Advanced Engineering Informatics 48 (2021) 101239

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Semantic information alignment of BIMs to computer-interpretable regulations using ontologies and deep learning

Peng Zhou [a], Nora El-Gohary [b,*]

[a] School of Management Science and Engineering, Central Univ. of Finance and Economics, 39 Xue Yuan South Road, Beijing 100081, China

[b] Dept. of Civil and Environmental Engineering, Univ. of Illinois at Urbana-Champaign, 205 N. Mathews Ave., Urbana, IL 61801, United States

ARTICLE INFO

Keywords:
Building information modeling
Semantic information alignment
Information extraction
Ontology
Deep learning
Automated compliance checking

ABSTRACT

A semantic information alignment method is proposed to align the representations used in building information models (BIMs) to the representations used in energy regulations. Compared to existing alignment efforts, which are either manual or semi-automated, the proposed method aims to automate the alignment process for supporting fully automated energy compliance checking. A first-level simple alignment method is proposed to align single design information instances to single regulatory concepts, in which (1) domain knowledge is used for interpreting the meaning of concepts to recognize candidate instances, and (2) deep learning is used for capturing the semantics behind the words to measure semantic similarity and select the matches. A final complex alignment method is proposed to recognize the instance groups belonging to a regulatory requirement, in which (1) supervised and unsupervised searching algorithms are used to identify the instance pairs, and (2) network modeling is used to group and link the instance pairs to the requirement. The proposed method showed 93.4% recall and 94.7% precision on the testing data.

1. Introduction

Automated energy regulatory compliance checking aims to automatically check the compliance of a building design [i.e., a given instance of a building information model (BIM)] with applicable energy regulations. To achieve a level of full automation, two prerequisites are essential: (1) transforming the regulations into computer-interpretable representations; and (2) aligning the instances of the BIM concepts and relations to the concepts and relations in the computer-interpretable regulatory representations, so that both the BIM and the regulations use the same concepts and relations to define and describe the same objects [1,2]. However, such alignment is challenging because BIMs and regulations use different data representations, conceptualizations, and terminology – to an extent they speak different languages [2].

To address this challenge, a significant number of efforts in the area of automated regulatory compliance checking have been undertaken, in which different techniques were used to model the BIMs and regulations into the same data representation. For example, [3] and [4] used semantic rule languages (Semantic Web Rule Language and N3Logic, respectively) to represent regulations and used Industry Foundation Classes (IFC) Web Ontology Language (OWL) ontologies to represent the BIMs. [2] and [5] utilized domain-specific modeling languages [Regulatory Knowledge Query Language (RKQL) and building environment rule and analysis (BERA) language, respectively] for regulation representation and IFC for BIM representation. [6] and [7] used first order logic and IFC to represent the regulations and BIMs, respectively. Despite the importance of these efforts, their methods for aligning the BIMs to the computer-interpretable regulations are not fully automated; the alignment is typically conducted in either a semi-automated way or a completely manual way.

To address this gap, this paper proposes a semantic information alignment method to automatically align the BIMs to the computer-interpretable regulations. The proposed method is novel in two ways. First, it proposes a first-level simple alignment method to align single design information instances to single regulatory concepts. Domain knowledge is used to interpret the meaning of concepts to recognize potential matching design information instances. An empirical method is used to analyze the patterns of semantic similarity to select the matching instances, in which a deep learning technique is used to measure the semantic similarity. Second, it proposes a final complex alignment method to recognize the groups of instances that belong to a regulatory requirement. Supervised searching and unsupervised

* Corresponding author.
E-mail address: gohary@illinois.edu (N. El-Gohary).

https://doi.org/10.1016/j.aei.2020.101239
Received 23 July 2020; Received in revised form 14 November 2020; Accepted 21 December 2020
Available online 20 April 2021

searching are used to identify the instance pairs, and network modeling is used to group and link the identified instance pairs to the associated regulatory concepts in the regulatory requirement. The proposed semantic information alignment method was tested in aligning design information instances to commercial building energy efficiency requirements.

2. Background

2.1. Industry Foundation Classes and buildingSMART Data Dictionary

The Industry Foundation Classes (IFC) is a widely accepted open specification for data exchange in the architectural, engineering, construction, and facility management (AEC and FM) domain [8]. The IFC specification is the "only existing public and non-proprietary, and well-developed data model for buildings and architecture existing today" [9]. The IFC schema is developed and maintained by the Model Support Group of buildingSMART [10]. The conceptual schema of IFC is written in EXPRESS data modeling language, registered as ISO 10303-11 [11], and IFC now has become the official ISO standard – ISO 16739:2018 [12]. A number of versions of IFC specification have been developed (e.g., IFC2x, IFC2x3 TC1, and IFC4.2). The latest version is IFC4.3 RC1 [13].

The buildingSMART Data Dictionary (bSDD), a semantic mapping tool that links concepts with similar meaning, could be exploited for mapping BIMs and regulations to the same data representation. bSDD, formally known as the International Framework for Dictionaries (IFD), is an ISO 12006-3-based library that contains objects and their properties for the building and construction industry [14]. It aims to help participants identify and share objects and properties regardless of human language [14]. bSDD incorporates the mapping to the IFC specification so that searching a concept or relationship in the bSDD may return the corresponding IFC concepts (which may refer to an IFC entity, enumeration type, etc.) or relationships. For example, searching a concept "slab" in bSDD would return the IFC concepts "IfcSlab" and "IfcSlabType". Each bSDD concept is assigned a name as a label [15]. There are two types of names: long name and short name. The long name refers to the full name. There may exist multiple long names as synonyms (e.g., thermal insulance, thermal resistance, coefficient of thermal insulation, R value). The short name refers to the abbreviation of the concept, and there may exist zero or more short names (e.g., "meter" has a short name "m").

