


## Article

# A Design for Safety (DFS) Semantic Framework Development Based on Natural Language Processing (NLP) for Automated Compliance Checking Using BIM: The Case of China

Yilun Zhou <sup>1,2</sup> , Jianjun She <sup>1,2,\*</sup>, Yixuan Huang <sup>1</sup>, Lingzhi Li <sup>1,2</sup>, Lei Zhang <sup>3</sup> and Jiashu Zhang <sup>4</sup>

- <sup>1</sup> College of Civil Engineering, Nanjing Tech University, Nanjing 210096, China; 201961126031@njtech.edu.cn (Y.Z.); hyx384286954@163.com (Y.H.); lilingzhi5566@njtech.edu.cn (L.L.)
- <sup>2</sup> Smart City Research Center, Nanjing Tech University, Nanjing 210096, China
- <sup>3</sup> Department of Engineering Management, School of Civil Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China; zhanglei-lzjtu@outlook.com
- <sup>4</sup> Department of Construction and Real Estate, School of Civil Engineering, Southeast University, Nanjing 210096, China; 230198129@seu.edu.cn
- \* Correspondence: shejj1016@126.com

**Abstract:** For design for safety (DFS), automated compliance checking methods have received extensive attention. Although many research efforts have indicated the potential of BIM and ontology for automated compliance checking, an efficient methodology is still required for the interoperability and semantic representation of data from different sources. Therefore, a natural language processing (NLP)-based semantic framework is proposed in this paper, which implements rules-based automated compliance checking for building information modeling (BIM) at the design stage. Semantic-rich information can be extracted from safety regulations by NLP methods, which were analyzed to generate conceptual classes and individuals of ontology and provide a corpus basis for rule classification. The data on BIM was extracted from Revit to a spreadsheet using the Dynamo tool and then mapped to the ontology using the Cellfie tool. The interoperability of different source data was well improved through the isomorphism of information in the framework of semantic integration, causing data processed by the semantic web rule language to be transformed from safety regulations to achieve the purpose that automated compliance checking is implemented in the design documents. The practicability and scientific feasibility of the proposed framework was verified through a 95.21% recall and a 90.63% precision in compliance checking of a case study in China. Compared with traditional compliance checking methods, the proposed framework had high efficiency, response speed, data interoperability, and interaction.

**Keywords:** building information modeling; design for safety; natural language processing; ontology; semantic web; automated compliance checking



**Citation:** Zhou, Y.; She, J.; Huang, Y.; Li, L.; Zhang, L.; Zhang, J. A Design for Safety (DFS) Semantic Framework Development Based on Natural Language Processing (NLP) for Automated Compliance Checking Using BIM: The Case of China. *Buildings* **2022**, *12*, 780. <https://doi.org/10.3390/buildings12060780>

Academic Editors: Yuan Chen, Hexu Liu, Xianfei Yin, Bo Xiao and Yinghua Shen

Received: 14 May 2022

Accepted: 4 June 2022

Published: 7 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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/>).

## 1. Introduction

The concept of design for safety (DFS) indicates that the design stage of a construction project is an important link in the project safety construction cycle, and engineering drawings or models produced in the design stage are the key documents used to guide subsequent construction. Noncompliance with the design documents will hurt the safety and quality of the construction project in the construction phase and the functions in the later operation phase, and it may even damage the safety of lives, property, and other public interests. At present, engineering designers, as the direct drafters of engineering drawings or models, often fail to foresee possible safety risks when carrying out design work, and problems with improper design and poor constructability are common. To control the quality of the design documents and maintain the bottom line of building safety, it is urgent to improve the safety risk control capability in the design stage.

Previously published data show that in recent years the quality level of design documents in China has been unsatisfactory, making it difficult to ensure their safety. The initial inspection pass rate of construction drawings is quite low, and the mandatory provisions of engineering construction standards are violated repeatedly, all of which indicate that the safety and compliance problems of design documents need to be effectively solved. An important reason for this phenomenon is that the current construction drawing inspection in China is still in the initial stage of digitalization, and it is difficult to avoid the problem of relying on a lot of manual repetitive work for the safety compliance inspection of design documents.

In solving this problem, BIM technology has shown great advantages in safety management due to the fact of its object-oriented modeling, digitization, visualization, and other characteristics. With the continuous introduction of policies and industry standards related to BIM technology, the popularity and application of BIM technology are also accelerating. Although the delivery method with the BIM model as the design result will gradually become the mainstream, the problem of data heterogeneity between BIM and DFS process limits the data interoperability system of automated compliance checking (ACC) based on computer technology. Many current research works aim to explore overcoming this difficulty, and some of them point to ontology technology. Ontology technology is an important technology for knowledge representation, knowledge management, knowledge sharing, and reuse. It has been widely used in coal mines, subways, medical care, and other fields. Ontology Web Language (OWL) [1] is a structured knowledge representation language, which can express the concept level clearly and accurately. Ontology provides a unified semantic basis that can be recognized by computers for information interaction in a domain, solving the problem of heterogeneous data interoperability.

Therefore, under the advocacy of developing intelligent drawing inspection, this paper proposes a DFS ontology framework based on semantic integration, which integrates the information from the building information model and the DFS process to realize the digitization of domain knowledge. DFS-related safety rules are used in NLP-based text mining to extract the conceptual classes contained in the framework for OWL ontology. Furthermore, BIM data are extracted and mapped to OWL ontology based on Dynamo-Cellfie methods to complement concept classes and generate individuals. In addition, the content that can grasp the regularity in the specification is transformed into the semantic web rule language, and the purpose of intelligent auxiliary inspection of the OWL ontology containing the semantic information of the design drawing is achieved by computer execution.

The rest of this paper is structured as follows. In Section 2, previous related research works are reviewed and their significance, gap, and reference value for this paper are explained. In Section 3, an ontology framework based on semantic integration is proposed, and methods for generating ontology concept classes and individuals based on BIM and NLP technologies and methods for transforming safety rules into computer-executable structured languages are introduced. In Section 4, a case study is applied to validate the proposed semantically integrated ontology framework. In Section 5, study conclusions are discussed, and study limitations and recommendations for future research are explored.

## 2. Related Works

This paper focuses on the application of BIM and semantic web technology in the safety management of the design phase. Previous related studies were reviewed as follows.

### 2.1. Design for Safety and BIM Using

The DFS concept was first developed in 1955 in the third edition of the Industrial Operation Accident Prevention Manual published by the US National Safety Council, which was the first to document instructions for considering the safe design from a designer's perspective. The hierarchy of project safety controls is highlighted and recommendations are made that designers should play an active role in eliminating hazards in construction projects [2].

The health and safety of the participants after the design stage of the construction project are considered in the design stage in advance, which is the biggest feature of DFS. The implementation of the DFS concept requires designers to strengthen their awareness of safety responsibilities, fully consider the constructability and safety compliance of the design results in the design stage, and start to identify the safety risks in the subsequent stages in advance. Effective measures need to be taken in the design stage to mitigate or eliminate potential hazards and to achieve the goal of safety management of construction projects.

In the practice of DFS, since Gambatese et al. [3] first loaded DFS recommendations into a computer database in 1997 and developed a design toolbox to ensure construction safety, the idea of assisting designers to identify the risks affecting construction safety in the design scheme and put forward design suggestions in the design stage has been widely referred to by subsequent researchers. Taking offshore platform and ship construction as examples, Wang et al. [4] described the characteristics of large-scale customized products, studied their design process, and a generic design for safety methodology was proposed. Weinstein et al. [5] established the relationship between specific safety risks and design suggestions, and completed risk assessments and scheme selections through risk identification of a design scheme of an actual subway project, verifying the feasibility of the method. Tanabe et al. [6] discussed a safe design approach for an onshore modular liquefied natural gas (LNG) liquefaction plant, where the project environment and specific design features of onshore and offshore power plants were identified and compared. Based on this comparison and consideration of modular features, a safety design approach for onshore modular LNG liquefaction plants was considered. These earlier studies aimed at revealing the inevitable link between design options and subsequent risks, and the potential of applying DFS practice to provide insightful information for safety-related problems, providing a solid foundation for this study.

The research that quantifies the relationship between design options and subsequent risks based on the DFS concept is also important. Sadeghi et al. [7] established a safety indicator used for the early design phase. The established safety indicator depends on two values that indicate the presence or absence of a hazard and the level of importance of the hazardous condition. These quantitative studies enhanced the scientific level of the DFS concept, providing the DFS concept with a wider range of applications. For example, Meacham [8] developed a sociotechnical system framework for performance-based design for fire safety (DFFS). The expansion of DFS application fields has greatly enriched the basic knowledge of DFS in different scenarios, which provides theoretical guidance for the combination of DFS concepts and modern information technology.

With the continuous development of BIM technology, integrating BIM technology has received more attention in the development of DFS tools [9]. Qi et al. [10] embedded the summarized safety recommendations in the BIM software through the secondary development form to achieve the purpose of identifying the unsafe state that may be caused by the construction personnel in the construction stage during the design stage. In addition, in a follow-up study [11], integrated BIM and third-party Solibri model checking software was used to develop a safety design tool to prevent falls from heights. However, the drawback of this approach is that due to the limitation of the development tools, only a few DFS rules can be integrated into the BIM software, which makes it difficult to obtain sound safety design recommendations comprehensively.

Hossain et al. [12] structured the DFS rules and divided them into six hierarchies to build a DFS rule knowledge base, and a safety risk intelligent inspection system integrating BIM and knowledge base was established, assisting designers to identify safety risks related to design elements for DFS. Tang et al. [13] developed an intelligent and safe design tool integrating the DFS method and BIM platform, which assisted subway station designers to apply the DFS concept in the design stage to mitigate emergency evacuation risks in the subway operation stage. Although these methods above strengthen the degree of computer execution of the DFS rules and lay a foundation for semantic understanding, it is

still difficult to solve the problem of low efficiency of rule execution caused by low data interoperability.

## 2.2. Semantic Web Technology and Automated Compliance Checking

Research on compliance inspection of design documents in the AEC field has transformed traditional design documents into digital design documents. Eastman et al. [14] proposed the four steps and the technology and framework of building design compliance checking, which is widely used for compliance checking research after object transformation.

A semantically richer building information model makes it possible to develop automated compliance checking systems for building regulations. The current research on automated compliance checking is difficult to implement independently of BIM. The Industry Foundation Class (IFC)-based format is considered a format well suited for automated compliance checking. Park et al. [15] proposed that the IFC improves the semantic interoperability of building information models, making information models easy to understand and sharing model data with heterogeneous computer systems. As a result, a set of rules-based systematic approaches have been developed to address the problems of building information representation based on the IFC of CAD objects, supplementary information extension of IFC models, interoperability, and semantic annotation of extensible information sets. Malsane et al. [16] analyzed residential fire safety-related building regulations and, based on this, explained a set of computer-implemented rules to implement automated compliance checking on semantically rich IFC models. Nevertheless, the IFC standard has a limited range of expression, which brings certain obstacles to data interaction.

Although semantic web technology is not commonly used in the AEC sector, it is of great value for integrating the complexity of BIM and heterogeneous data from different sources. Semantic web technology is considered to activate the potential of BIM, describing building regulations and BIM in the form of a hierarchical data structure through the ontology description language, OWL, to fully expand the range of semantic expressions of the model. Concept classes, relations, functions, axioms, and individuals can all be described in the form of hierarchical data structures through the ontology description language OWL [17]. Recently, many publications have recognized the importance of semantic web technology and BIM integration for automated compliance checking research.

Mohamed et al. [17] proposed an ontology system integrating BIM information and semantic web technology to enrich the semantics of existing building facilities. Zhong et al. [18] proposed an ontology framework integrating BIM and environmental monitoring to implement compliance checking. Dibley et al. [19] proposed the OntoFM, an ontology framework that supports real-time building monitoring. The development of the ontology framework lays the foundation for realizing information sharing and semantic interoperability. The ontology approach is also applicable to the development of an automated compliance checking framework incorporating the DFS concept. Jiang et al. [20] proposed a grey-box checking technique and a BIM-based automated code compliance check method to ensure the accuracy of design work that utilizes multi-ontology fusion. These studies all provided information for the development of the ontology framework and its underlying ontology in this paper.

