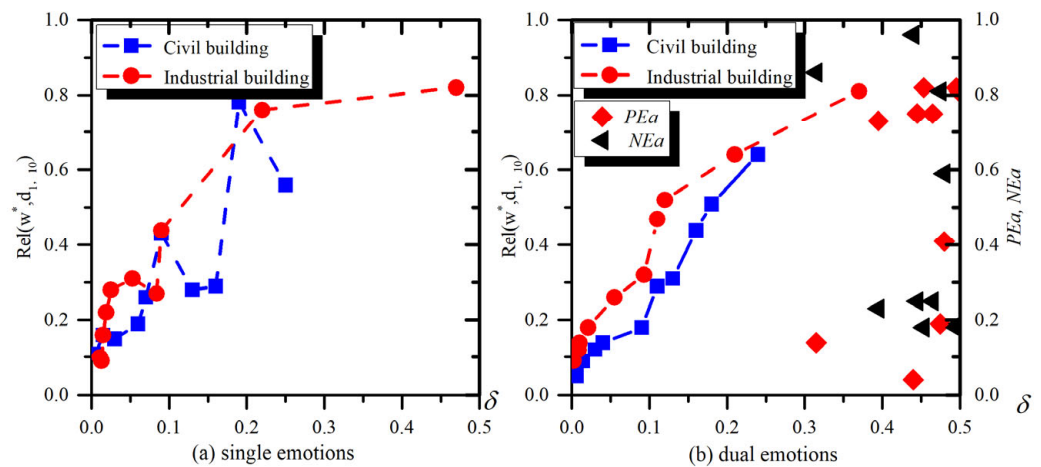


Figure 4. Matching of object elements to single- and dual-emotion documents in two buildings.

The results from the double-emotion documents indicate that as δ_i increases, the correlation curve gradually rises and becomes smoother. Notably, the maximum matching degrees of “steel template” (Rel = 0.642) and “lifting” (Rel = 0.813) demonstrate excellent correlation effects, respectively. This also shows that the sentiment attribute refines the RD association rule, which makes the rule compatible with the influence of both positive and negative sentiments of documents to achieve more accurate associations. Moreover, the scatter plot in Figure 4b shows the sentiment attributes when the association is maximized. For instance, the sentiment attributes of “steel template” at the time of maximum relevance are 0.14 (*PEa*) and 0.86 (*NEa*), respectively, and the relevance increased by 0.042 compared to a single-emotion document, which suggests that there is a possibility of improving the accuracy of document association when considering the double emotion of a document. Thus, the weighting coefficients also strike a balance between document keyword share and quantitative matching, making the associated documents more compatible with the actual SRs.

4.3. Knowledge Extraction Results for Documents

To facilitate the improved TF-IDF algorithm to extract knowledge from documents, the collected documents in this study were divided into a training set and a test set. The test set comprises six accident case documents with high relevance, as matched above, while the training set consists of the remaining 150 accident case documents. For improved comparative analysis, this study selected the cause of the accident in each document as the corpus and used the corresponding document label as the keyword. To comprehensively compare the performance of different algorithms for the extraction of tacit knowledge keywords, this study also applied the TextRank algorithm and the TF-IDF algorithm for testing, corresponding to the evaluation metrics of precision, recall, and F-measure, and the test results are recorded in Figure 5a,b,c, respectively.

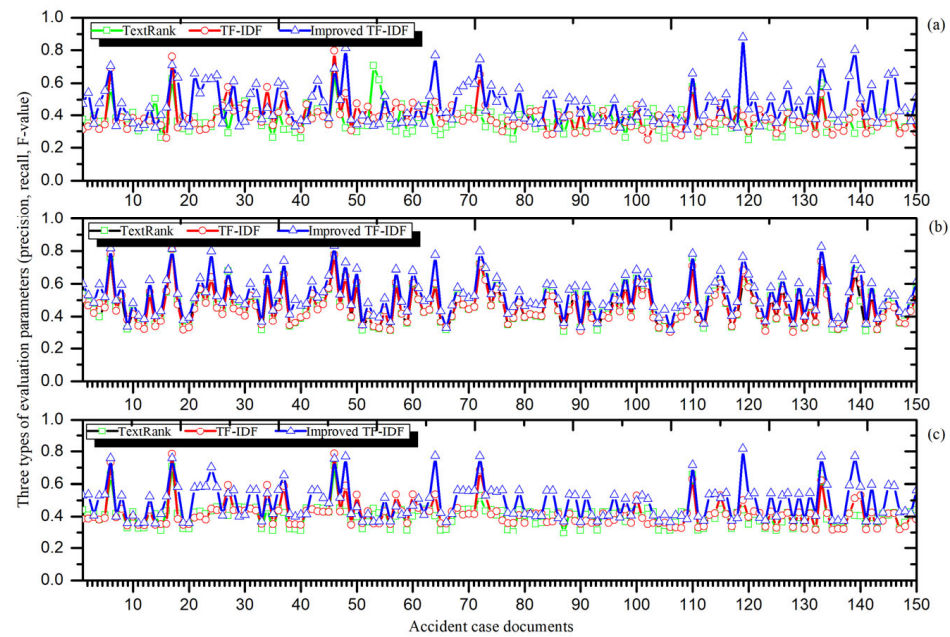


Figure 5. The effect of different algorithms for the extraction of object elements.