There are nine types of bSDD concepts: "activity", "actor", "classification", "document", "measure", "property", "subject", "unit", and "value". For the definition of each type, the readers are referred to [15]. The "subject" and "property" type bSDD concepts were used in this research to search for the IFC concepts. "Subject" may either refer to a physical object (e.g., a building element duct) or a logical object (e.g., space, submittal) [15]. "Property" refers to an attribute of an object (e.g., R-value is an attribute of an object duct). bSDD concepts are connected by relationships. There are 25 relationships defined in the bSDD. The most commonly-used is the "specialization" relationship, which means one concept is a subconcept of another. For example, a duct is a specialized building element indicating that the concept "duct" is a subconcept of "building element". The "specialization" relationship was used in this research to search for the superconcepts of a bSDD concept. bSDD offers an open source Representational State Transfer (REST) model-based application programming interface (API) for parsing the bSDD concepts and relationships, and searching the matched concepts and relationships in other classifications (e.g., IFC) [15].

2.2. BIM information extraction

BIM information extraction aims to extract information from BIM models and prepare the extracted information for different BIM uses (e.g., building energy analysis, regulatory compliance checking). Depending on the data representations of the BIM models, there are two main BIM information extraction approaches: (1) a proprietary software API-based approach that uses the software API to extract information from BIM models in proprietary softwares; and (2) an open standard data model parsing approach that uses either open source IFC toolboxes [e.g., OpenIFCTools, Java Standard Data Access Interface (JSDAI), xBIM, IFC Engine DLL] or custom-developed parsing algorithms to extract information from BIM models in open specifications (e.g., IFC, ifcXML, gbXML). Table 1 summarizes the BIM information extraction efforts in the recent five years, which used either approach for

Table 1
BIM information extraction efforts in the recent five years.

BIM extraction approach	Extraction tool	Application
Proprietary software API-based approach		**Example efforts in supporting general applications**
	Autodesk Dynamo API	Building energy performance visualization and management [16]
	Navisworks API	Construction risk knowledge management [17]
	Autodesk Revit API	Construction workface planning [18]
		Construction-specific information management [19]
	Feature Manipulation Engine (FME)	Conversion of IFC schema to IndoorGML schema [83]
		Example efforts in supporting regulatory compliance checking
	Autodesk Revit API	Building thermal envelope energy checking [20]
	bim + REST API	Fire safety checking [21]
Open standard data model parsing approach		**Example efforts in supporting general applications**
	OpenIFCTools	Partial building information model extraction [22]
	JSDAI	Dimensional quality assurance of full-scale precast concrete elements [23], and mapping IFC schema to CityGML schema [24]
	xBIM	Indoor and outdoor combined route planning [25]
	IFC Engine Dynamic Link Library (DLL)	BIM semantic information enrichment [26], and indoor space path planning [27]
	Custom-developed parsing algorithms	Construction-specific information management [19], building energy analysis [28–32], partial model extraction [33], safety risk identification [34], automated construction schedules generation [35], automated cost estimation [36], and interior utility network analysis [37]
		Example efforts in supporting regulatory compliance checking
	JSDAI	General building design checking [6]
	IFC Engine DLL	High-rise and complex building evaluation checking [38]
	Custom-developed parsing algorithms	Compliance checking of fire safety [39], building sustainability [3], deep foundation design [62], building accessibility and visibility [5], building energy efficiency [41], general construction conformity [42], general building design [43], and building envelope design [44,45]

supporting different applications. The open standard data model parsing approach is used in this research because an open standard BIM model (e.g., IFC) can ensure platform independency of the proposed semantic information alignment method. An open source toolbox, rather than a custom-developed parsing algorithm, is used to extract the BIM information because open source toolboxes have better scalability across different versions of the IFC specification. The open source toolbox JSDAI was selected because it is Java-based, and fits better with the Java-implementation of the proposed semantic information alignment method.

3. State of the art and knowledge gaps in semantic information alignment for regulatory compliance checking in construction

To check the regulatory compliance of a given instance of a BIM, the BIM information representations should be aligned with those of the regulatory requirements so that both representations can "speak the same language" (or at least translate well). There are a significant number of regulatory compliance checking efforts in the AEC and FM domain, in which four main ways were mainly used to address the alignment problem. In the first approach, concepts and terms of the BIM (e.g., IFC concepts) are used in representing the regulatory requirements (i.e., "write the regulatory requirements using the BIM language and terminology"). The regulatory requirements may be represented as Jess rules (e.g., [46]), Jena rules (e.g., [41,47]), Semantic Web Rule Language (SWRL) rules (e.g., [48]), conceptual graph-represented rules (e.g., [7]), EXPRESS rules (e.g., [49]), BIM-server advanced queries (e.g., [50]), SPARQL Protocol and RDF Query Language (SPARQL) queries (e.g., [42,51]), BIM software API functions (e.g., [52–54]), or custom-developed computer programs (e.g., [55]). In the second, a separate mapping scheme is developed to map the concepts and terms of the regulatory requirements to those of the BIM (i.e., "translate the regulatory language and terminology to the BIM language and terminology"). The mapping scheme may be represented as a mapping ontology (e.g., [3]), N3Logic rules (e.g., [4]), JBoss rules (e.g., [45]), procedural mapping algorithms or functions (e.g., [56,57]), or a set of black box mapping files in the industrial efforts [2] (e.g., DesignCheck [58], SMARTCodes [59], ePlanCheck [60], Solibri Model Checker [61]). In the third approach, the regulatory concepts and terms are used to develop the BIM models (e.g., [38,62]) or extend the representations of the BIM models (e.g., [6]) (i.e., "extend the BIM language and terminology with regulatory conceptualizations and terminology"). In the fourth, the users of the regulatory compliance checking systems are required to specify the alignment between the BIM information and the regulatory requirements using predefined functions or languages (i.e., "conduct a manual translation"), such as high-level query functions [55], Language-Integrated Query (LINQ) [44,84], Regulatory Knowledge Query Language (RKQL) [2], Visual Code Checking Language (VCCL) [63], KBim Visual Language (KBVL) [85], and building environment rule and analysis (BERA) language [5].