The rules for automated compliance checking also need to contain rich semantic information to perform rule checking based on the interpretation of the ontology framework. Uhm et al. [21] applied context-free grammar (CFG) to natural language processing to perform computer interpretation processing on specification language, providing guidance and basic data for the development of a general automated design compliance check system. Zhang et al. [22] proposed methods based on semantic natural language processing techniques and expressive data-based techniques to automatically extract and transform regular language in canonical texts and to represent information based on semantic logic for fully automated reasoning. These efforts are a beneficial enhancement for automatic rule-based compliance checking and construction of semantic expressions, but the challenge

is that matching with the underlying ontology that integrates BIM and domain knowledge may be difficult due to the differences in the frameworks.

### *2.3. Natural Language Processing in the AEC Sector*

The natural language processing method has not yet been widely adopted in the AEC industry [23], but it has attracted the attention of researchers to some practical problems. For example, the use of text mining methods to capture key information in reports or regulations and documents in this text format are abundant in the AEC domain.

Existing research related to the application of NLP methods in the AEC field can be roughly divided into three categories: capturing requirements, extracting features, and acquiring knowledge. Hosseini et al. [24] implemented text mining on BIM-related natural language text in recruitment websites to extract keywords and co-occurrences, capturing the competencies and skills required by BIM workers. Zhou et al. [25] conducted a BIM app-related study in which the feedback, complaints, and expectations of users were extracted from natural language comments on BIM apps to describe their needs for BIM apps. These studies revealed the potential of NLP methods to support demand insights in the AEC field. Lin et al. [26] proposed an NLP-based method to capture key objects and their specifications in natural language-based sentences. Wang et al. [27] developed a question answering system for BIM information extraction using NLP methods, including natural language understanding, information extraction, and natural language generation modules. Compared to the completely unstructured natural language of users, applying feature extraction methods from semi-structured corpora requires simpler tools. Nonetheless, neither type of research is aimed at acquiring reusable domain knowledge.

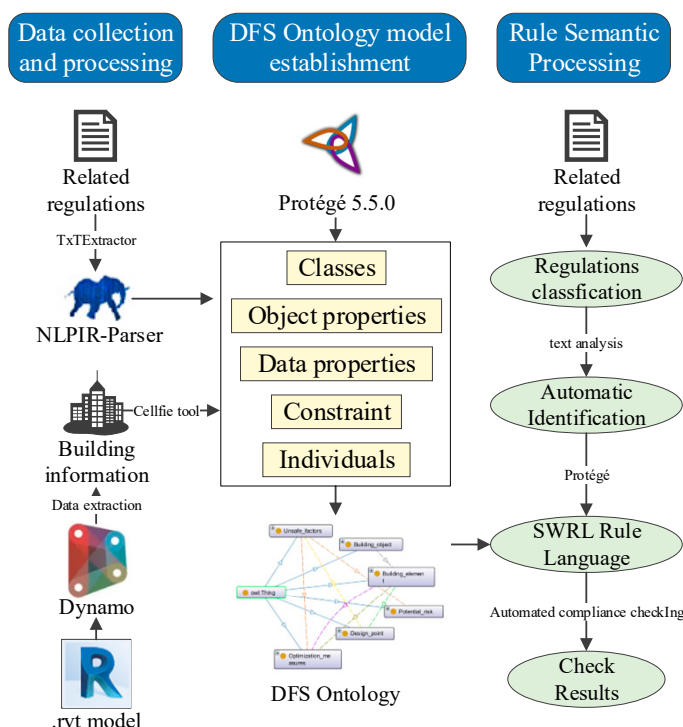
Current research focuses extensively on the use of NLP methods to acquire knowledge. Although many studies still use off-the-shelf tools, structured ordinance text-processing tools can characterize domain knowledge with appropriate refinements. Zhang et al. [28] proposed a semantic, rule-based NLP method for automatic information extraction from construction specification documents. Al Qady et al. [29] proposed a natural language processing system that was effectively used for contract text processing and adopted the technique for concept relation identification using shallow parsing (CRISP) to extract semantic knowledge from construction contract documents, improving electronic document management functions such as document classification and retrieval. The NLP-based knowledge acquisition methods proposed by these studies provide a reference for this paper, but emphasizing the importance of incorporating domain knowledge is still a problem that needs to be considered in this paper.

## **3. Methodology for the Proposed Framework**

### *3.1. Proposed Framework*

As illustrated in Figure 1, the proposed framework was split into three key modules, and the final output was the result of ACC which proposed safety recommendations for the design scheme. The proposed framework drew on a classic data mining process advocated by Shearer [30] and incorporated safety regulation analyses of the AEC sector, building information modeling technology, and semantic-based ontology modeling to improve the process. First, raw data were obtained from safety regulations and BIM models. The regulatory text was transformed into the .txt format, which can be processed by NLP through TxTExtractor. Then, the regulatory text was processed with NLP to obtain a user dictionary based on the new word corpus and the Chinese word-segmented regulations. As for the BIM, after the .rvt model was preprocessed in the Revit software, the data could be extracted through Dynamo to obtain the building information in a table format. Second, the results of text processing are used to create classes, properties, and constraints in the DFS ontology, and the table containing building information is converted into individuals in DFS ontology through the Cellfie tool. Third, safety regulations were divided into three types, and the linguistic characteristics of these different types were analyzed, through which different types of safety regulations could be extracted. At last, the safety regulations

were transformed into the SWRL language for a semantic query to realize the automated compliance checking of the DFS ontology.



**Figure 1.** Proposed framework for the development of design for safety (DFS) ontology and automated compliance checking (ACC) rules based on natural language processing (NLP).

### 3.2. Data Collection and Processing for DFS Ontology

This section, first, illustrates a method for Chinese word segmentation of safety regulation text through NLPIR. The result of segmentation contained a list of DFS specification words, according to which a user dictionary containing AEC professional words was created. Therefore, professional vocabulary tags were displayed after the words in the word segmentation results of the text. Such annotated words were considered key vocabularies to describe the DFS ontology, and they were integrated and analyzed to create the classes, properties, and constraints of the DFS ontology. Second, a method for preprocessing the .rvt model using Revit software was provided and building information was extracted from the processed model by the Dynamo tool. Building information was used to create the individuals of the DFS ontology, because it reflects the actual condition of a specific building component.

#### 3.2.1. DFS Regulations Collection and Text Mining

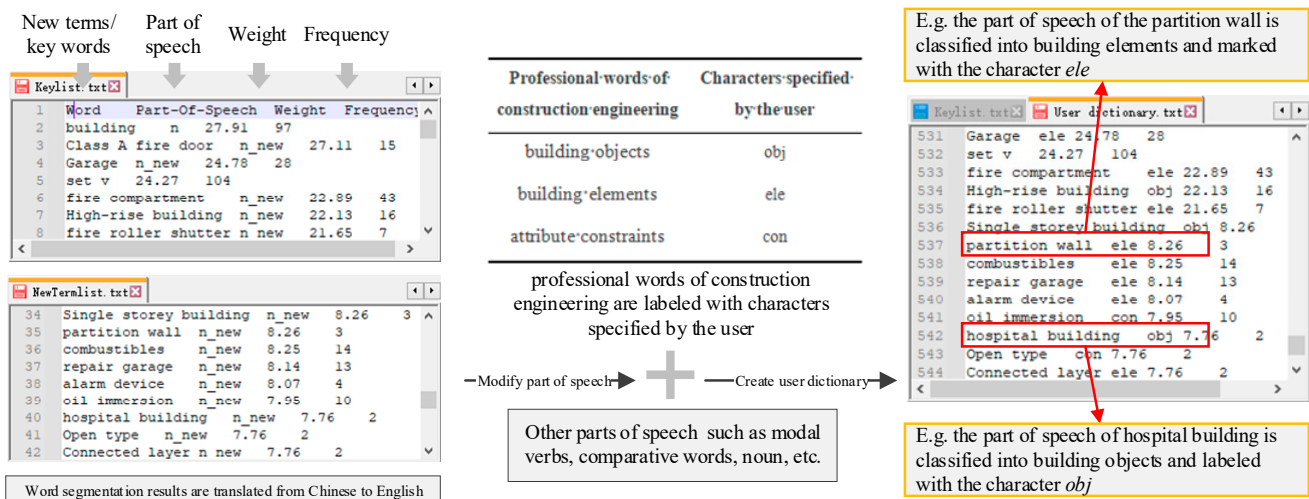
DFS is a systematic project that requires a continuous improvement process. Taking China as an example, the existing laws, regulations, and standards have made certain requirements for the safety of design work. Although there are some shortcomings, certainly, the requirements of these laws, regulations, and standards and regulations made by the State Council and industry departments are formulated according to the current national conditions of China. Laws and regulations determine the order of the construction market, which also determines the need to adhere to the DFS principle of construction projects.

Therefore, the content of safety regulations needs to be analyzed and classified in preparation for the conversion of unstructured text regulations in natural language into structured language. There are many types of existing construction safety regulations, among which, Compulsory Provisions of Engineering Construction Standards (Housing

Construction Part) is a universal standard that can cover 11 aspects of design work. This section mainly analyzes the design-related content of the regulation.

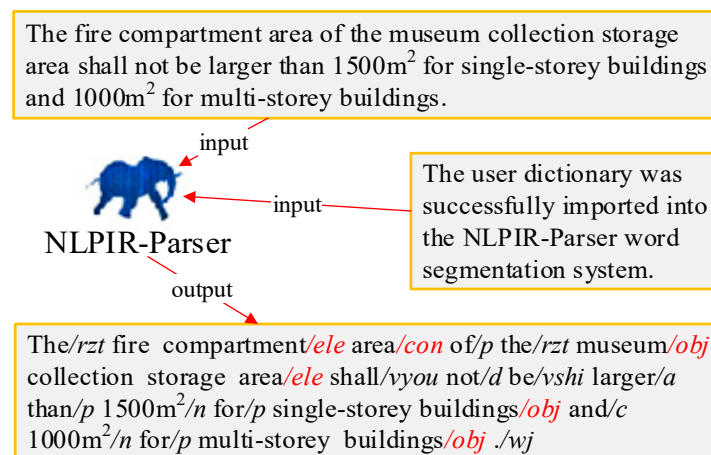
The NLPIR Chinese word segmentation system is used in this section to segment and tag the building design regulation text. The Chinese lexical analysis system NLPIR was developed by the Institute of Computing Technology, Chinese Academy of Sciences. The main functions of the system include Chinese word segmentation, part-of-speech tagging, and keyword capture. According to recent research [31], the 973 expert group evaluation results show that the correct rate of word segmentation was as high as 97.58%, and the processing speed of word segmentation and part-of-speech tagging was 31.5 KB/s; therefore, the tool had high processing efficiency in word segmentation and part-of-speech tagging. However, the NLPIR tool was difficult for AEC professional vocabulary recognition. Therefore, the new word list generated after rule text processing was used to add a user vocabulary corpus to form a user dictionary containing AEC professional vocabulary and specified annotation to segment the rule text according to the part of speech and the specified annotation of the user dictionary.

Figure 2 illustrates the process of user dictionary creation, and the results of word segmentation and part-of-speech tagging of regulatory text based on the user dictionary. Taking the chapters of building layout and fire zoning as an example, NewTermlist containing 49 new nouns and Keylist containing 497 keywords were generated in NLPIR's new word discovery function. The part-of-speech of each word was checked, and words determined to be professional words of construction engineering were labeled with characters specified by the user. Finally, the parts of speech used to label professional words of construction engineering were summarized into three types: building objects, building elements, and attribute constraints. Other parts of speech, such as modal verbs and comparative words, can be automatically tagged to the corresponding words by NLPIR without adding them to the user dictionary.