The results depicted in Figure 5 show that compared to the TextRank algorithm and TF-IDF algorithm, the improved TF-IDF algorithm is better than the other two types of algorithms in terms of precision rate and recall, and it performs well in analyzing almost all the accident case documents. From the mean value, Figure 5a,b shows that the precision rate of the traditional two types of algorithms is 0.3691 and 0.3857, while the recall rate is 0.4651 and 0.4613. In contrast, the precision rate of the improved TF-IDF algorithm is 0.4618, and the recall rate is 0.5278. The results show that there is a significant improvement in both precision rate and recall rate. Besides, Figure 5c shows that the F-value of the improved TF-IDF algorithm is improved by nearly 20% compared to the conventional method. This reflects to some extent that the improvement of the traditional TF-IDF algorithm in this study can accurately grasp the critical contents of retrieval and accurately obtain the tacit knowledge of SRs. Moreover, the knowledge extracted using the improved TF-IDF algorithm needs to be compared with other algorithms to reduce the error.

In this study, the improved TF-IDF algorithm was utilized to analyze the associated accident case documents and extract the effective tacit knowledge of SRs. To facilitate the establishment of the domain ontology knowledge base, this study selected 10 safety risk factors based on the weight ordering. The results in Figure 6a show that four factors were extracted from the first document, and the factor with the highest weight was “irregularity” (0.5142), followed by “substandard materials”. The safety risk factors in the second document are “unqualified riser”, “uninspected”, and “insufficient safety awareness”, with a maximum weight of 0.4405.

The importance of the third document is relatively low, with a maximum weight of only 0.3281. Also, most of the potential consequences are “casualties”. The factors categorized as larger accidents at the accident level are “irregularity” and “substandard materials”, both from the first document. For industrial buildings, Figure 6b shows that four factors were also extracted from this document, with a maximum weight of 0.7233 corresponding to “operating without a license” and the weight corresponding to “irregularity” is also higher (0.5883). The last two documents extracted three factors, and the weights were lower. Compared to civil buildings, the extracted tacit knowledge in industrial buildings is more dependent on a particular document.

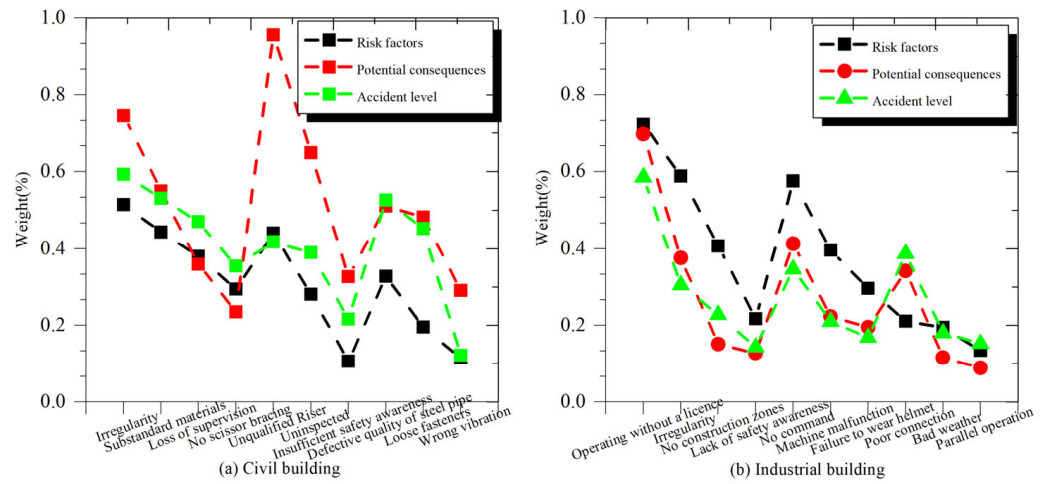


Figure 6. Extracted tacit knowledge corresponding to SRs.

To derive effective response measures, this study employed an improved TF-IDF algorithm to extract tacit knowledge from accident case documents. The extraction results, as shown in Figure 7, identify the response measures with the highest weights for the corresponding factors. Additionally, more than three response measures were selected for each factor through a combination of algorithmic analysis and manual evaluation. For civil buildings, the response measure corresponding to the safety risk factor “irregularity” is “induction training”, with a weight of 0.4502. The response measure corresponding to the safety risk factor “substandard materials” is “induction training”, with a weight of 0.4502. Accordingly, “substandard materials” corresponds to “check quality”. For industrial buildings, “operating without a license” corresponds to “licensed to work”, and “no construction zones” corresponds to “setting up construction exclusion zones”. These responses were taken from accident case documents and are the most critical.