Despite the importance of these efforts, their information alignment approaches are limited in one or more of the following three ways. First, all of these approaches require some degree of manual effort (e.g., [2,5,44,55,63] require manual specification of the alignment, by domain experts, using predefined functions/languages). Manual approaches are typically time-consuming, costly, and unscalable [3,64]. Second, many of these efforts are somewhat rigid (e.g., [3,4,45,56–61] use pre-defined mappings or mapping rules). Rigid approaches lack sufficient flexibility and adaptability to allow for successful implementation across BIM instances, different types of regulations, and changes or updates to the BIM or the regulations [1,2]. Third, several of these efforts (especially those by software vendors such as [58–61]) use proprietary methods. Proprietary methods lack the needed transparency to enable the users to check the correctness of the alignment [2].

4. Proposed method for fully automated semantic information alignment

To address the aforementioned gaps, this research proposes a fully automated semantic information alignment method to align the BIM information to the regulatory information. The proposed method aims to align the IFC-represented design information instances to the regulatory information. The proposed method is novel in two ways. First, it captures domain knowledge to automatically interpret the meaning of concepts and recognize the candidate design information instances that are potentially matched to the regulatory concepts, and uses deep learning to capture the semantics behind the words and accordingly measure semantic similarity and select the matches. Second, it uses supervised and unsupervised searching algorithms to automatically identify the relationships that create instance pairs, and uses network modeling to model and group the instance pairs that are linked to the associated concepts in a regulatory requirement. By that, the proposed method breaks down the complex information alignment task into two relatively simpler tasks – individual-individual matching and group-group matching. Accordingly, the proposed method includes two primary sub-methods: (1) a method for first-level, simple alignment (individual-individual matching): matching single design information instances to single regulatory concepts; and (2) a method for final, complex alignment (group-group matching): recognizing the regulatory concepts that belong to one requirement, and linking the matched design information instances to these associated regulatory concepts.

4.1. First-level simple alignment

The first-level simple alignment method aims to align single design information instances (i.e., the instances of the IFC entities in an instance of a BIM, which the authors call thereafter "BIM instances") to single regulatory concepts. There are two types of regulatory concepts: object concepts and property concepts. An object concept refers to an object such as a building element (e.g., duct). A property concept refers to a property of an object (e.g., thermal resistance is a property of a duct). Accordingly, two types of BIM instances are defined: object instances and property instances. An object instance refers to an instance of an IFC entity that matches or aligns to an object concept, including instances of IfcProduct, IfcProductType, IfcSystem, and IfcMaterial. A property instance refers to an instance of an IFC entity that matches or aligns to a property concept, including instances of IfcSimpleProperty, IfcPhysicalSimpleQuantity, IfcMaterialProperties, and IfcMaterialLayer.

The proposed method includes two primary steps: concept interpretation and matching, and semantic similarity analysis. Concept interpretation and matching aims to interpret the meaning of regulatory concepts and accordingly to select an initial set of candidate matches. A match is defined, in this paper, as a BIM instance that is matched or aligned to a concept. Semantic similarity analysis aims to assess the semantic similarity for each candidate pair (BIM instance and regulatory concept), and accordingly to select the matches for regulatory concepts. A domain ontology (called commercial building energy ontology) is used to support both steps. Fig. 1 illustrates the method for first-level simple alignment.

4.1.1. Concept interpretation and matching

Concept interpretation and matching aims to interpret the meaning of a regulatory concept to recognize all candidate matches (BIM instances). The proposed approach uses ontology and bSDD to capture domain knowledge for the interpretation of the meaning of concepts to automatically recognize the candidate matches. This step includes: object concept interpretation and matching, and property recognition.

4.1.1.1. Object concept interpretation and matching. Object concept interpretation aims to recognize all candidate object instances

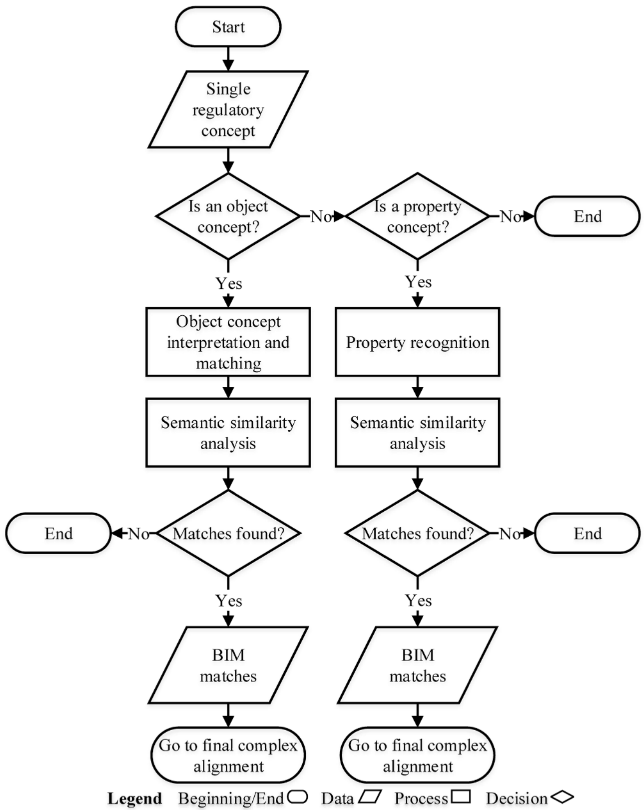


Fig. 1. Method for first-level simple alignment.

(candidate matches) for a regulatory object concept. Three methods for object concept interpretation and matching are proposed and used, as shown in Fig. 2: (1) concept interpretation using bSDD direct searching for finding perfect matches; (2) concept interpretation using ontology-based concept decomposition for finding parent matches; and (3) concept interpretation using superconcept information for finding parent matches. The three methods are conditionally dependent: if a preceding method fails to find a match, the following method is used. If all three methods fail, no match to that concept has been found.