**Figure 2.** User dictionary creation based on NewTermlist and Keylist.

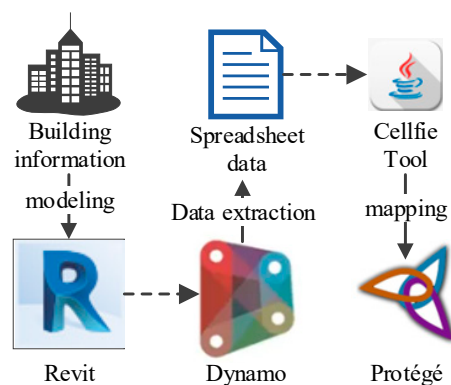
The user dictionary was imported in the word segmentation function interface of NLPIR, according to which the system processes the input regulation text. As shown in Figure 3, the rules text was segmented according to the part of speech and the professional characters specified by the user.



**Figure 3.** Text segmentation using the user dictionary.

### 3.2.2. Data Interoperability and Information Extraction of BIM

At present, no plug-ins have been developed to directly convert model data into the OWL format; therefore, data interoperability was the key to completing the inference, which required preprocessing of BIM. The model information extraction and conversion methods proposed by the current research efforts were not perfect. The information interaction based on IFC standards has problems, such as lack of logic and limited-expression range, and the business model checker was limited to specific rules checking, resulting in poor scalability and low knowledge utilization of building information. There are several ways of outputting information that can be applied to Revit in which project parameters can be extracted to Excel spreadsheets through Dynamo tools. Moreover, the Cellfie tool can directly convert tabular data into the OWL language and map it to the existing ontology to realize semi-automatic data structure conversion, which provides a feasible method for data interoperability between modeling software and ontology application software. Figure 4 provides the process of realizing data interoperability, where the solid line represents the transfer of data, and the dashed line represents the transition of the software scene.



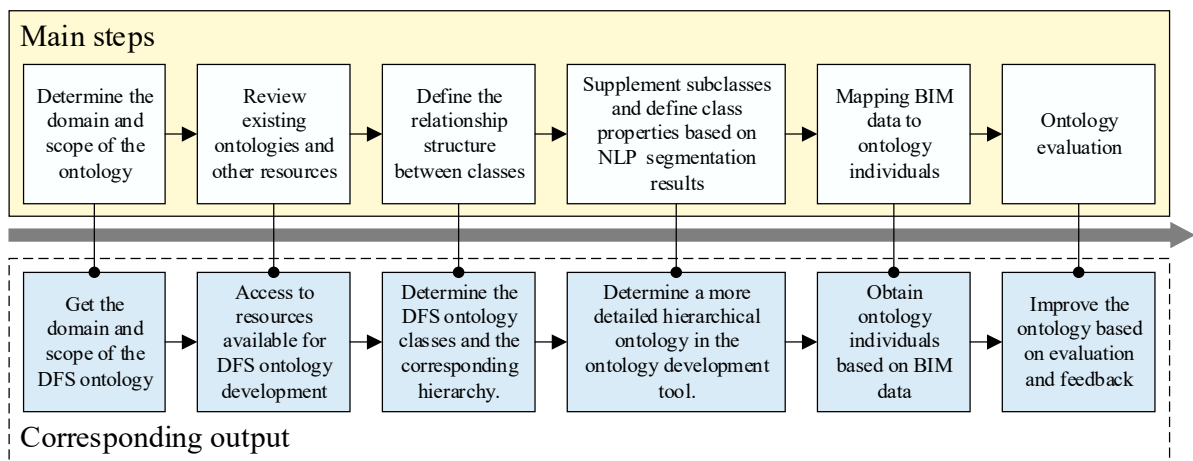
**Figure 4.** Implementation path of data interoperability.

The information of the preprocessed model was extracted in batches by the Dynamo visual programming tool. The Dynamo toolbar contains a series of code blocks, which can quickly process files and access commands. The button on the far right exports a snapshot of the workspace, which is important for documentation and sharing. Dynamo's basic function input can be accessed through the left search bar or selected from a loaded library of functions, and nodes are organized hierarchically in libraries, categories, and subcategories according to whether the node creates data, performs an action, or queries data. The loaded function library includes very rich code content and can also satisfy the operation performed on the data such as mathematical or geometric transformation.



### 3.2.3. DFS Ontology Development and Mapping

In the field of construction engineering safety, ontology theory has been researched and practiced to a certain extent, but there is still no unified standard to regulate the development of DFS ontology. Various scholars have put forward different ontology development principles according to their research. Influenced by these research efforts, the DFS ontology development method incorporating BIM technology and NLP technology is demonstrated in Figure 5.

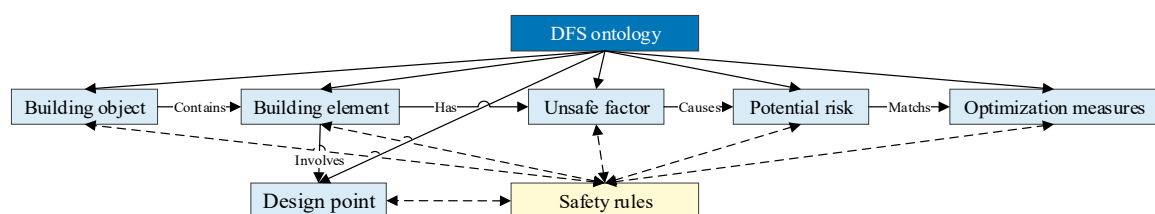


**Figure 5.** Seven steps to ontology development incorporating BIM and NLP.

The development of the DFS ontology should express the safety knowledge related to the design as much as possible and assist the designers to conduct safety compliance checks to meet the needs of safety design. Therefore, focusing on the field of construction safety design, the safety knowledge of DFS ontology development is obtained from safety regulations, project practice manuals, and other texts.

Reviewing the existing research efforts related to the development of building DFS ontology, the building safety ontology model proposed by Zhang et al. [32] and the building safety automated compliance checking ontology proposed by Huang [33] have important reference values for the DFS ontology developed in this section. Based on previous research, the characteristics of NLP technology and BIM technology are combined into the ontology proposed in this section.

The user dictionary obtained by NLP processing of the text of safety regulations contains key terms for constructing the compliance inspection process from the perspective of DFS, where building objects, building elements, design points, unsafe factors, potential risk, and optimization measures are extracted to constitute the main classes of DFS ontology and the object properties or relationships among them as explained in Figure 6.



**Figure 6.** The main classes and their connections in DFS ontology.

The word segmentation result of NLP on the text of safety regulations is naturally a subclass of the above main class of ontology. The classification standard of BIM components also serves as the basis for expanding the subclasses of building-related ontology. Correspondingly, building objects can be divided into subclasses according to function, height,

and fire resistance grade. For example, building objects can be divided into public buildings and civil buildings according to functions. The hierarchical structure of building elements mainly refers to the IfcBuildingElement classification standard. For example, building elements include stairs and ramps and structural elements include beams and columns. The design points are subdivided into subclasses such as fire protection design, structural design and selection, and seismic design. According to the summary of the description of different unsafe factors in the design content, the unsafe factors are divided into three subclasses: attribute constraints, element settings, and spatial relationships. The attribute constraint subclass included geometric properties, physical properties, and material properties, the element setting class was composed of measures taken, supporting settings, and other subclasses, and the spatial relationship class was composed of spatial distance and spatial position. Potential hazards were classified according to the most important and common types of hazards, namely, falls from heights, object strikes, fires, and collapses. The classes of optimization measures were initially divided into subclasses of optimization measures and safety measures, and further refinement could be divided into the adjustment of the overall layout, the adjustment of structural form, the addition of protection systems and safety training, etc., which could be further expanded in combination with expert opinions. The expanded and refined hierarchical ontology is presented in Figure 7.

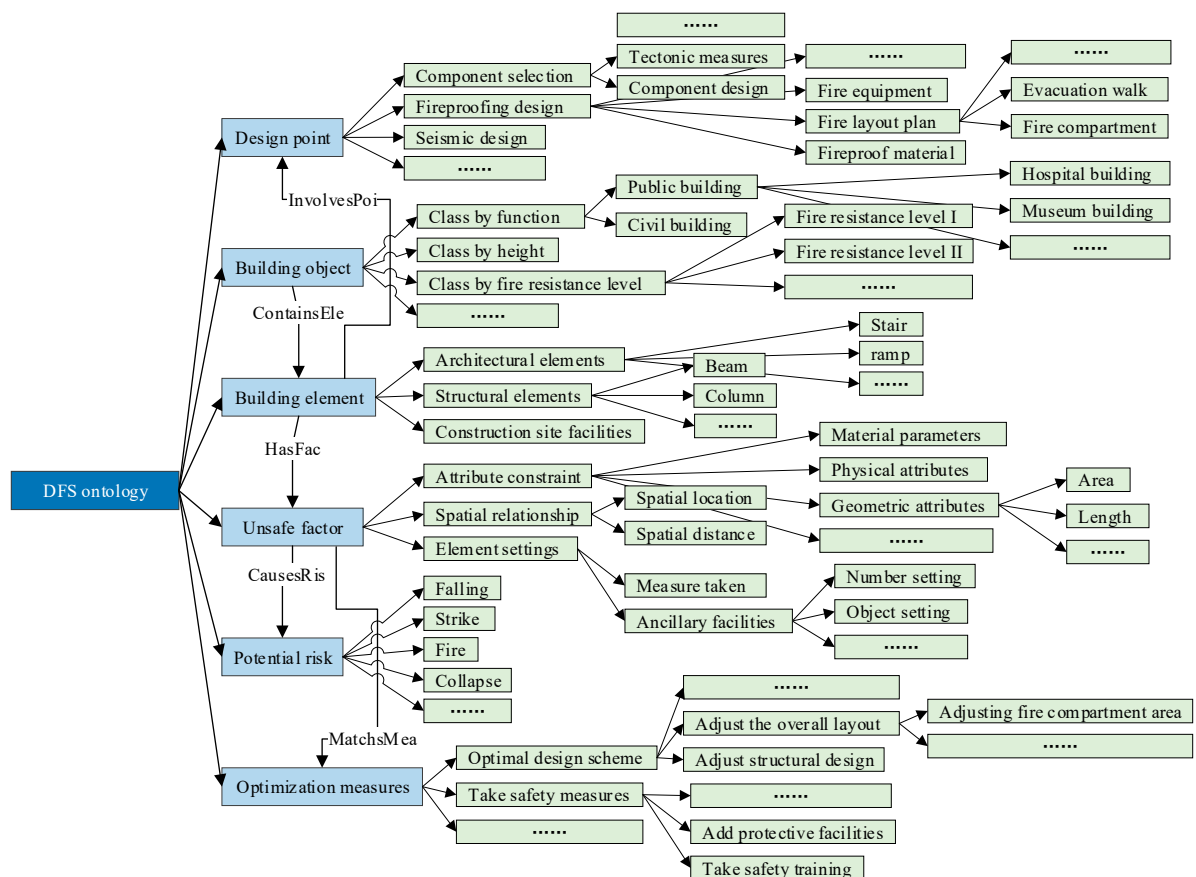
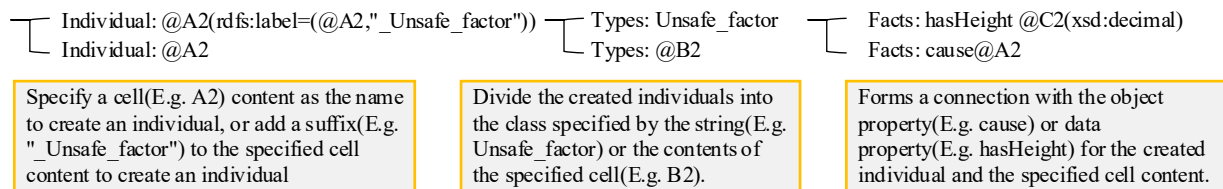


Figure 7. Generation of the hierarchical DFS ontology.