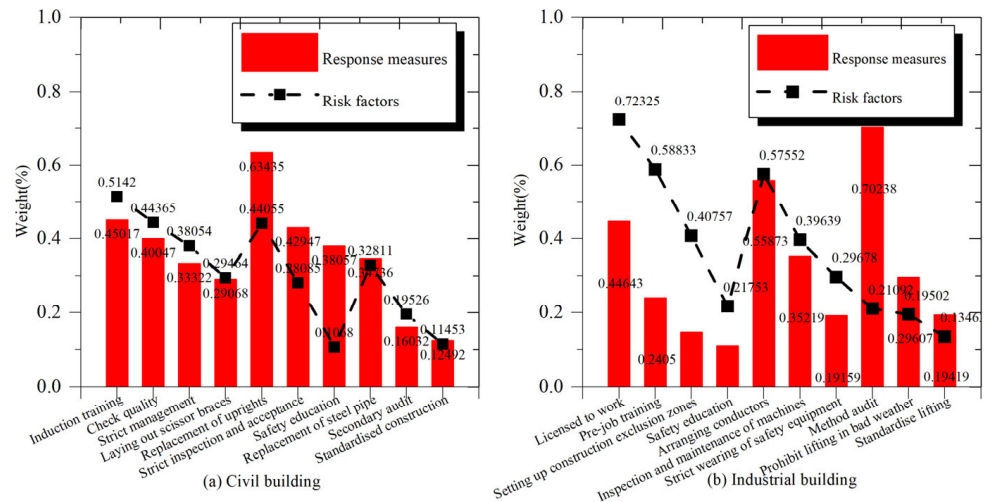


Figure 7. Maximum weighted response measures corresponding to each risk factor.

4.4. Results of Ontological Knowledge Expression

(1) Determination of relationships between classes and attributes

To standardize the representation for each type of knowledge within SRs, it is essential to define the relationships between each class and its corresponding attributes. The study was conducted in two distinct stages.

Stage 1 involved determining the relationship between classes. This study identified the mechanism of SRs in conjunction with the classical construction safety manage-

ment paradigm, and this was used to determine the relationships between classes. For instance, SRs induce risk factors once violated, and risk factors trigger potential consequences. These are the relationships between the classes, which also serve as the basis for attribute determination.

Stage 2 determined the attributes of the class. For SRs and risk factors, this study determined the object attributes of “unsatisfied leads to”. For example, the object element “steel template” is associated with “irregularity” by “unsatisfied leads to”. For industrial buildings, the object element “lifting” is associated with “operating without a license” via “unsatisfied leads to”. Accordingly, three attributes were identified between the risk factors and the other classes. The object attribute identified between risk factors and potential consequences is “may trigger”, while the object attribute identified between potential consequences and accident level is “categorized as”. Additionally, some subcategories are represented as data attributes, such as “yuan” for economic loss and “day” for delay. This study obtained the relationship between the classes and realized the canonical representation of the extracted knowledge.

(2) Import of instances

Next, this study imported instances of each class for constructing a complete domain ontology knowledge base for the knowledge transformation of SRs. Thus, instances corresponding to each class were imported into Protégé, and annotations were added to each class. According to the distribution of knowledge, the SRs class belongs to explicit knowledge and includes 10 pieces of information corresponding to five keywords (object elements). Moreover, six strongly correlated documents were obtained from the matching results of the accident case documents for extracting the tacit knowledge of SRs. Further, 10 safety risk factors (i.e., risks that may be triggered when SRs are violated) were extracted for both types of buildings. Accordingly, potential consequences, accident levels, and response measures were extracted corresponding to the identified risk factors. At this stage, instances of this knowledge can be generated and integrated into the platform.

In Figure 8a,b, instances of the imported knowledge are shown for civil and industrial buildings, respectively, and the corresponding annotations are added. According to the knowledge transformation framework of SRs, explicit and tacit knowledge can be extracted, and the correlation between individual knowledge can be clarified. To reveal the correlations between the knowledge, two requirement elements were used as examples in this study. For two types of buildings, “steel template” and “lifting” were selected as object elements, and the risk factors extracted from the former are “irregularity” and “substandard materials”, respectively. The risk factors extracted from the former are “irregularity”, “substandard materials”, and “loss of supervision” (the top three weights), and their potential consequences are “casualties” and three response measures. The latter corresponded to the factors “operating without a license”, “irregularity”, and “no construction zones”, and the response measures also consisted of three items. The details are recorded in Table 1.

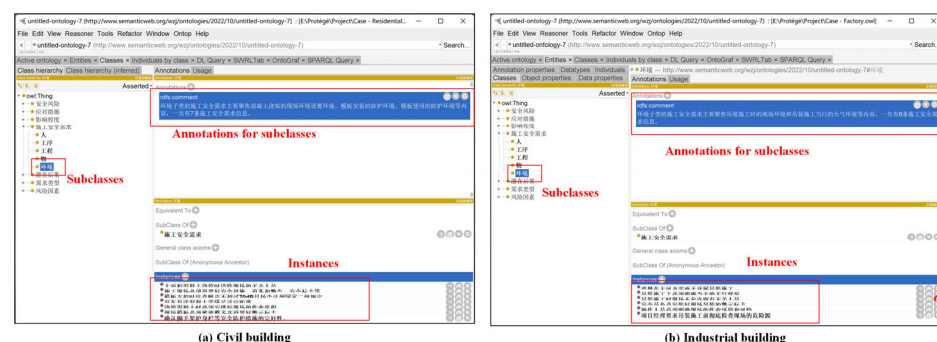


Figure 8. Instances of ontological knowledge for SRs.