Method 1: Concept interpretation using bSDD direct searching for finding perfect matches. This method uses the bSDD API to directly search the bSDD to find the perfectly matching IFC entities (100% match to the morphologically-analyzed regulatory concept). The instances of these matching IFC entities (including instances of their subentities) are the matches for the regulatory object concepts. First, the object concepts (i.e., the terms in the concept name) are morphologically analyzed to collapse the different derivational (e.g., affixes like "ly", "ion") and inflectional forms (e.g., plural, progressive) of each term to its base form. Then, three techniques are used for searching: (1) using potentially-equivalent concept names: a set of potentially-equivalent concept names are defined based on the original and base forms of the terms. For example, the following three concept names are potentially equivalent to the original concept name "lighting fixtures": "lighting fixture", "light fixtures", and "light fixture". Using this technique, instances of the IfcLightFixture and IfcLightFixtureType entities would be recognized as perfect matches to "lighting fixtures"; (2) using bSDD synonyms: synonyms of bSDD concept names are used in the search for matching IFC entities. For example, the synonyms of the bSDD concept "lighting fixture" include "luminaire". Using this technique, instances of the IfcLightFixture and IfcLightFixtureType entities would be recognized as perfect matches to "luminaire"; and (3) using ontology-based equivalent concepts: similar to bSDD synonyms, equivalent concepts in the ontology are used in the search for matching IFC entities. For example, in the ontology, the concept "beam" is equivalent to the concept "girder". Using this technique, instances of the IfcBeam and IfcBeamType entities would be recognized as perfect matches to "girder".

Method 2: Concept interpretation using ontology-based concept decomposition for finding parent matches. If Method 1 fails to find any matching IFC entities, Method 2 is used. This method uses ontology-based concept decomposition to find matching IFC entity parents. The instances of those IFC entity parents (including instances of their subentities) are the candidate matches for the regulatory object concepts. A given object concept (e.g., "metal-framed roof") is decomposed into two parts: (1) a core inner concept carrying the most important meaning of the given concept (e.g., "roof"); and (2) the remaining part of that given concept (e.g., "metal-framed"), which could be viewed as the property information of that core inner concept. The core inner concept name is then used to recognize the matching parents using Method 1.

Two steps are proposed and used to decompose a concept based on the ontology. First, ontology parsing is used to find all potential inner

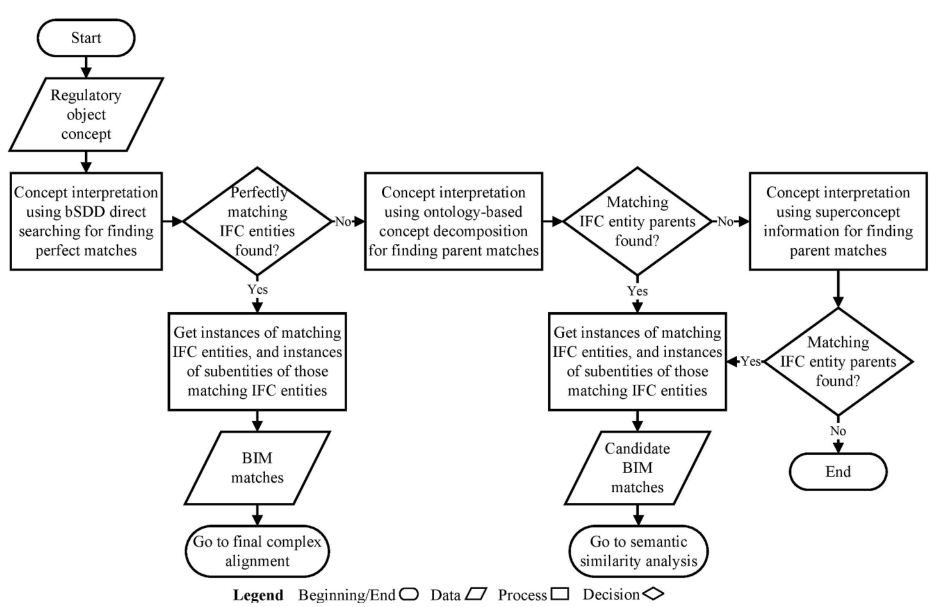


Fig. 2. Method for object concept interpretation and matching.

concepts for a given concept (using morphological analysis as necessary). Second, conflict resolution methods are used to select the core inner concept from those potential ones. Three conflict resolution methods are proposed and used: (1) iterative concept name reduction: if there are two (or more) equivalent concepts among those inner concepts, the longer concept(s) (i.e., concept with the longer name) is replaced by the shorter one. This is based on the hypothesis that if a concept uses more terms to express the same meaning, it contains redundant information. For example, the first round of ontology parsing finds the following inner concepts for the concept “metal building roof assembly”: “metal”, “metal building”, “metal building roof”, “building”, “roof,” “roof assembly”, among which “roof assembly” and “roof” are equivalent. Accordingly, the concept “metal building roof assembly” is reduced to “metal building roof”; (2) concept removal: any inner concept that is inside another inner concept is removed. For example, among the four inner concepts of “metal building roof” – “metal”, “metal building”, “building”, and “roof” – the concepts “metal” and “building” are removed; and (3) concept selection: if there are multiple inner concepts remaining (after iterative concept name reduction and concept removal), the rightmost inner concept (with reference to its position in the given concept) is selected, because the rightmost term in a concept carries the most important meaning. For example, among the remaining inner concepts, “metal building” and “roof”, the rightmost concept “roof” is selected as the core one. Accordingly, searching for the concept “roof” in the bSDD (using Method 1) would return IfcRoof and IfcRoofType. As such, using Method 2, the IfcRoof and IfcRoofType entities would be recognized as the matching parents of “metal building roof assembly” (thus filtering out the instances of the non-matching IFC entities).

Method 3: Concept interpretation using superconcept information for finding parent matches. If Method 2 fails to find any matching IFC entities, Method 3 is used. This method uses superconcept information (in the ontology and the bSDD) to find matching IFC entity parents. The instances of these IFC entity parents (including instances of their subentities) are the candidate matches for the object concepts. The search is conducted in a recursive manner: if a lower-level superconcept fails to find a matching parent, a higher-level superconcept is used in the search – until the root concept is reached. For example, using Method 3, the IfcMaterial entity would be recognized as the matching parent of “radiant panel” (using the search term “material”, where “material” is a superconcept of “radiant panel” in the ontology).