Properties in ontology include object properties and data properties. Object properties exist in the relationship between two classes and two individuals, while data properties describe the relationship between classes or individuals and values. Although the relationship between the main classes was defined in the previous steps, with the change in the range and domain of ontology properties [34], the object properties and data properties also need to be expanded to improve the connection of subclasses in the ontology. For example, the subclass room of the building element was connected to the subclass area of

the unsafe factor by the object property hasArea, while it was not connected to the subclass weight of the class Unsafe\_factor. The subclass Area of the class Unsafe\_factor had the data property hasArea and the domain was the area value.

The mapping of BIM data to ontology instances was realized by the built-in Cellfie tool, which needs to use transformation rules to perform the semantic conversion of tables to OWL ontology. This grammatical rule follows MappingMaster DSL [35]. MappingMaster is a domain-specific language (DSL) that defines the content of spreadsheets to OWL ontology. This section implements the mapping of BIM model information to DFS ontology by constructing data conversion rules by editing the transformation axioms on the transformation rules on the Cellfie module interface and defining the unique identifier of the instance corresponding to each cell address in the Excel spreadsheet by defining the axioms. The cell content was connected to the ontology class and properties through types and facts. There are multiple data import modes based on transformation rules, which can satisfy the mapping of different cell objects. The transformation rule language statements mainly used in this paper are explained in Figure 8.



**Figure 8.** Transformation rule language statements of spreadsheet data to ontology.

### 3.3. Semantic Analysis for Automated Compliance Checking Rules

Since the judgment rules of a safety compliance check are structured language, the parts of the three types of safety knowledge, attribute constraint class, element setting class, and spatial relationship class that can be digitally converted were selected in this section to build safety compliance check rules.

The basic expressions of the selected three types of safety rules are shown in Table 1, based on which the language structure of the basic expressions of each type was analyzed. New parts of speech and their corresponding character annotations were added to the user dictionary, which cannot be directly converted into ontology classes but reflect the structural characteristics of basic expressions. Table 2 shows the basis for judging the three types of safety norm knowledge, which is based on the fact that different types of safety rules contain different types and numbers of parts of speech.

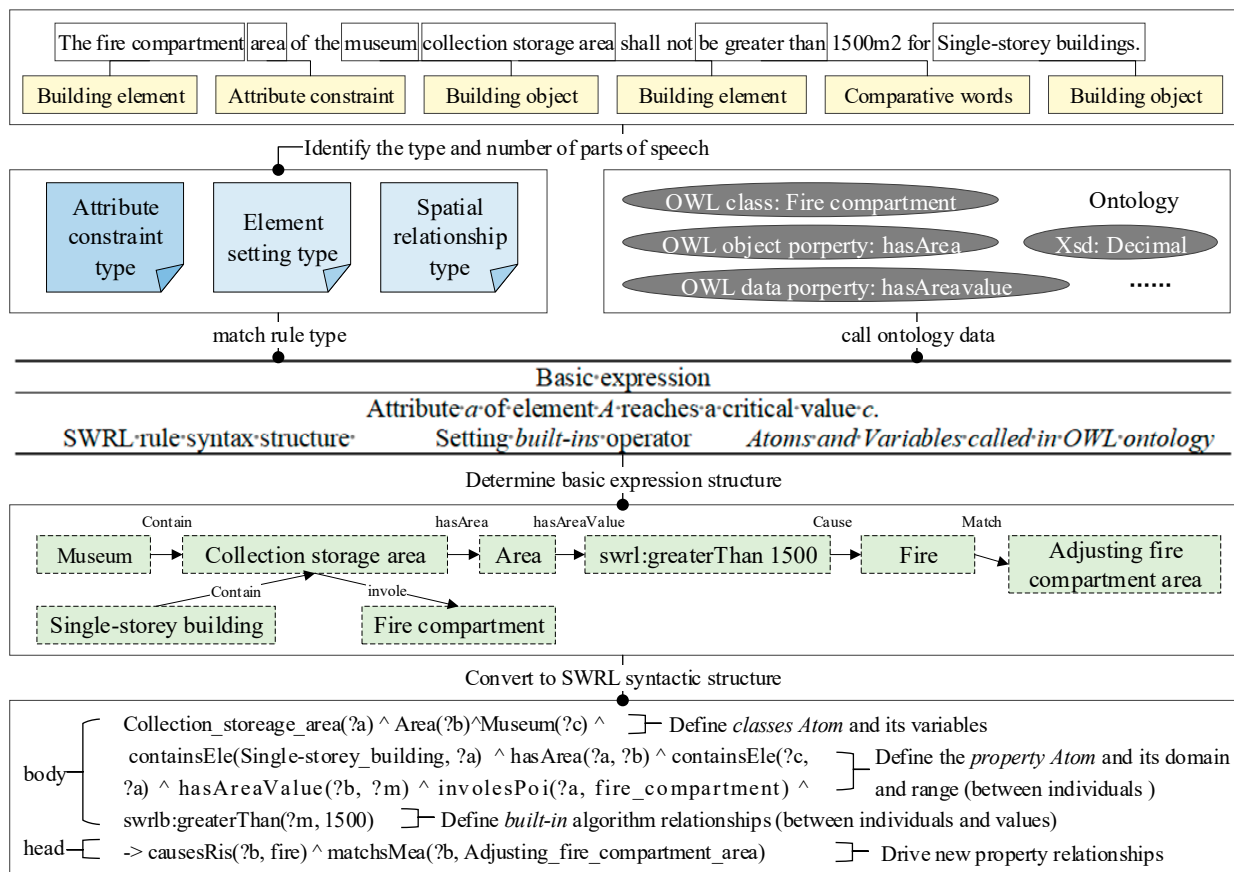
**Table 1.** Knowledge types of a building safety design code.

Type	Premise	Constraint	Basic Expression
Attribute constraint	Building element A	Has attribute a	Attribute a of element A reaches a critical value c
Element setting	Building element A	With building element B as an ancillary facility	Element A contains element B
	Building element A	Adopt safety measures a	Element A requires measure a
Spatial relationship	Building elements A and B	Spatial distance between A and B	The distance between element A and element B reaches a critical value c
	Building elements A and B	The relative position in the space between A and B	Element A and Element B have a spatial position relationship adjacent/on/below/contain/cover, etc.

**Table 2.** Judgment basis for knowledge types of safety regulations.

Type		Building Object /obj	Building Element /ele	Attribute Constraints /pro	Measures Taken /mea	Spatial Distance /spd	Spatial Location /spl	Comparative Words /com
Attribute constraint type		≥1	≥1	≥1	0	0	0	1
Element setting	Measures taken	≥1	≥1	0	1	0	0	0
	Ancillary facilities	≥1	≥2	0	0	0	0	0
Spatial relationship	Spatial distance	≥1	≥2	0	0	1	0	1
	Spatial location	≥1	≥2	0	0	0	1	1

Accordingly, the process of the structured processing of safety knowledge in this paper is summarized in Figure 9. The safety knowledge in natural language was transformed into the compliance checking rules in structured language that could be recognized by the computer.



**Figure 9.** Transforming safety regulations expressed in natural language into structured language.

Primarily, it consists of identifying the element part-of-speech type and quantification of the normative text, and then matching the three types of safety knowledge through the type and quantity of the part-of-speech. The structured expression of the regulation was obtained according to the basic expression corresponding to the safety knowledge type. SWRL (semantic web rule language) was used to construct the structured expression of DFS compliance check rules. It is a standard rule language developed by W3C, which can integrate DFS knowledge expressed in OWL ontology into rule statements. The SWRL syntax structure consists of an inference premise (body) and an inference result (head). The

body derives the head through the built-in logical comparison relationship of SWRL and is connected to the head by the symbol  $\rightarrow$ . A model-theoretic semantics is given on the W3C official website [36] to provide the formal meaning for OWL ontology including rules written in this abstract syntax, which is not detailed in this section.

## 4. Case Study

### 4.1. Case Study Description

To evaluate the feasibility of the framework in practical applications, a construction project, Southern New City Medical Center, undertaken by the Third Construction Co., Ltd. of the China State Construction Engineering Corporation (CSECC), Nanjing, China, was used as a case study in this section, whose construction area is over 310,000 square meters including outpatient buildings, emergency buildings, medical technology buildings, ward buildings, scientific research and administrative complexes, underground parking lots, and many other facilities. The corresponding building information model was created in Revit, where the model of the fourth floor of an inpatient building was used for the case study.

### 4.2. BIM Preprocessing and Information Extraction Module

Autodesk Revit software was used for the building information modeling and model preprocessing in this section by assigning values to the properties of the created building model and using the method of adding project parameters to complete the model information. The purpose of this step was to make the building component model contain the information needed to create the DFS ontology. The new project parameter was defined as a type parameter, and the component type to which it was attached was selected in the filter as illustrated in Figure 10. For example, the model type to which the property partition belongs was the room model. The model type to which the property fire resistance limit belongs was the wall, fire door, etc. The specific values of these properties were entered in the property column of the corresponding component model.

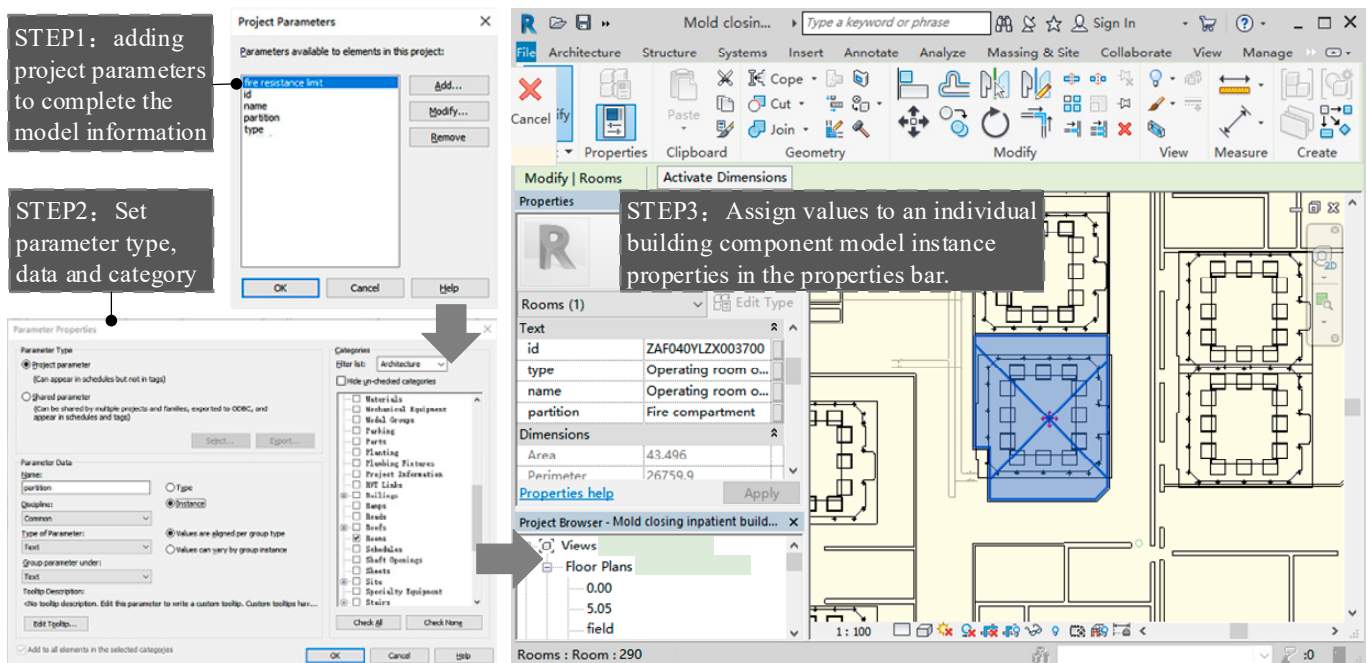


Figure 10. Process of BIM preprocessing in Revit.