Table 1. Knowledge details of selected SRs.

Case	SRs	Risk Factors	Potential Consequences	Accident Level	Response Measures
Civilian building	Steel template	Personnel, Facilities, Management	Casualties, Economic losses, Construction disruptions	Larger accidents or major accidents	1. Procure 2. Comprehensively 3. Check.
Industrial building	Lifting	People, Weather, Management	Casualties, Economic losses, Construction disruptions	General or major accidents	1. The project. 2. Immediately 3. Establish.

Accordingly, the various types of knowledge of SRs are stored through the domain ontology knowledge base, and the knowledge is represented according to different classes. Meanwhile, the instances corresponding to each class are imported into the knowledge base, adding substance to each class. The domain ontology knowledge base constructed in this study based on Protégé is capable of both extracting the various types of knowledge and describing the complex relationships between them. Further, SRs are transformed into effective knowledge that can be used for safety management through information recording and relationship grooming.

5. Discussion

A knowledge transformation framework was developed in this study for realizing the transformation of SRs from abstract information to canonical knowledge, which is accurately accomplished through document matching, knowledge extraction, and knowledge representation. Overall, the main findings of this study are as follows:

- (1) The relevance of the retrieved accident case documents according to the RD association rules can reach 0.771 and 0.840, showing that the method has excellent performance in document matching. Moreover, the document association accuracy of the object elements needs to take into account the affective attributes of the documents so that the associated accident case documents are closely related to the SRs for accurate knowledge extraction. In other words, embedding document sentiment attributes into the corresponding association rules is consistent with the previous studies proposing to combine semantic sentiment fitness to assess document similarity [47,57]. Since the behavior elements in the information structure of SRs have sentiment preferences, the combined consideration of document sentiment attributes also ensures the applicability of the established RD association rules;
- (2) The improved TF-IDF algorithm improved the precision and recall by 20%, which shows that this study improved the traditional TF-IDF algorithm by combining word properties, word positions, and word spans to accurately grasp the key of knowledge extraction, thus obtaining more accurate tacit knowledge. Meanwhile, this study calculated the weights of the corresponding tacit knowledge according to the improved TF-IDF algorithm and sequentially obtained the tacit knowledge in the accident case documents. Traditional TF-IDF algorithms ignore the effects of word properties, word positions, and word spans on document topics, but project documents that harbor knowledge of various types of SRs have these features [52,58]. Once the extraction method fails to respond accurately to these features, it may result in biased knowledge extraction. In contrast, the improved TF-IDF algorithm confirms the interference of document features on knowledge extraction based on its excellent performance. Further, it was shown that combining document features for knowledge extraction can either obtain complete information that satisfies the requirements or uncover tacit knowledge that is covered by “false topics”;

- (3) In this study, the analysis of the extracted explicit and tacit knowledge was completed, and the classes and subclasses in the domain ontology were identified, including the relationships between different classes and two attributes (object and data). Moreover, instances in the tacit knowledge were obtained and imported into Protégé, thus creating a complete knowledge base of the domain ontology. As the structural characteristics of information in SRs have been defined in previous studies, the corresponding types of knowledge also have specific information structures [1,2,7,8]. The results of the study show that dealing with explicit and tacit knowledge based on the components (i.e., internal) and connecting relationships (i.e., external) of the knowledge is a reasonable model that can express the information core of the knowledge completely and accurately. Moreover, the domain ontology knowledge base achieves normative documentation for each type of knowledge in SRs, which facilitates knowledge application and value addition.

This paper introduces a semi-automated model to develop a domain ontology knowledge base for knowledge transformation of SRs and tries to achieve a complete record and standardized representation of requirement information. In this regard, the main contributions of this study are as follows. First, this study developed a multi-stage knowledge transformation framework that analyzes the explicit knowledge expressed by the three elements in SRs and condenses the knowledge transformation process based on the mechanism of action when requirements are violated. Meanwhile, this study also specified a multi-stage knowledge expression model for relationships between classes, ontological attributes, and instances based on this framework. Secondly, an RD association rule associating SRs and project documents was designed for matching documents and can obtain the most relevant documents based on their sentiment attributes. The RD association rule can accurately capture project documents directly related to SRs and use them as raw data for knowledge extraction. Thirdly, word properties, word positions, and word spans were fused into the keyword weights, and an improved TF-IDF algorithm was developed in this study. This algorithm improves precision and recall compared to the TextRank algorithm and traditional TF-IDF algorithm and can extract tacit knowledge based on weight ordering. Finally, a domain ontology knowledge base was developed in this study, which achieves the normative documentation and representation for explicit and implicit knowledge of SRs. Moreover, this knowledge base can reveal the association relationship between the classes in conjunction with the mechanism of action when SRs are violated, enabling dynamic updating after the instance is imported.