4.1.1.2. Property recognition. Property recognition aims to recognize all candidate property instances (candidate matches) for a regulatory property concept. The recognition of candidates is conducted in an indirect way, as per Fig. 3: (1) the object concepts associated with a given property concept are identified; (2) the matching object instances are recognized; and (3) all the property instances that are associated with each of the matching object instances are captured as candidate property

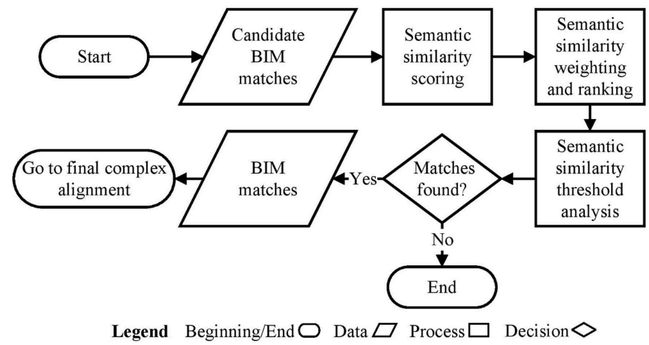


Fig. 4. Method for semantic similarity analysis.

instances. Property instances without any quantitative property values are excluded from the candidate list, because this research only focuses on checking the compliance with quantitative requirements.

4.1.2. Semantic similarity analysis

Semantic similarity analysis aims to assess the semantic similarity for each candidate pair (BIM instance and regulatory concept), and accordingly to select the aligned BIM instances (i.e., matches) for regulatory concepts. The proposed approach is novel in two ways. First, it uses deep learning to capture the semantics behind the words for enhanced assessment of the semantic similarity. Second, it uses an empirical way to analyze the patterns of semantic similarities for enhanced recognition of concept matches. The semantic similarity analysis includes three sequential steps, as per Fig. 4: semantic similarity scoring, semantic similarity weighting and ranking, and semantic similarity threshold analysis.

4.1.2.1. Semantic similarity scoring. Semantic similarity scoring aims to assess the semantic similarity between a BIM instance (a candidate match) and a regulatory concept. The proposed scoring method: (1) describes each instance by its entity, property, and material information, all which are called ‘instance descriptors’ thereafter; (2) describes each concept by its name, regulatory definition, and its equivalency or synonym information, all which are called ‘concept descriptors’ thereafter; (3) assesses term-to-term semantic similarity (i.e., similarity between a term in a BIM instance descriptor and a term in a regulatory concept descriptor) using a deep learning technique; and (4) assesses instance-concept similarity based on all term-to-term similarities.

The following types of instance descriptors are proposed and used for similarity assessment of object pairs (object instance and object concept): (1) corresponding IFC entity information (e.g., the entity attribute “name” and its value “Basic Roof:EPDM – 4 1/2" – lsf 3:9773671” for an IfcRoof instance); (2) property information (e.g., property name “Heat Transfer Coefficient (U)” and value

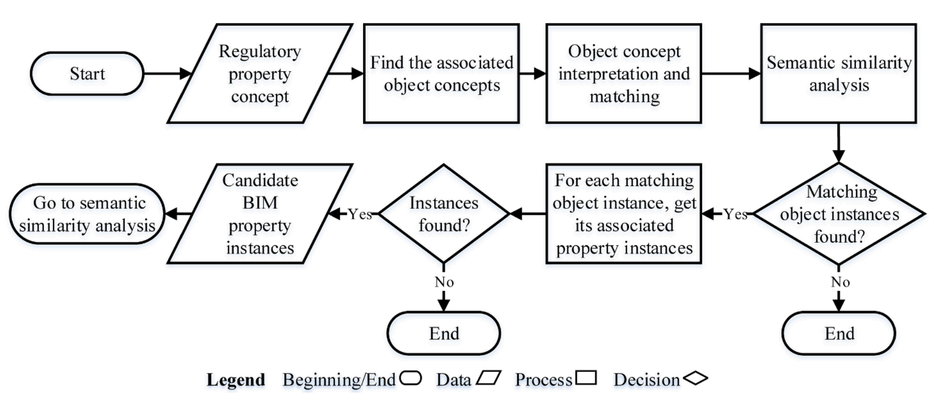


Fig. 3. Method for property recognition.

"0.374015748031496" for the IfcRoof instance); (3) material information (e.g., "Insulation/Thermal Barriers – Batt insulation #ThermalAssetName: Glass Fiber Batt" is the name of a material layer for the IfcRoof instance). If the object instance is an aggregated type instance, the material information of the aggregated instances is used as the material information of the object instance; and (4) the instance name and corresponding IFC entity name (without the prefix "Ifc") of the instances that are spatially located inside the object. For example, if an IfcLamp instance is spatially located inside an IfcSpace instance, the IfcLamp instance name "LED lamp" and its entity name "Lamp" are used as descriptors for the IfcSpace instance. For property pairs (property instance and property concept), the name of the property instance [e.g., "Heat Transfer Coefficient (U)"] is used as its descriptor.

The following types of concept descriptors are proposed and used for similarity assessment of candidate pairs (both object pairs and property pairs): (1) concept name (e.g., u factor), (2) names of its equivalent concepts in the ontology (e.g., u value), (3) names of its synonyms in the bSDD (e.g., heat transfer), (4) names of its acronyms (e.g., LED is an acronym of light emitting diode), and (5) its quantitative definitions. The construction domain-specific acronyms were manually collected from a number of regulatory documents. The quantitative definitions were extracted from the regulatory documents using the proposed ontology-based information extraction method in [65] and represented using four semantic information elements (comparative relation, compliance checking attribute, quantity value, and quantity unit/reference). For example, the concept "low-sloped roof" is described using the following semantic information elements, which were extracted from the 2012 International Energy Conservation Code [66]: "slope" (compliance checking attribute), "less than" (comparative relation), and "2/12" (quantity value).