The BIM data extraction visual programming language based on Dynamo was divided into three modules according to its functions:

(1) Module 1 was created to select the models that had information that needed to be extracted and to obtain their elementIDs and code block Select Model Elements;

(2) Module 2 was created to extract data for model elements by parameter name, including a code block that sorts the parameter name characters by column, and a corresponding number of code blocks `Element.GetParameterValueByName` and a code block `List.create`. The model element obtained by module 1 and the output result of the code block containing the parameter name characters were used as the input values of the code block `Element.GetParameterValueByName`, and then the `List.create` code block was connected to arrange the obtained parameter information;

(3) Module 3 was created to generate an Excel spreadsheet and write the BIM data extracted by the previous modules in the specified header sequence including a core code block `Excel.WriteToFile` and several code blocks connected to it. The header, file path, written data content, and layout of the spreadsheet were determined in this module.

Eventually, building information from the .rvt model in Revit was exported to an Excel spreadsheet by Dynamo, containing the information specified. As illustrated in Figure 11.

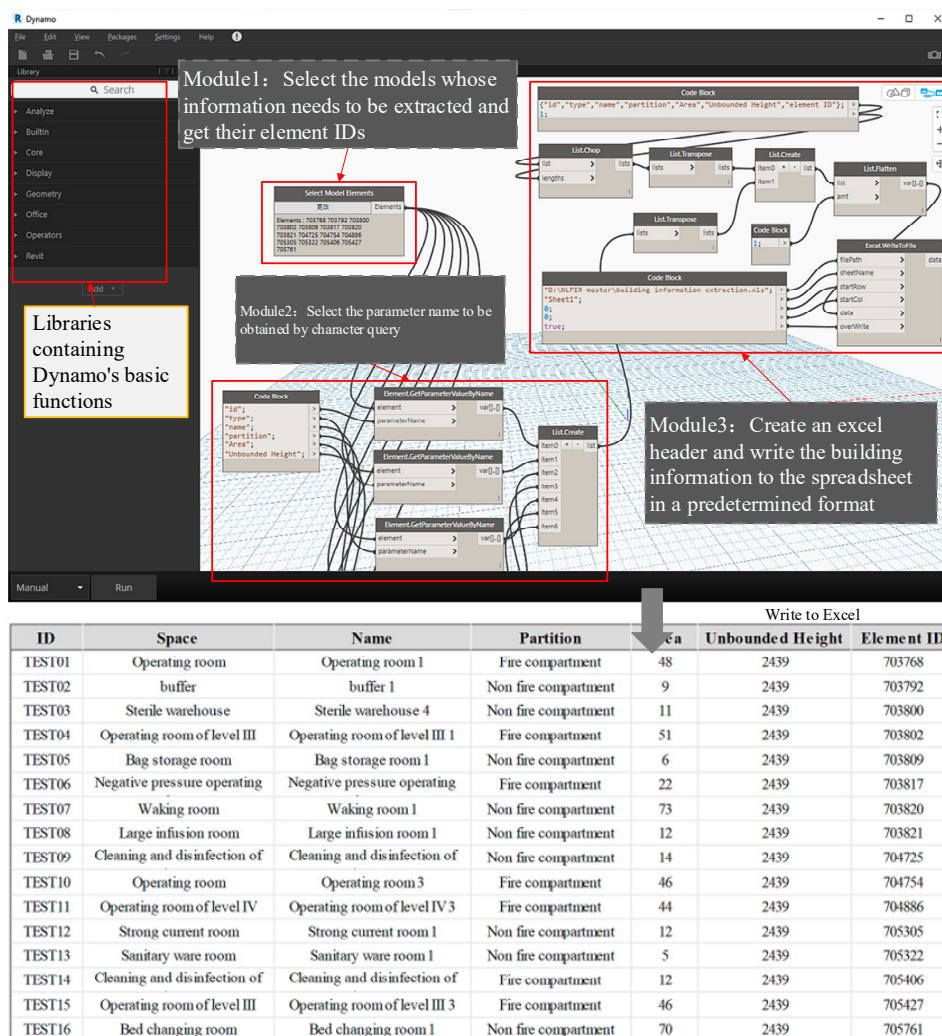


Figure 11. BIM data extraction via Dynamo.

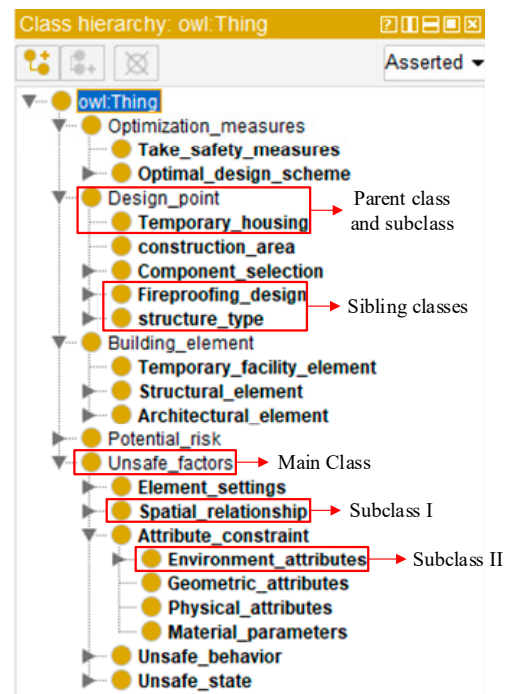
#### 4.3. DFS Ontology Development Module

The development tool of ontology was Protégé 5.5.0 developed by Stanford University School of Medicine, which has the advantages of being open source, easy to use, convenient for modification, and convenient for storage [37]. In this section, the development of DFS ontology is introduced, which was split into three main steps: (1) Define and enumerate the classes, subclasses, and sibling classes of the ontology, which are the sum of the concepts in the field with the same characteristics to describe the concepts in the ontology. (2)

Define properties and semantic relationships to express relationships between concepts. Properties in the ontology are descriptions of the characteristics of the classes, including object properties and data properties, and determine the classes that the properties act on by defining the domains and ranges. (3) Create individuals at the bottom of the hierarchical ontology, including those created directly in Protégé and those obtained by mapping data from BIM through Cellfie tools, which are the most basic parts of ontology.

#### 4.3.1. Classes of the DFS Hierarchical Ontology

As noted in Figure 12, the domain of the DFS ontology is to construct a building safety design hierarchy, limited in scope to the text mining results of the specified safety regulations. The main classes were the key components extracted from the relevant rules, and subclasses I and subclasses II were further subdivisions and extensions of the main classes based on the NLP results. According to the previous content, six main classes, including Design\_point, Building\_object, Building\_element, Unsafe\_factor, Potential\_risk, and Optimization\_measures, were created in the classes hierarchy of Protégé.



**Figure 12.** The class hierarchy of DFS ontology.

#### 4.3.2. Definition of Properties and Related Constraints

Properties in the ontology included object properties and data properties, which constrain the defined classes. Object properties exist by the relationship between two classes or instances, while data properties describe the relationship between classes or individuals and values. Set the domain to determine the subject of the property, and set the range to determine the object of the property in defining a property [35]. Defining properties is a key step in developing an ontology. It is not only related to whether the relationship between concepts in the field can be accurately described but also whether the logical reasoning of the ontology can be accurately realized.

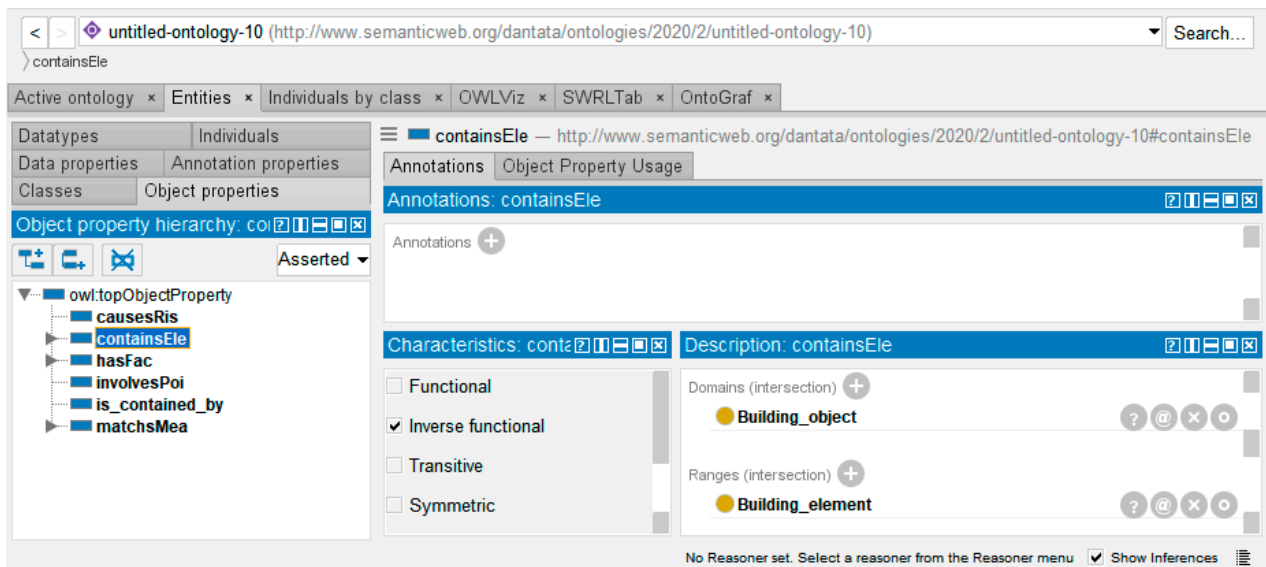
Moreover, as a mapping relationship linking domains and ranges, object properties have seven main properties that can affect the transitivity of properties, thereby affecting the inference engine to form previously unnoticed but legal connections between other classes. Therefore, in addition to the object property relationship is a connecting the subclass and the superclass, the five main object properties between the six main classes defined in this

section and their domains, ranges, and characteristics are illustrated in Table 3. The Protégé wiki [38] can be consulted for a detailed explanation of property characteristics.

**Table 3.** Definition of object properties of the DFS ontology.

Object Property	Domain	Range	Characteristic
containsEle	Building_object	Building_element	Inverse functional
hasFac	Building_element	Unsafe_factor	Inverse functional
causesRis	Unsafe_factor	Potential_risk	Inverse functional
matchesMea	Unsafe_factor	Optimization_measures	Inverse functional
involvesPoi	Building_element	Design_point	Inverse functional

An instance of object properties defined in Protégé is shown in Figure 13. The property containsEle has the inverse function characteristic, and the domain and range of it are defined as Building\_object and Building\_element in the corresponding tab.



**Figure 13.** Object properties of DFS ontology in Protégé.

A data property describes the relationship between a class or an individual and a value, of which the class is the domain, and the value is the range. The data types mainly used in this section included the xsd:integer type for describing integers, the xsd:decimal type for describing decimals, the xsd:string type for describing strings, and the xsd:boolean type for describing true or false. According to the previous NLP processing results of the safety regulation text, the attributes involved in digitization were defined as data properties, some of which are shown in Table 4.

**Table 4.** Definition of data properties of the DFS ontology.

Data Property	Domain	Range	Characteristic
hasEleID	Building_element	xsd:integer	Functional
hasHeight	Geometric_attributes	xsd:decimal	Functional
hasWidth	Geometric_attributes	xsd:decimal	Functional
hasLength	Geometric_attributes	xsd:decimal	Functional
hasAreaValue	Geometric_attributes	xsd:decimal	Functional
hasMaterial	Material_parameters	xsd:string	Functional
hasFireproofLevel	Fireproof_door	xsd:integer	Functional
...	...	...	...



According to the above definitions, the data properties of the DFS ontology classes were established in Protégé, an instance of which can be seen in Figure 14 below. The data property hasAreaValue has the function characteristic, and the domain and range of it are defined as Geometric\_attributes and xsd:decimal in the corresponding tab.