The results of this study also provide implications for project management on construction sites. The multi-stage knowledge framework proposed in this study can assist project managers in acquiring knowledge of SRs associated with construction projects and clarifying safety preferences for current techniques and segments. Meanwhile, the mechanism of action when SRs are violated can identify potential safety risks to formulate countermeasures in advance. Also, project managers can use the domain ontology knowledge base in Protégé 5.5.0 to manage the various types of knowledge in SRs and evaluate implementation deviations through knowledge retrieval and analysis. Moreover, project managers can leverage the strategy generation within the knowledge base to facilitate safety-related decision making.

6. Conclusions

As SRs present redundant information characteristics and possess the ability of knowledge migration, a large amount of latent valuable knowledge is left unexplored. Moreover, a large amount of knowledge fails to achieve normative documentation and representation, resulting in low utilization of this knowledge. Therefore, this study analyzed the

knowledge transformation mechanism of SRs and developed a multi-stage knowledge transformation framework for discovering and documenting various types of knowledge in SRs. On the one hand, this study utilized the framework to obtain complete requirements information and identified three elements for expressing ontological knowledge for representing explicit knowledge. Meanwhile, this study discovered the tacit knowledge when SRs are violated through knowledge migration and identified the knowledge categories and components. On the other hand, this study identified three stages (including document matching, knowledge extraction, and knowledge representation) to realize the transformation of SRs from abstract information to canonical knowledge, which facilitates the application of this knowledge for safety management. Also, a semi-automated model was introduced to develop a domain ontology knowledge base for completing the normative documentation and representation for each type of knowledge in the SRs, and the dynamic updating of the knowledge base was realized by instance import.

Although this study attempted to uncover and transform complete knowledge about SRs, there are still limitations. This study only extracted knowledge from project documents, while other types of data generated by the safety management process were not utilized. These multi-source heterogeneous data (e.g., images, audio, etc.) also contain a large amount of valuable requirement information waiting to be uncovered. With the maturity of different types of data processing technologies, it will be possible in the future to obtain information and perform knowledge transformations from multiple types of data. Meanwhile, this study only improved the TF-IDF algorithm but failed to embed the characteristics of SRs into other algorithms, therefore lacking the comparison of the knowledge transformation effect between multiple algorithms. For instance, the TextRank algorithm extracts knowledge based on recommender relationships, and there is a derived relationship between the various types of knowledge in SRs. The next study can embed this derived relationship into the TextRank algorithm to design recommendation rules to complete the knowledge extraction under specific relationships. The LDA topic model tends to combine topic clustering to obtain topic words, which can be improved based on the knowledge of the same type of SRs in the future to achieve the topic extraction of tacit knowledge without specific clues. Furthermore, there is still a large amount of work requiring manual intervention (e.g., document tagging) in the knowledge transformation of SRs, which is both time-consuming and prone to bias. Autonomous labeling by artificial intelligence (AI) brings convenience to this work and provides new ideas for future research.

Author Contributions: Conceptualization, Z.W., M.L. and G.M.; methodology, Z.W. and M.L.; software, Z.W. and M.L.; validation, Z.W.; writing—original draft preparation, Z.W. and M.L.; writing—review and editing, Z.W. and M.L.; project administration, Z.W. and M.L.; funding acquisition, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: The research described in the paper is funded by National Natural Science Foundation of China (Grant No: 72401250), China Postdoctoral Science Foundation (Grant No: 2024M761993), and Natural Science Foundation of Yangzhou Municipality (Grant No: YZ2024165). We also appreciate the detailed suggestions and comments from the editor and the anonymous reviewers.

Institutional Review Board Statement: The study has been conducted according to the Declaration of Helsinki. Ethical review and approval were waived for this study, as it did not include any ethical issues, and the data did not collect any personally identifiable data.

Informed Consent Statement: Informed consent was obtained from the provider for all data used in this study. Participants were allowed to withdraw from the survey at any time during the course of the survey and were also informed that the data would be used for academic purposes only.

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

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

Appendix A

Table A1. Notations table in the equation.

Variables	Notion of Variables	Variables	Notion of Variables
w_{opi}	Refers to viewpoint words (corresponding to behavioral elements in SRs).	SO_o^+, SO_o^-	Refers to the positive and negative semantic tendency.
w_{seeds}	Refers to seed viewpoint words.	$v_{w_{opi}}^i$	Refers to the affinity vector of w_{opi} in row i .
$v_{j \leftarrow seeds}^+$	Refers to the row j vectors corresponding to the positive and negative tendency seed viewpoint words in the affinity matrix, respectively.	S_p, S_n	Refers to the positive and negative inclination viewpoint word sets.
$v_{j \leftarrow seeds}^-$			
$w_{1st}^{d_i}, w_{2nd}^{d_i}$	Refers to the word counts of the two topic words.	w^{d_i}	Refers to the total number of words in the document.
$Dis(w_i, w_j), Dis_{max}$	Refers to the corresponding semantic distance and the maximum distance from the vocabulary to the root node.	PEa, NEa	Refers to the proportion of positive and negative semantics of viewpoint words, respectively.
$w_p^{d_i}, w_N^{d_i}$	Refers to the number of positive and negative viewpoint words.	$w_{P+N}^{d_i}$	Refers to the total number of viewpoint words of the document.
w_i^d, w_{Total}^d	Refers to the number of words (w_i) and total number in the document.	$w_{i_cla}^d$	Refers to the total number of similar words in the document.
$w_{i_loc}^{d*}, w_{i_loc}^d$	Refers to the total number of words in article titles and subsection headings, respectively.	l_{w_i}, L	Refers to the number of paragraphs in which the word is located and the total number of document paragraphs, respectively.