The term-to-term semantic similarity is assessed using deep learning. A deep learning technique (i.e., Word2vec) [40] is first used to learn a vector representation for each term (excluding stopwords and low-frequncey terms) from a number of energy regulatory documents. Stopwords (e.g., "at", "the") were excluded because they do not carry important meaning, while low-frequency terms (terms with a frequency less than five) were excluded because accurate vector representations cannot be learned for low-occurrence terms [76]. Then, term-to-term semantic similarity is calculated as the cosine similarity of their corresponding vectors, as per Equation (1). The similarity between two same terms is always 1 regardless of their term frequencies. The similarity is set to zero if at least one term has a frequency less than five. A positive cosine similarity value indicates two terms are similar, while a negative value means two terms are dissimilar. To avoid negative similarity values, the semantic similarity scoring function in Equation (1) is further transformed to Equation (2) using an exponential function. The value 100 was empirically selected as the base of the exponential function to give more power to similar terms, while discounting the semantic similarity of dissimilar terms.

The total instance-concept similarity (TS_{ic}) is assessed by aggregating all possible pairs of term-to-term similarities, as per Equation (3). Duplicate terms (after stemming, morphological analysis, and removing stopwords) are not double-counted to avoid potential TS_{ic} inflation.

$$TS_{ic} = \sum_{term_m \in instance, term_n \in concept} S_{tr}(term_m, term_n) \quad (3)$$

4.1.2.2. Semantic similarity weighting and ranking. Semantic similarity weighting and ranking aims to weight the term-to-term semantic similarity for different terms, weight the total instance-concept similarity for different candidate matches (i.e., BIM instances) for a regulatory concept, and accordingly rank the candidate matches. The weighting aims to capture the factors that impact the degree of similarity such as positions of matching terms and lengths of instance descriptors.

Three term-to-term semantic similarity weighting functions are proposed and used to capture the fact that different terms may have different powers in indicating the matching degree between a BIM instance and a regulatory concept – based on the degree of match and position. First, if the descriptors of a concept share common terms (100% match) with the descriptors of a BIM instance, it is likely that the BIM instance is related to that concept. Thus, higher term-to-term semantic similarity should be given to those common terms. Equation (4) shows the weighted term-to-term semantic similarity for the common terms (S_{ct}), in which one bonus point is empirically given to the term-to-term semantic similarity (S_{tr}) of those common terms. Second, if there are two or more common terms and these common terms are adjacent in both descriptors, this indicates more confidence that the BIM instance is related to that concept. Thus, the term-to-term semantic similarity of those common adjacent terms should be further increased. Equation (5) shows the weighted term-to-term semantic similarity for the common adjacent terms (S_{cat}), which is double of the S_{ct}. Third, since the carried meaning of terms in a concept decreases from right to left, the term-to-term semantic similarity should be further adjusted based on the term positions in the concept. Equation (6) shows the proposed term position weighting (TPW) function, which is logarithmic to avoid overweighting the rightmost terms in the concept. Fig. 5 shows an example of applying those three term-to-term semantic similarity weighting functions for assessing the matching degree between the BIM instance "IfcDuctSegment (#4946670)" and the regulatory concept "supply and return air duct".

An instance-concept semantic similarity weighting function is proposed to capture the impact of the instance descriptor lengths on the total instance-concept similarity (TS_{ic}). Since longer instance descriptors tend to result in higher TS_{ic} than shorter ones (i.e., longer descriptors result in aggregating more pairs of term-to-term semantic similarities), the TS_{ic} should be weighted by the average length of all instance descriptors. Equation (7) shows the weighted total instance-concept semantic similarity ($TS_{weighted}$), where the TS_{ct} is the weighted total term-to-term semantic similarity of common terms and the TS_{nct} is the total term-to-term semantic similarity of noncommon terms. Only the TS_{nct} is weighted by the average length of all instance descriptors to avoid discounting the TS_{ct}. Equations (8) and (9) show the equations used for calculating the TS_{ct} and the TS_{nct}, respectively, where k is the total number of common terms.

$$sim(term_m, term_n) = \begin{cases} 1, \textit{if} \ term_m = term_n \\ \textit{cosine}(term_m, term_n), \textit{if} \ \text{frequencies of both } term_m \ \textit{and} \ term_n \geq 5, term_m \neq term_n \\ 0, \textit{if} \ 1 \leq \text{frequency of either } term_m \ \textit{or} \ term_n < 5, term_m \neq term_n \end{cases} \quad (1)$$

$$S_{tr}(term_m, term_n) = 100^{sim(term_m, term_n) - 1} \quad (2)$$

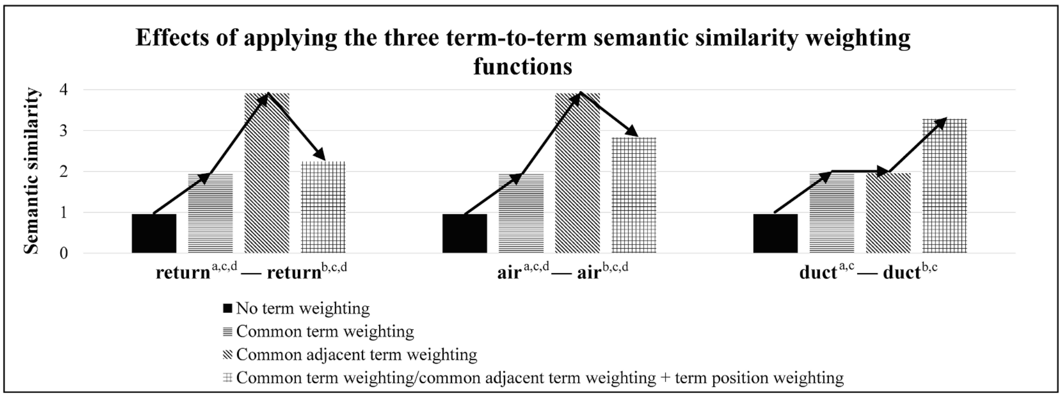
$$S_{ct}(term_n, term_n) = S_{tr}(term_n, term_n) + 1 \quad (4)$$

$$S_{cat}(term_n, term_n) = 2{*}S_{ct}(term_n, term_n) \quad (5)$$

$$TPW(term_n) = \frac{\log(n+1)}{\sum_{m=1}^{concept\ length}\log(m+1)}, term_n\ \text{is the}\ n^{th}\ \text{term from left in a concept} \tag{6}$$

$$TS_{weighted} = TS_{ct} + \frac{TS_{nct}}{\text{instance descriptor length}} * \text{average length of all instance descriptors} \tag{7}$$

$$TS_{ct} = \left[\sum_{term_n \in common\ nonadjacent\ terms} TPW(term_n) * S_{ct}(term_n, term_n) + \sum_{term_m \in common\ adjacent\ terms} TPW(term_m) * S_{cat}(term_m, term_m)\right] * k,$$
$$k\ \text{is the total number of common terms} \tag{8}$$



[a]This term comes from the concept descriptors below.
[b]This term comes from the BIM instance descriptors below.
[c]Common term between concept descriptors and BIM instance descriptors.
[d]Common adjacent term between concept descriptors and BIM instance descriptors.