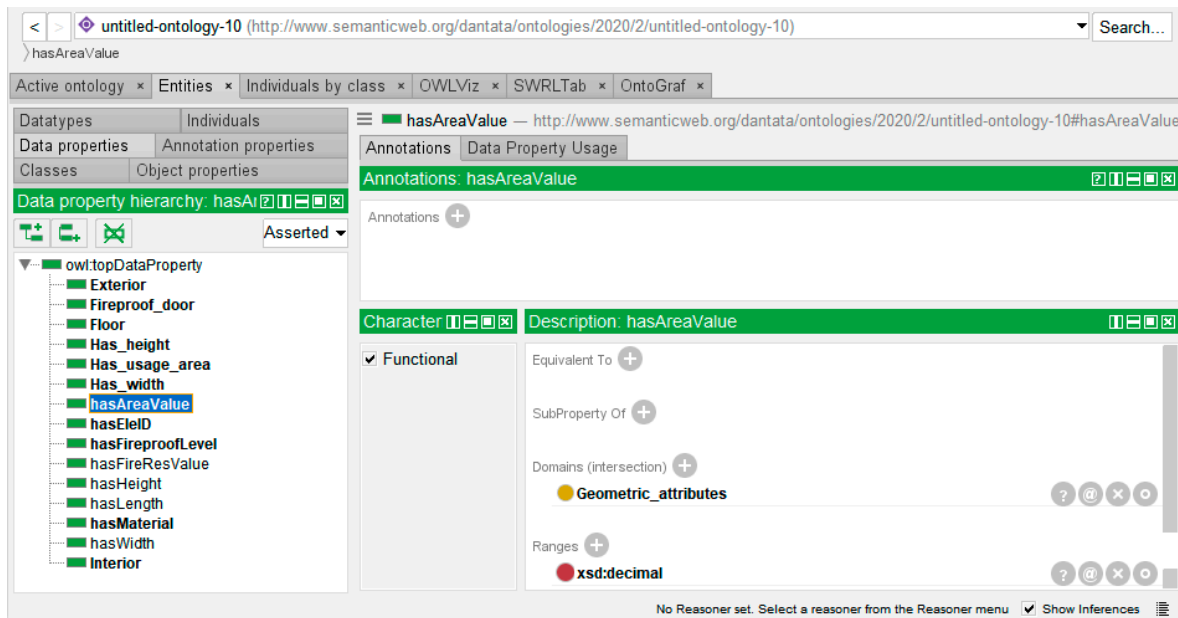


Figure 14. Data properties of the DFS ontology in Protégé.

The conceptual classes, individuals, and connections between them in DFS ontology are presented in a visual view, a part of which is shown in Figure 15. The types of arcs and nodes are illustrated on the right side of the figure.

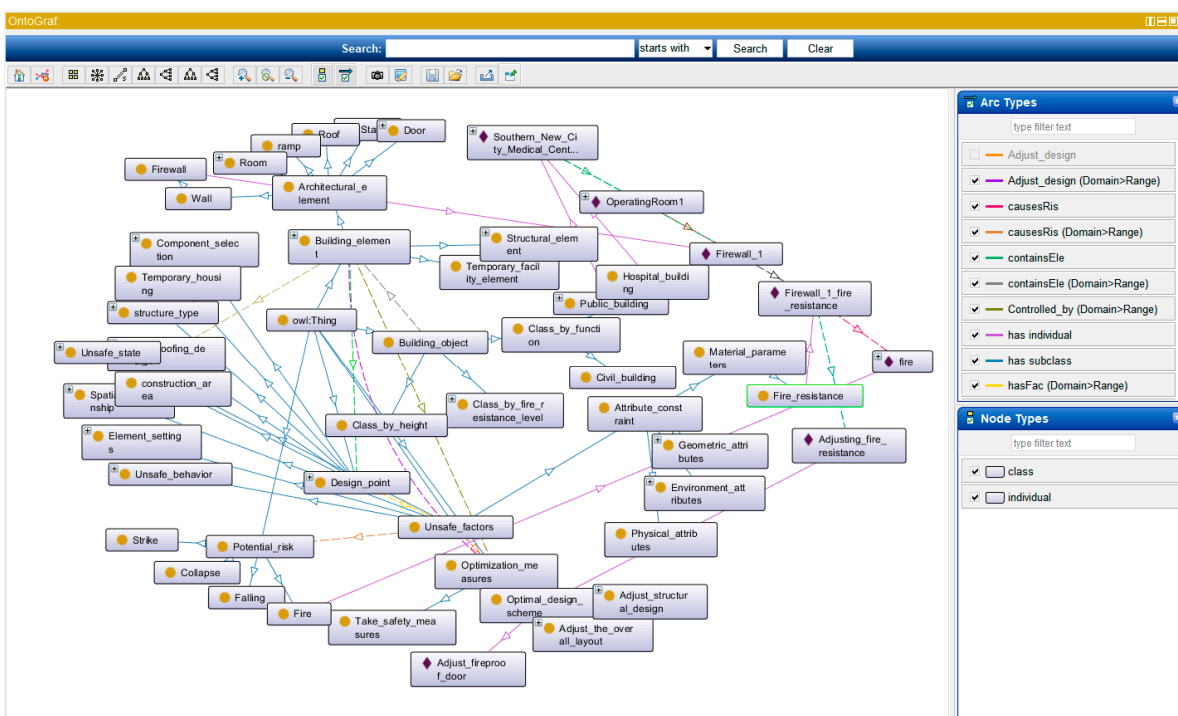


Figure 15. Ontology visualization in Ontograf of Protégé.

#### 4.3.3. Individuals Creation and BIM Data Mapping

In an ontology, individuals are concrete representations of classes that are indispensable parts of implementing ontology reasoning as inspection datasets. Furthermore, the properties defined by the DFS ontology have meaning only if they exist, depending on the individuals. Before the mapping step based on BIM data, some individuals of related concept classes need to be added in the individual by class of Protégé to form a complete inference chain. For example, individuals, such as fire and falling hazards, are the basic concepts derived from NLP-based text mining, but there is no corresponding information in BIM data. In addition, other information to be checked for safety compliance is mapped from the output data of the BIM project instance.

According to the MappingMaster DSL mentioned above, the language of mapping BIM data to ontology individuals is edited in the transformation rules editor of Protégé. The transformation rules used in this section can be generated as follows:

Create classes according to column B, which are subclasses of Room.

Class: @B\*

SubClassOf: Room

Create partition individuals named with the content in column D that belong to a class named Fire\_layout\_plan.

Individual: @D\*

Types: Fire\_layout\_plan

Create room individuals named with the content in column C, which are connected to elementIDs in column G through the data property hasEleID and connected to partitions in column D through the object property involvesPoi. The name of the classes to which these individuals belong is in column B. An individual named Southern\_New\_City\_Medical\_Center, of class Hospital\_building, was created previously, and is connected with the individuals above through the object property containsEle. Therefore, the room individuals above were connected with Southern\_New\_City\_Medical\_Center through object property is\_contained\_by (inverse property of containsEle).

Individual: @C\*

Types: @B\*

Facts: hasEleID@G\*(xsd:integer),

involvesPoi @D\*,

is\_contained\_by Southern\_New\_City\_Medical\_Center,

hasArea@C\*(rdfs:label = (@C\*,"\_Area"))

Create area individuals, names of which are contents in column C adding a suffix \_area. Define the value on column E as the data property value and connect them through the data property hasAreaValue. The name of the classes to which these individuals belong is Area.

Individual: @C\*(rdfs:label = (@C\*,"\_Area"))

Types: Area

Facts: hasAreaValue @E\*(xsd:decimal)

The edited conversion rules can be stored in the database. When reusing or updating instance data, rules can be edited. The individuals generated by the Cellfie tool are rendered in Figure 16.

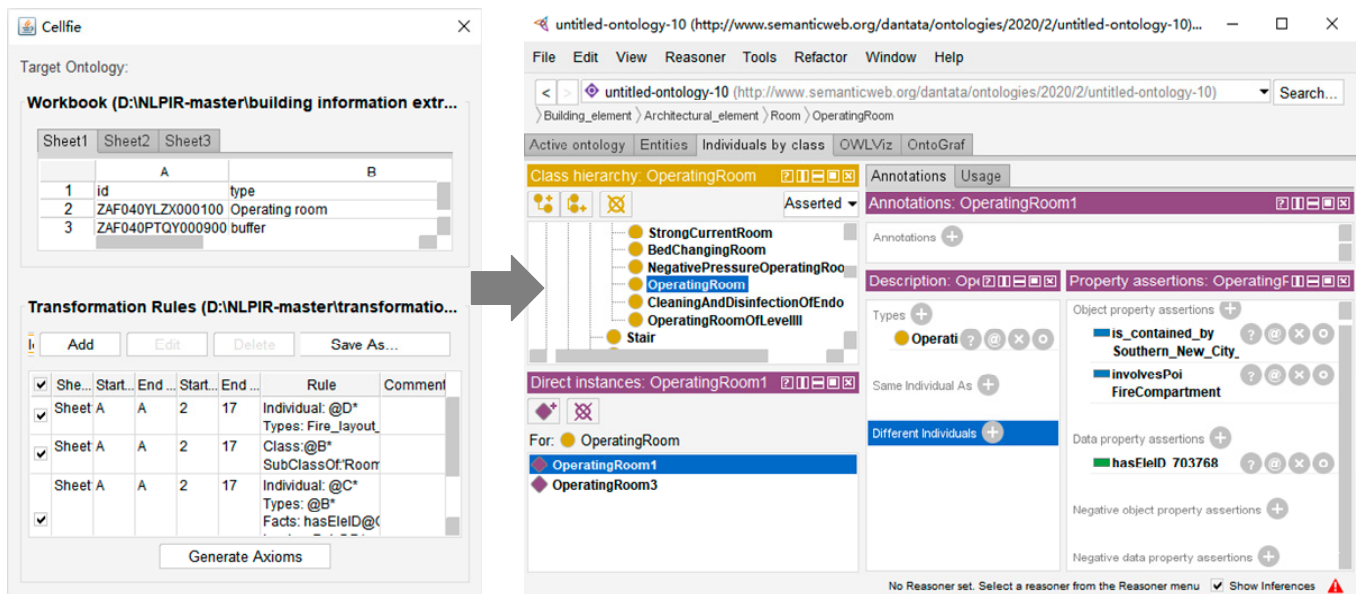


Figure 16. Mapping spreadsheet data to ontology individuals via Cellfie tools.

#### 4.4. Semantic Transformation Module

The editing of the rules in the SWRLTab of Protégé is completed through the method proposed previously. The SWRL statements used in this section to implement the safety compliance check on the ontology developed above can be generated as follows for example, wards, delivery rooms, operating departments, rooms for precision, and valuable medical equipment in the fire compartment should be separated from other parts by noncombustible bodies with a fire-resistance rating of no less than 1.0 h.