References

1. Wu, Z.; Ma, G. NLP-based approach for automated safety requirements information retrieval from project documents. *Expert Syst. Appl.* **2024**, *239*, 122401. [\[CrossRef\]](#)
2. Wang, X.; El-Gohary, N. Deep learning-based relation extraction and knowledge graph-based representation of construction safety requirements. *Autom. Constr.* **2023**, *147*, 104696. [\[CrossRef\]](#)
3. Zou, Y.; Kiviniemi, A.; Jones, S.W. Retrieving similar cases for construction project risk management using Natural Language Processing techniques. *Autom. Constr.* **2017**, *80*, 66–76. [\[CrossRef\]](#)
4. Wang, H.-H. Semi-automated identification of construction safety requirements using ontological and document modeling techniques. *Can. J. Civ. Eng.* **2015**, *42*, 756–765. [\[CrossRef\]](#)
5. Kim, I.; Lee, Y.; Choi, J. BIM-based Hazard Recognition and Evaluation Methodology for Automating Construction Site Risk Assessment. *Appl. Sci.* **2020**, *10*, 2335. [\[CrossRef\]](#)
6. Blanc, F.; Ottimofiore, G.; Myers, K. From OSH regulation to safety results: Using behavioral insights and a “supply chain” approach to improve outcomes—The experience of the health and safety Executive. *Saf. Sci.* **2022**, *145*, 105491. [\[CrossRef\]](#)
7. Venkatesh, P.; Ergan, S. Classification of Challenges in Achieving BIM-Based Safety-Requirement Checking in Vertical Construction Projects. *J. Constr. Eng. Manag.* **2023**, *149*, 04023131. [\[CrossRef\]](#)
8. Wang, X.; El-Gohary, N. Deep Learning-Based Named Entity Recognition and Resolution of Referential Ambiguities for Enhanced Information Extraction from Construction Safety Regulations. *J. Comput. Civ. Eng.* **2023**, *37*, 04023023. [\[CrossRef\]](#)
9. De Melo, R.R.S.; Costa, D.B.; Alvares, J.S.; Irizarry, J. Applicability of unmanned aerial system (UAS) for safety inspection on construction sites. *Saf. Sci.* **2017**, *98*, 174–185. [\[CrossRef\]](#)
10. Tam, V.W.Y.; Fung, I.W.H. Tower crane safety in the construction industry: A Hong Kong study. *Saf. Sci.* **2011**, *49*, 208–215. [\[CrossRef\]](#)
11. Yang, H.; Chew, D.A.S.; Wu, W.; Zhou, Z.; Li, Q. Design and implementation of an identification system in construction site safety for proactive accident prevention. *Accid. Anal. Prev.* **2012**, *48*, 193–203. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Penaloza, G.A.; Saurin, T.A.; Formoso, C.T. Identification and assessment of requirements of temporary edge protection systems for buildings. *Int. J. Ind. Ergon.* **2017**, *58*, 90–108. [\[CrossRef\]](#)
13. Fung, I.W.H.; Lee, Y.Y.; Tam, V.W.Y.; Fung, H.W. A feasibility study of introducing chin straps of safety helmets as a statutory requirement in Hong Kong construction industry. *Saf. Sci.* **2014**, *65*, 70–78. [\[CrossRef\]](#)