Concept	BIM instance	Concept descriptors	BIM instance descriptors
supply and return air duct	#4946670= IFCDUCTSEGMENT('1VrzP4okL1AOes_uCjegtj',#42,'Rectangular Duct:Rectangular Duct - 1":9757602',$,'Rectangular Duct:Rectangular Duct - 1":9757654',#4946649, #4946665,'9757602',$);	supply and **<u>return air duct</u>**	**IFC entity information:** {name,Rectangular Duct:Rectangular Duct - 1":9757602,object type, Rectangular Duct:Rectangular Duct - 1":9757654} **Property information(partial):** {Insulation,#4946674,Insulation Thickness,0.025399999999999992,{{null, metre, '1'}}}; {Insulation,#4946675,Insulation Type,Batt Insulation 1",{{null, one, '1'}}}; {Mechanical,#4946472,System Type,**<u>Duct</u>** System: **<u>Return Air,</u>**{{null, one, '1'}}}; {Mechanical,#4946686,Area,1.623749010560001,{{null, metre, '2'}}}; {Mechanical,#4946687,Bottom Elevation,4.1148,{{null, metre, '1'}}}; {Mechanical,#4946688,Equivalent Diameter,0.4460769616132811,{{null, metre, '1'}}}; {Mechanical,#4946690,System Classification,Return Air,{{null, one, '1'}}}; {Mechanical,#4946691,System Name,Mechanical Return Air 1,{{null, one, '1'}}};

Fig. 5. Example of applying the three term-to-term semantic similarity weighting functions.

$$TS_{nct} = \sum_{term_m \in instance, term_n \in concept, term_m \neq term_n} S_{tr}(term_m, term_n) \tag{9}$$

After semantic similarity weighting, the candidate matches (BIM instances) for a given regulatory concept are ranked by the $TS_{weighted}$, in a decreasing order.

4.1.2.3. Semantic similarity threshold analysis. Semantic similarity threshold analysis aims to analyze the degree of instance-concept similarity, based on a set of thresholds, in order to select the matches from the set of ranked candidate matches for a regulatory concept. Two sets of thresholds were experimentally set: one for the object concepts and one for the property concepts, as shown in Table 2, respectively.

Threshold analysis for object pairs: In order to select the matches (from a set of candidate matches) for an object concept, the following four threshold types are proposed and used together to indicate the similarity cut-off (i.e., the level of similarity that indicates a match):

- The minimum TS_{Oct} threshold defines the threshold for matching – a candidate pair must have a TS_{ct} larger than the minimum TS_{Oct} threshold to be a match. This threshold helps define the cutoff of similarity indicated by the degree of sharing common terms (e.g., a

Table 2
Threshold types and values for regulatory concepts.

Concept type	Threshold type	Value
Object concept	Minimum TS_{Oct}[1]	TS_{ct} (t[2])
	Minimum $TS_{Oweighted}$	0.5
	Minimum normalized $TS_{Oweighted}$	0.5
	Minimum normalized $TS_{Oweighted}$ difference	0.06
Property concept	Minimum $TS_{Pweighted}$	$(1/\sum_{n=1}^{\text{Concept length}} n) * 2$
	Minimum normalized $TS_{Pweighted}$	0.5
	Minimum normalized $TS_{Pweighted}$ difference	0.05

[1] Only applies to the BIM instances recognized using Method 2 (Section 4.1.1.1).
[2] Calculated using Eq. (8), where common terms ∈ core inner concept. A core inner concept carries the most important meaning of a given concept (e.g., "roof" is the core inner concept for the given concept "metal-framed roof").

pair sharing 3 adjacent common terms would have a higher total similarity than a pair sharing 2 non-adjacent common terms).

- The minimum $TS_{Oweighted}$ threshold helps define the cutoff of similarity indicated by the weighted total instance-concept similarity ($TS_{weighted}$), where a $TS_{weighted}$ lower than the threshold is probably not reflecting a large-enough degree of similarity to indicate a match.
- The minimum normalized $TS_{Oweighted}$ threshold is a normalized version of the former threshold – normalized to the range of (0, 1) by dividing the $TS_{weighted}$ of each by the highest $TS_{weighted}$ among all pairs. The normalization helps adjust the threshold values to a common scale.
- The minimum normalized $TS_{Oweighted}$ difference threshold helps define the cutoff of similarity indicated by the difference in similarity between two adjacently-ranked pairs. The higher the difference, the larger the similarity jump from one pair to the other. The threshold helps indicate whether this jump is large enough to mean that the pair with the higher normalized $TS_{weighted}$ is similar but the following pair is not similar enough to be a match. Given a set of ranked pairs, a difference value larger than the threshold indicates that the pairs before the cutoff are sufficiently different than the pairs after – the ones before are similar (i.e., matches), but the ones after are dissimilar.