Rule 1:

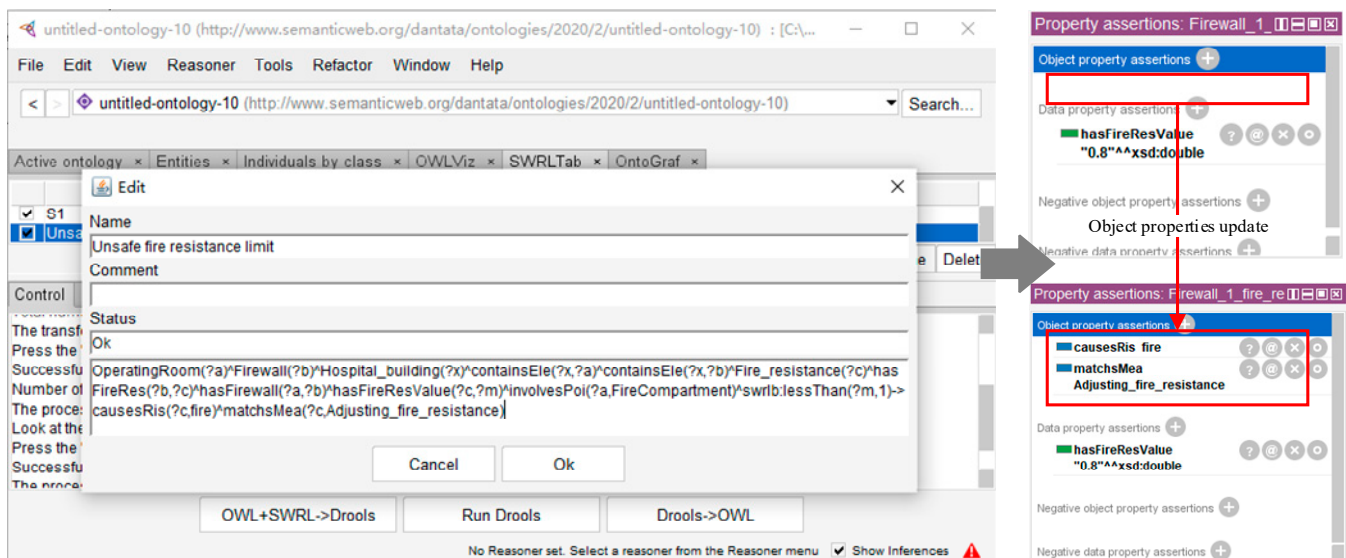
```
OperatingRoom(?a)^Firewall(?b)^Hospital_building(?x)^containsEle(?x,?a)^containsEle(?x,?b)^Fire_resistance(?c)^hasFireRes(?b,?c)^hasFirewall(?a,?b)^hasFireResValue(?c,?m)^involvesPoi(?a,FireCompartment)^swrlb:lessThan(?m,1)->causesRis(?c,fire)^matchsMea(?c,Adjusting_fire_resistance)
```

The usage area of the front room of the smoke-proof stairwell in public buildings should not be less than 6.0 m<sup>2</sup>.

Rule 2:

```
Public_building(?x)^Stairwell_Anteroom(?a)^Area(?c)^involvesPoi(?a,FireCompartment)^Smoke-proofStairwell(?b)^hasAntroom(?b,?a)^containsEle(?x,?a)^containsEle(?x,?b)^hasArea(?a,?c)^hasAreaValue(?c,?m)^swrlb:lessThan(?m,6)->causesRis(?c,fire)^matchsMea(?c,Adjust_usage_area)
```

The SWRL editing interface is presented in Figure 17. The Drools inference engine in Protégé is invoked, which is a rule inference engine based on description logic and supports the Java language and has good compatibility and inference functions. Run the inference on the engine interface, and the inference result is returned to the individuals of the ontology to update the DFS ontology. e.g., after rule 1 runs, the results fed back from Drools to the OWL ontology showed that two object properties were added to the Firewall\_1\_fire\_resistance individual.



**Figure 17.** Ontology reasoning of the Drools inference engine based on SWRL.

#### 4.5. Framework Testing and the Result

The NLP-based semantic framework proposed for DFS was tested with the above case. The SWRL sentences used for the test were the previously proposed Rule 1 and Rule 2, which were designed to identify noncombustible objects with noncompliant fire ratings in hospital buildings and fire stairwell front rooms in public buildings that do not meet the area value requirements. Based on the project parameters of Revit, design information related to compliance checking rules was added to the BIM model. Four information sets with 50 items were generated based on this design information, corresponding to the scenarios that were (1) compliant fire resistance with rule 1, (2) noncompliant fire resistance with rule 1, (3) compliant area value with rule 2, and (4) noncompliant area value with rule 2.

For compliance checking of these information sets, precision, recall, and the F1-measure were used to evaluate the NLP-based semantic framework. The set of items retrieved could be divided into related items and irrelevant items, and the set of related items in the database could be divided into retrieved items and non-retrieved items. The related items in the retrieved itemset are identical to the retrieved items in the related itemset in the database. The number of items in these information sets is listed in Table 5.

**Table 5.** The number of items in the information sets corresponding to the scenarios (1)–(4).

Information Sets for the Scenarios (1)–(4)	The Number of Items Contained in the Relevant Item Sets in the Database	The Number of Items Contained in the Retrieved Item Sets	The Number of Intersection Sets.
(1)	16	16	15
(2)	16	16	15
(3)	3	4	3
(4)	15	14	14

The ratio of the number of these items to the number of items contained in the retrieved item sets was defined as precision, and the ratio of these items to the number of items contained in the relevant item sets in the database was defined as recall. The F1-measure was defined as the harmonic mean of precision and recall. The specific interpretation of these indicators and the test results based on the data in Table 5 were summarized in Figure 18. The test results indicate that the precision, recall, and the F1-measure of the NLP-based semantic framework for DFS were 95.21%, 90.63%, and 92.44%.

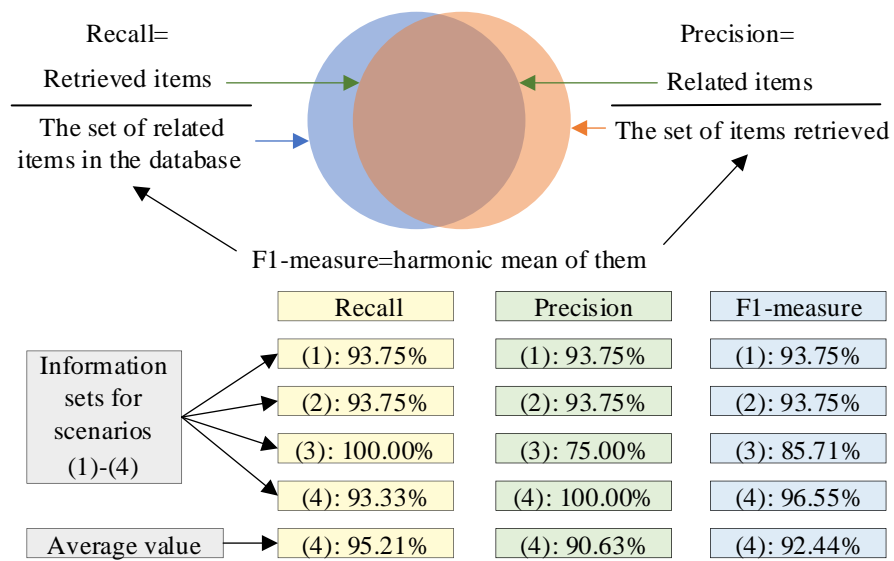


Figure 18. The specific interpretation recall, precision, and the F1-measure and the test results.

An error analysis was conducted to determine the reasons for compliance check deviations in the framework tests. Due to the limitations of semi-automatic data isomorphism technology, information errors may be caused by incorrect operations in the process of information transformation. Figure 19 provides an excerpt from the error analysis table, and errors in TEST24 were caused by the room of cleaning and disinfection of the endoscope being erroneously identified as a non-fire compartment, while TEST30 and TEST33 were caused by improper manipulation during data entry. Furthermore, recall was more significant than precision for automated compliance checking frameworks. This was because the noncompliance in the design work should be identified as much as possible, while precision deviation can be improved using analysis and elimination. The proposed framework has higher recall than precision, which indicates the efficiency and applicability of the framework for automated compliance checking.

TEST24	Cleaning and disinfection of endoscope	Cleaning and disinfection of endoscope 3	Fire compartment	0.8	Misidentified as a non-fire compartment
TEST25	Cleaning and disinfection of endoscope	Cleaning and disinfection of endoscope 4	Fire compartment	1.5	
TEST26	Strong current room	Strong current room 4	Non-fire compartment	1.2	
TEST27	Strong current room	Strong current room 5	Non-fire compartment	1.2	
TEST28	Strong current room	Strong current room 6	Non-fire compartment	1.2	
TEST29	Sanitary ware room	Sanitary ware room 3	Non-fire compartment	1.2	
TEST30	Operating room	Operating room 7	Fire compartment	0.8	Incorrectly changed when entering information into BIM
TEST31	Operating room	Operating room 8	Fire compartment	0.8	
TEST32	Operating room	Operating room 9	Fire compartment	0.8	
TEST33	Anteroom of smoke-proof stairwell	Anteroom 1	Fire compartment	5.5	Incorrectly changed when entering information into BIM

Figure 19. An excerpt from the detection error analysis table.

The results of checking the safety compliance of building design documents with the above developed SWRL statement were compared with the database-based information retrievals and the traditional and manual information contradiction methods. The comparison results are illustrated in Table 6.

**Table 6.** A comparison of compliance checking methods in different items.

Comparison Items	The NLP-Based Semantic Framework	The Information Retrievals Based on Database	The Manual Contradistinction Measures
Types of compliance detection for DFS	According to Table 1, the proposed framework can identify noncompliance based on 5 digitized rule types, covering most of the compliance checking regulations	Due to the highly structured retrieval language, it is possible to retrieve according to the rules of direct digitization, but it is difficult to execute the retrieval rules of indirect digitization	Noncompliance can be identified by almost all of the compliance detection regulations based on manual judgment
Detection speed and efficiency.	Batch processing of ontology data, and has a quite fast detection response speed and high detection efficiency	Batch processing of spreadsheet data, and has a quite fast retrieval response speed and high retrieval efficiency	Processing building design document inspection items one by one, and has a slow detection response speed and low detection efficiency
The safety performance improvement	The proposed semantic framework can understand noncompliance in design documents and match appropriate solutions based on the SWRL syntax and logical connections among ontology concept classes formed through properties to improve the safety of the design document	Although the solution can be matched to the noncompliance of building design through SQL-based retrievals, the relationship between noncompliance and resolution measures is difficult to understand by the database, which may lead to biased safety performance improvement measures.	Inspectors can accurately understand the meaning of noncompliance in design documents and obtain solutions according to regulations to improve safety performance.

To sum up, the NLP-based semantic framework combined the advantages of computer retrieval technology and human recognition. The database-based automatic compliance checking method had high information processing efficiency, but the checking results may be biased due to the limitations of SQL retrieval syntax and methods. For example, this approach failed to identify building types based on building characteristics and, thus, failed to provide safety performance enhancements for hospital building design. Rules-based manual contradistinction compliance checking method was considered to have high precision but low recall and efficiency. The NLP-based semantic framework not only had the corresponding speed and efficiency of database-based computer retrieval technology but also understood the design documents similar to how the human brain processes information; thus, it identified more types of noncompliance and provided appropriate solutions to improve safety performance via the semantic web. The comparison indicated that the method based on semantic knowledge proposed in this paper was more accurate and effective than the traditional compliance checking method based on paper documents, which implies that the proposed framework may substantially reduce the problem of wrong information in the compliance checking of building design documents.

## 5. Conclusions

Design for safety (DFS) is an effective way to improve safety management in the construction industry, which has been verified in previous research work. With improvement in China's construction digitization level, the process of BIM-related drafting standards, data standards, and delivery standards is also advancing. In this environment, the design work puts forward higher requirements, and it is necessary to meet the complicated mandatory safety regulations to improve the safety and reliability of design documents. Computer-based automated compliance checking technology has improved the efficiency of the DFS process, but there is still a lack of an efficient method for the semantic and knowledge-based description of BIM information in the design phase of construction projects.

To solve these problems, we made some efforts. In this paper, an NLP method was proposed for dealing with safety regulations, and the results of text mining and data from BIM were used to develop an OWL ontology for DFS. The developed ontology provided

a semantically isomorphic description of the design phase information contained in BIM, the information of the DFS process, and the rule-based and knowledge-based constraint information. This process homogenized information from different sources, improved data interoperability and, in turn, optimized the efficiency of DFS-based automated compliance checks. The work conducted is specifically summarized as:

(1) A DFS automated compliance checking framework based on semantic integration was proposed to improve the format and information management of construction design documents and regulations;

(2) A DFS ontology hierarchy was developed based on NLP text mining of safety regulations and deconstruction of BIM design documents;

(3) A visual programming code written by Dynamo was used to extract the BIM data into a spreadsheet, where the data were mapped to the individuals in the OWL ontology for DFS through the Cellfie tool;

(4) SWRL-based safety compliance checking rules were written to support semantic-integration-based DFS knowledge and constraint expression;

(5) A case study was used to validate the proposed DFS automated compliance checking framework, and 95.21% recall and 90.63% precision in ACC were achieved.