14. Zailani, B.M.; Moda, H.; Ibrahim, Y.M.; Abubakar, M. Improving the antecedents of non-compliance to safety regulations toward an optimized self-regulated construction environment in Nigeria. *Int. J. Occup. Saf. Ergon.* **2023**, *29*, 1212–1219. [[CrossRef](#)]
15. Agapiou, A. A Systematic Review of the Socio-Legal Dimensions of Responsible AI and Its Role in Improving Health and Safety in Construction. *Buildings* **2024**, *14*, 1469. [[CrossRef](#)]
16. Lu, Y.; Li, Q.; Zhou, Z.; Deng, Y. Ontology-based knowledge modeling for automated construction safety checking. *Saf. Sci.* **2015**, *79*, 11–18. [[CrossRef](#)]
17. Famakin, I.O.; Aigbavboa, C.; Molusiwa, R. Exploring challenges to implementing health and safety regulations in a developing economy. *Int. J. Constr. Manag.* **2023**, *23*, 89–97. [[CrossRef](#)]
18. Azhar, S.; Choudhry, R.M. Capacity building in construction health and safety research, education, and practice in Pakistan. *Built Environ. Proj. Asset Manag.* **2016**, *6*, 92–105. [[CrossRef](#)]
19. Patel, D.A.; Jha, K.N. Structural Equation Modeling for Relationship-Based Determinants of Safety Performance in Construction Projects. *J. Manag. Eng.* **2016**, *32*, 05016017. [[CrossRef](#)]
20. Mostofi, F.; Togan, V. Construction safety predictions with multi-head attention graph and sparse accident networks. *Autom. Constr.* **2023**, *156*, 105102. [[CrossRef](#)]
21. Simpeh, F.; Bamfo-Agyei, E.; Amoah, C. Barriers to the implementation of COVID-19 safety regulations: Insight from Ghanaian construction sites. *J. Eng. Des. Technol.* **2022**, *20*, 47–65. [[CrossRef](#)]
22. Raatikainen, M.; Mannisto, T.; Tommila, T.; Valkonen, J. Challenges of Requirements Engineering—A Case Study in Nuclear Energy Domain. In Proceedings of the 19th IEEE International Requirements Engineering Conference (RE), Trento, Italy, 29 August–2 September 2011; pp. 253–258.
23. Shen, Y.; Xu, M.; Lin, Y.; Cui, C.; Shi, X.; Liu, Y. Safety Risk Management of Prefabricated Building Construction Based on Ontology Technology in the BIM Environment. *Buildings* **2022**, *12*, 765. [[CrossRef](#)]
24. Chen, D.; Zhou, J.; Duan, P.; Zhang, J. Integrating knowledge management and BIM for safety risk identification of deep foundation pit construction. *Eng. Constr. Archit. Manag.* **2023**, *30*, 3242–3258. [[CrossRef](#)]
25. Shen, Q.; Wu, S.; Deng, Y.; Deng, H.; Cheng, J.C.P. BIM-Based Dynamic Construction Safety Rule Checking Using Ontology and Natural Language Processing. *Buildings* **2022**, *12*, 564. [[CrossRef](#)]
26. Ma, G.; Ding, Y.; Ma, J. The Impact of Airport Physical Environment on Perceived Safety and Domestic Travel Intention of Chinese Passengers during the COVID-19 Pandemic: The Mediating Role of Passenger Satisfaction. *Sustainability* **2022**, *14*, 5628. [[CrossRef](#)]
27. Ebekozi, A.; Aigbavboa, C.; Samsurijan, M.S.; Radin Firdaus, R.B.; Rohayati, M.I. Investigating Safety Violations on Nigerian Construction Sites. *Int. J. Constr. Manag.* **2024**, *24*, 1454–1464. [[CrossRef](#)]
28. Yang, Z.; Yuan, Y.; Zhang, M.; Zhao, X.; Tian, B. Assessment of Construction Workers' Labor Intensity Based on Wearable Smartphone System. *J. Constr. Eng. Manag.* **2019**, *145*, 04019039. [[CrossRef](#)]
29. Fini, A.A.F.; Akbarnezhad, A.; Rashidi, T.H.; Waller, S.T. Enhancing the safety of construction crew by accounting for brain resource requirements of activities in job assignment. *Autom. Constr.* **2018**, *88*, 31–43. [[CrossRef](#)]
30. Jacobs, D.E.; Forst, L. Occupational Safety and Health and Healthy Housing: A Review of Opportunities and Challenges. *J. Public Health Manag. Pr.* **2017**, *23*, e36–e45. [[CrossRef](#)]
31. Tang, S.; Roberts, D.; Golparvar-Fard, M. Human-object interaction recognition for automatic construction site safety inspection. *Autom. Constr.* **2020**, *120*, 103356. [[CrossRef](#)]
32. Li, X.; Yang, D.; Yuan, J.; Donkers, A.; Liu, X. BIM-enabled semantic web for automated safety checks in subway construction. *Autom. Constr.* **2022**, *141*, 104454. [[CrossRef](#)]
33. Koh, P.B.; Abas, N.H.; Deraman, R. Investigation on the Compliance of Occupational Safety and Health (OSH) Legislations among Contractors and Potential Interventions to Improve Construction Safety Performance. *Int. J. Sustain. Constr. Eng. Technol.* **2022**, *13*, 293–299. [[CrossRef](#)]
34. Zheng, X. Intelligent framework for multiple-attribute decision-making under probabilistic neutrosophic sets and its applications. *Int. J. Knowl.-Based Intell. Eng. Syst.* **2023**, *27*, 503–513. [[CrossRef](#)]
35. Zhang, Y.; Xing, X.; Antwi-Afari, M.F. Semantic IFC Data Model for Automatic Safety Risk Identification in Deep Excavation Projects. *Appl. Sci.* **2021**, *11*, 9958. [[CrossRef](#)]
36. Jiang, S.; Feng, X.; Zhang, B.; Shi, J. Semantic enrichment for BIM: Enabling technologies and applications. *Adv. Eng. Inform.* **2023**, *56*, 101961. [[CrossRef](#)]
37. Boje, C.; Guerriero, A.; Kubicki, S.; Rezguy, Y. Towards a semantic Construction Digital Twin: Directions for future research. *Autom. Constr.* **2020**, *114*, 103179. [[CrossRef](#)]
38. Wang, H.; Xu, S.; Cui, D.; Xu, H.; Luo, H. Information Integration of Regulation Texts and Tables for Automated Construction Safety Knowledge Mapping. *J. Constr. Eng. Manag.* **2024**, *150*, 04024034. [[CrossRef](#)]
39. Pandithawatta, S.; Ahn, S.; Rameezdeen, R.; Chow, C.W.K.; Gorjian, N. Systematic Literature Review on Knowledge-Driven Approaches for Construction Safety Analysis and Accident Prevention. *Buildings* **2024**, *14*, 3403. [[CrossRef](#)]