The efforts in this paper had a positive effect on the practice of the DFS concept and the development of the integration of building design and computing technology based on ontology and BIM. Existing typical safety regulations were used for NLP text mining, from which new concept classes for enriching the DFS ontology knowledge base and safety regulation knowledge structured rule expressions for compliance checking were extracted. In addition, not only the content of the DFS ontology knowledge base was enriched, but the proposed method still took effect after the safety regulations were updated or the BIM model was iterated, which can meet the needs of semantic ontology expansion. Moreover, the proposed Dynamo–Cellfie-based BIM data extraction and ontology mapping methods provide a solution for realizing data interoperability in compliance checking.

Although the current efforts were demonstrated in theory and practice, and the primary goals of DFS compliance checking can be achieved, the gap in this paper and subsequent research recommendations still need to be explained. The process of BIM model preprocessing, adding data, and establishing the mapping relationship requires a lot of manual work, which is still unresolved in this paper. Before implementing DFS compliance checks, the required data need to be added to building model elements to ensure that the BIM can contain the semantic information required for ontology-based compliance checks. Moreover, with the development of the AEC sector, DFS ontology has a process of continuous updating. What this paper conducted was preliminary work based on the current DFS concept. Existing DFS theoretical research requires the participation of a large number of scholars and practitioners from academia and industry to achieve safety management goals. With the progress of DFS-related research results, the DFS ontology framework proposed in this paper will face considerations that need to be improved and adjusted. In addition, this paper only uses the basic part of the NLP method, and subsequent research should consider deepening the application of NLP. For example, combining machine learning to mine conceptual classes terms for DFS ontology from text to eliminate the reliance on the creation of user dictionaries that require tedious manual work.

**Author Contributions:** Conceptualization, Y.Z. and J.S.; methodology, Y.Z. and Y.H.; software, Y.Z. and Y.H.; validation, Y.Z.; formal analysis, Y.Z. and Y.H.; investigation, Y.Z. and J.S.; resources, J.S.; data curation, Y.Z.; writing—original draft preparation, Y.Z.; writing—review and editing, Y.H., J.Z., L.L., L.Z. and J.S.; visualization, Y.Z.; supervision, J.S. and L.L.; project administration, J.S.; funding acquisition, J.S. and L.L. 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: NSFC-71901120) and the Special Project of the Ministry of Housing and Urban–Rural Development of The People’s Republic of China (grant number: ZLAQ(2019)AQ-4).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data, models, and code generated or used during the study appear in the submitted article.

**Acknowledgments:** The authors appreciate the Third Construction Co., Ltd., of the China State Construction Engineering Corporation (CSCEC), the provider of the models used in this paper, and the technical supporter.

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

## References

1. Bayoudhi, L.; Sassi, N.; Jaziri, W. A Hybrid Storage Strategy to Manage the Evolution of an OWL 2 DL Domain Ontology. *Procedia Comput. Sci.* **2017**, *112*, 574–583. [CrossRef]
2. Workplace, N. Prevention through Design. Available online: <http://www.dbp.org.uk/> (accessed on 26 April 2022).
3. Gambatese, J.A.; Hinze, J.W.; Haas, C.T. Tool to Design for Construction Worker Safety. *J. Archit. Eng.* **1997**, *3*, 32–41. [CrossRef]
4. Wang, J.; Ruxton, T. A design-for-safety methodology for large engineering systems. *J. Eng. Des.* **1998**, *9*, 159–170. [CrossRef]
5. Weinstein, M.; Gambatese, J.; Hecker, S. Can design improve construction safety?: Assessing the impact of a collaborative safety-in-design process. *J. Constr. Eng. Manag.* **2005**, *131*, 1125–1134. [CrossRef]
6. Tanabe, M.; Miyake, A. Safety design approach for onshore modularized LNG liquefaction plant. *J. Loss Prev. Process Ind.* **2010**, *23*, 507–514. [CrossRef]
7. Sadeghi, L.; Mathieu, L.; Tricot, N.; Al Bassit, L. Developing a safety indicator to measure the safety level during design for safety. *Saf. Sci.* **2015**, *80*, 252–263. [CrossRef]
8. Meacham, B.J. A Sociotechnical Systems Framework for Performance-Based Design for Fire Safety. *Fire Technol.* **2022**, *58*, 1137–1167. [CrossRef]
9. Love, P.E.D.; Edwards, D.J.; Han, S.; Goh, Y.M. Design error reduction: Toward the effective utilization of building information modeling. *Res. Eng. Des.* **2011**, *22*, 173–187. [CrossRef]
10. Qi, J.; Issa, R.; Hinze, J.; Olbina, S. Integration of Safety in Design through the Use of Building Information Modeling. In Proceedings of the International Workshop on Computing in Civil Engineering, Miami, FL, USA, 19–22 June 2011.
11. Qi, J.; Issa, R.R.A.; Olbina, S.; Hinze, J. Use of Building Information Modeling in Design to Prevent Construction Worker Falls. *J. Comput. Civ. Eng.* **2014**, *28*, A4014008. [CrossRef]
12. Hossain, M.A.; Abbott, E.L.S.; Chua, D.K.H.; Qui, N.T.; Goh, Y.M. Design-for-Safety knowledge library for BIM-integrated safety risk reviews. *Autom. Constr.* **2018**, *94*, 290–302. [CrossRef]
13. Tang, Y.C.; Xia, N.N.; Lu, Y.; Varga, L.; Li, Q.M.; Chen, G.C.; Luo, J. BIM-based safety design for emergency evacuation of metro stations. *Autom. Constr.* **2021**, *123*, 103511. [CrossRef]
14. Eastman, C.; Lee, J.M.; Jeong, Y.S.; Lee, J.K. Automatic rule-based checking of building designs. *Autom. Constr.* **2009**, *18*, 1011–1033. [CrossRef]
15. Park, J.W.; Yun, G.L. *Developing a Performance-Based Design System with Semantic Interoperability*; Springer: Berlin/Heidelberg, Germany, 2013.
16. Malsane, S.; Matthews, J.; Lockley, S.; Love, P.E.D.; Greenwood, D. Development of an object model for automated compliance checking. *Autom. Constr.* **2015**, *49*, 51–58. [CrossRef]
17. Mohamed, A.G.; Abdallah, M.R.; Marzouk, M. BIM and semantic web-based maintenance information for existing buildings. *Autom. Constr.* **2020**, *116*, 103209. [CrossRef]
18. Zhong, B.T.; Gan, C.; Luo, H.B.; Xing, X.J. Ontology-based framework for building environmental monitoring and compliance checking under BIM environment. *Build. Environ.* **2018**, *141*, 127–142. [CrossRef]
19. Dibley, M.; Li, H.J.; Rezgui, Y.; Miles, J. An ontology framework for intelligent sensor-based building monitoring. *Autom. Constr.* **2012**, *28*, 1–14. [CrossRef]
20. Jiang, L.; Shi, J.Y.; Wang, C.Y. Multi-ontology fusion and rule development to facilitate automated code compliance checking using BIM and rule-based reasoning. *Adv. Eng. Inform.* **2022**, *51*, 101449. [CrossRef]
21. Uhm, M.; Lee, G.; Park, Y.; Kim, S.; Jung, J.; Lee, J.K. Requirements for computational rule checking of requests for proposals (RFPs) for building designs in South Korea. *Adv. Eng. Inform.* **2015**, *29*, 602–615. [CrossRef]
22. Zhang, J.S.; El-Gohary, N.M. Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking. *Autom. Constr.* **2017**, *73*, 45–57. [CrossRef]
23. Marzouk, M.; Enaba, M. Text analytics to analyze and monitor construction project contract and correspondence. *Autom. Constr.* **2019**, *98*, 265–274. [CrossRef]
24. Hosseini, M.R.; Martek, I.; Papadonikolaki, E.; Sheikhhoshkar, M.; Banihashemi, S.; Arashpour, M. Viability of the BIM Manager Enduring as a Distinct Role: Association Rule Mining of Job Advertisements. *J. Constr. Eng. Manag.* **2018**, *144*, 04018085. [CrossRef]
25. Zhou, S.; Ng, S.; Lee, S.; Xu, J.; Yang, Y. A Domain Knowledge Incorporated Text-Mining Approach for Capturing User Needs on BIM Applications. *Eng. Constr. Archit. Manag.* **2019**, *27*, 458–482. [CrossRef]



26. Lin, J.R.; Hu, Z.Z.; Zhang, J.P.; Yu, F.Q. A Natural-Language-Based Approach to Intelligent Data Retrieval and Representation for Cloud BIM. *Comput.-Aided Civ. Infrastruct. Eng.* **2016**, *31*, 18–33. [[CrossRef](#)]
27. Wang, N.; Issa, R.R.A.; Anumba, C.J. NLP-Based Query-Answering System for Information Extraction from Building Information Models. *J. Comput. Civ. Eng.* **2022**, *36*, 04022004. [[CrossRef](#)]
28. Zhang, J.S.; El-Gohary, N.M. Semantic NLP-Based Information Extraction from Construction Regulatory Documents for Automated Compliance Checking. *J. Comput. Civ. Eng.* **2016**, *30*, 04015014. [[CrossRef](#)]
29. Al Qady, M.; Kandil, A. Concept Relation Extraction from Construction Documents Using Natural Language Processing. *J. Constr. Eng. Manag.* **2010**, *136*, 294–302. [[CrossRef](#)]
30. Shearer, C. The CRISP-DM model: The new blueprint for data mining. *J. Data Warehous.* **2000**, *5*, 13–22.
31. Zhang, H.; Miao, J.; Liu, Z.; Wesson, I.L.; Shang, J. NLP-Parser: Making Chinese and English Semantic Analysis Easier and Complete. In Proceedings of the 15th International Conference on the Statistical Analysis of Textual Data, Rome, Italy, 12–15 June 2020.
32. Zhang, S.; Boukamp, F.; Teizer, J. Ontology-based semantic modeling of construction safety knowledge: Towards automated safety planning for job hazard analysis (JHA). *Autom. Constr.* **2015**, *52*, 29–41. [[CrossRef](#)]
33. Huang, Y.; She, J.; Ye, S. Security Risk Automatic Identification System Based on BIM-Ontology from Perspective of DFS. *J. Civ. Eng. Manag.* **2021**, *38*, 200–207. [[CrossRef](#)]
34. O'Connor, M.J.; Das, A.K. SQWRL: A query language for OWL. In Proceedings of the 5th International Workshop on OWL: Experiences and Directions (OWLED 2009), Chantilly, VA, USA, 23–24 October 2009.
35. O'Connor, M. MappingMasterDSL. Available online: <https://github.com/protegeproject/mapping-master/wiki/MappingMasterDSL> (accessed on 26 April 2022).
36. Member, W.C. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. Available online: <https://www.w3.org/Submission/SWRL/> (accessed on 26 April 2022).
37. Noy, N.F.; Fergerson, R.W.; Musen, M.A. The knowledge model of Protege-2000: Combining interoperability and flexibility. In *Knowledge Engineering and Knowledge Management, Proceedings: Methods, Models, and Tools*; Dieng, R., Corby, O., Eds.; Lecture Notes in Artificial Intelligence; Springer: Berlin/Heidelberg, Germany, 2000; Volume 1937, pp. 17–32.
38. The Protégé team at the Stanford Center for Biomedical Informatics Research. Properties. Available online: <https://protegewiki.stanford.edu/wiki/Special:Properties> (accessed on 26 April 2022).