40. Li, H.; Yang, R.; Xu, S.; Xiao, Y.; Zhao, H. Intelligent Checking Method for Construction Schemes via Fusion of Knowledge Graph and Large Language Models. *Buildings* **2024**, *14*, 2502. [[CrossRef](#)]
41. Erfani, A.; Cui, Q. Natural Language Processing Application in Construction Domain: An Integrative Review and Algorithms Comparison. In Proceedings of the International Conference on Computing in Civil Engineering (I3CE), Orlando, FL, USA, 12–14 September 2022; pp. 26–33.
42. Stepien, M.; Jodehl, A.; Vonthron, A.; Koenig, M.; Thewes, M. An approach for cross-data querying and spatial reasoning of tunnel alignments. *Adv. Eng. Inform.* **2022**, *54*, 101728. [[CrossRef](#)]
43. Hao, J.; Zhao, Q.; Yan, Y.; Wang, G. A Review of Tacit Knowledge: Current Situation and the Direction to Go. *Int. J. Softw. Eng. Knowl. Eng.* **2017**, *27*, 727–748. [[CrossRef](#)]
44. Huang, H.; Wu, H.; Wei, X.; Gao, Y.; Shi, S. Mapping sentences to concept transferred space for semantic textual similarity. *Knowl. Inf. Syst.* **2019**, *60*, 1353–1376. [[CrossRef](#)]
45. Rogers, C.S.; Wingfield, A. Stimulus-independent semantic bias misdirects word recognition in older adults. *J. Acoust. Soc. Am.* **2015**, *138*, EL26–EL30. [[CrossRef](#)]
46. Al Qady, M.; Kandil, A. Document Management in Construction: Practices and Opinions. *J. Constr. Eng. Manag.* **2013**, *139*, 06013002. [[CrossRef](#)]
47. Fu, M.; Qu, H.; Huang, L.; Lu, L. Bag of meta-words: A novel method to represent document for the sentiment classification. *Expert Syst. Appl.* **2018**, *113*, 33–43. [[CrossRef](#)]
48. De Clercq, D.; Wen, Z.; Song, Q. Innovation hotspots in food waste treatment, biogas, and anaerobic digestion technology: A natural language processing approach. *Sci. Total Environ.* **2019**, *673*, 402–413. [[CrossRef](#)]
49. Yi, J.; Yang, G.; Wan, J. Category Discrimination Based Feature Selection Algorithm in Chinese Text Classification. *J. Inf. Sci. Eng.* **2016**, *32*, 1145–1159.
50. Wang, Y. Research on the TF–IDF algorithm combined with semantics for automatic extraction of keywords from network news texts. *J. Intell. Syst.* **2024**, *33*, 20230300. [[CrossRef](#)]
51. Abedi, M.M.; Sacchi, E. A machine learning tool for collecting and analyzing subjective road safety data from Twitter. *Expert Syst. Appl.* **2024**, *240*, 122582. [[CrossRef](#)]
52. Zhang, J.; El-Gohary, N.M. Automated Information Transformation for Automated Regulatory Compliance Checking in Construction. *J. Comput. Civ. Eng.* **2015**, *29*, B4015001. [[CrossRef](#)]
53. Malekitabar, H.; Ardeshir, A.; Sebt, M.H.; Stouffs, R.; Teo, E.A.L. On the calculus of risk in construction projects: Contradictory theories and a rationalized approach. *Saf. Sci.* **2018**, *101*, 72–85. [[CrossRef](#)]
54. Lu, W.; Liu, Z.; Huang, Y.; Bu, Y.; Li, X.; Cheng, Q. How do authors select keywords? A preliminary study of author keyword selection behavior. *J. Informetr.* **2020**, *14*, 101066. [[CrossRef](#)]
55. Zhou, H.; Zhao, Y.; Shen, Q.; Yang, L.; Cai, H. Risk assessment and management via multi-source information fusion for undersea tunnel construction. *Autom. Constr.* **2020**, *111*, 103050. [[CrossRef](#)]
56. Wu, Y.; Lu, P. Comparative Analysis and Evaluation of Bridge Construction Risk with Multiple Intelligent Algorithms. *Math. Probl. Eng.* **2022**, *2022*, 2638273. [[CrossRef](#)]
57. Phu, V.N.; Chau, V.T.N.; Dat, N.D.; Tran, V.T.N.; Nguyen, T.A. A valences-totaling model for English sentiment classification. *Knowl. Inf. Syst.* **2017**, *53*, 579–636. [[CrossRef](#)]
58. Al Qady, M.; Kandil, A. Document Discourse for Managing Construction Project Documents. *J. Comput. Civ. Eng.* **2013**, *27*, 466–475. [[CrossRef](#)]

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