The four thresholds are used in the following manner to collectively indicate which candidate pairs meet the different types of similarity cutoffs – the pairs that meet all four thresholds, in the following way, are likely to be matches. First, all candidates that do not meet both the minimum TS_{Oct} threshold and the minimum $TS_{Oweighted}$ threshold are filtered out. Second, the remaining candidates that meet both the minimum normalized $TS_{Oweighted}$ threshold and the minimum normalized $TS_{Oweighted}$ difference threshold cutoff qualify as matches. For example, Fig. 6 shows that there are 72 candidate BIM instances for the concept

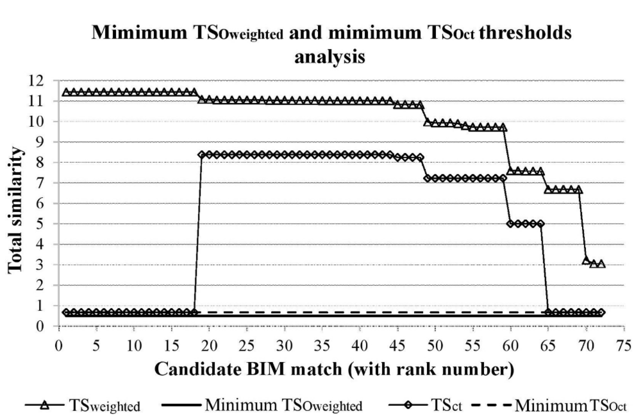


Fig. 6. Example of minimum $TS_{Oweighted}$ and minimum TS_{Oct} thresholds analysis.

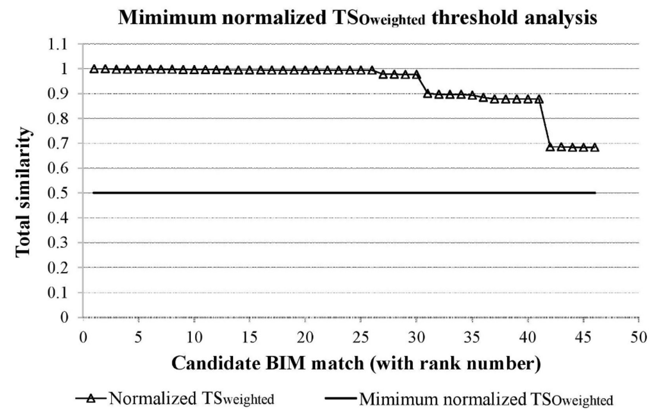


Fig. 7. Example of minimum normalized $TS_{Oweighted}$ threshold analysis.

"supply and return air duct", 46 of them (ranked 19 to 64) meet both the minimum TS_{Oct} threshold and the minimum $TS_{Oweighted}$ threshold. All 72 instances meet the minimum $TS_{Oweighted}$ threshold, which indicates that they all have a large-enough degree of total similarity to indicate a potential match – they are all duct instances. However, only 46 of those instances (ranked 19 to 64) meet the minimum TS_{Oct} threshold, which indicates that they all have a large-enough degree of common-term similarity. These 46 instances have a higher degree of sharing terms, compared to the other 26 that only share the term "duct". Figs. 7 and 8 show that only 41 of the 46 candidate instances meet both the minimum normalized $TS_{Oweighted}$ threshold and the minimum normalized $TS_{Oweighted}$ difference threshold cutoff. Fig. 7 shows that all 46 instances meet the minimum normalized $TS_{Oweighted}$ threshold, which indicates they all have a large-enough degree of normalized total similarity. Fig. 8 further shows that there are two large-enough similarity jumps: a larger jump (between rank 41 and 42) and a smaller jump (between rank 30 and 31). The 41 instances share two common adjacent terms "supply air" or "return air" and share another common non-adjacent term "duct", while the other five instances (42 to 46) have relatively lower common-term similarity; they only share two non-adjacent common terms "air" and "duct".

Threshold analysis for property pairs: The threshold analysis for property pairs is similar to that for object pairs. Three similar threshold types are proposed and used to select the matching BIM instances for a property concept: minimum $TS_{Pweighted}$, minimum normalized $TS_{Pweighted}$, and minimum normalized $TS_{Pweighted}$ difference thresholds. All property candidates must meet all three thresholds to qualify as matches.

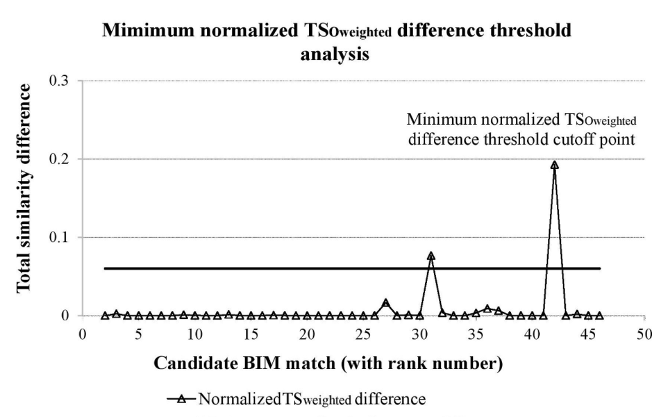


Fig. 8. Example of minimum normalized $TS_{Oweighted}$ difference threshold analysis.

4.2. Final complex alignment

Final complex (group-group) alignment aims to recognize the object concepts that belong to one requirement (called thereafter 'concept group'), find the matches to each concept in that concept group (using the methods in Section 4.1), and accordingly recognize the instance groups (which could be one or more) that are linked to that concept group (i.e., recognize the instance groups that belong to one requirement). The proposed final complex alignment approach is novel in two ways. First, it uses supervised and unsupervised searching to find the relationships that create instance pairs. Second, it uses network modeling to model and link concept groups and their associated instances, where each concept group and its associated instance groups are modeled as a network of linked concept pairs and instance pairs. The proposed final complex alignment method is, thus, composed of three main steps, illustrated in Fig. 9: supervised searching, unsupervised searching, and network construction.

4.2.1. Supervised searching

Supervised searching aims to recognize, in a supervised manner (i.e., using known or predefined relationships), the relationships that create instance pairs. Two types of relationships were empirically predefined and used for searching: (1) object-material usage relationship, which is defined as a relationship (an instance of the IfcRelAssociatesMaterial entity) that links two IfcMaterial object instances, or links one IfcMaterial object instance and one IfcProduct (or IfcTypeProduct) object instance; and (2) spatially-contained relationship, which is defined as a relationship (an instance of the IfcRelContainedInSpatialStructure entity or an instance captured using bounding box geometric assessment) that links two IfcProduct (or IfcTypeProduct) object instances, where one object instance is spatially contained in the other (e.g., a lighting fixture is spatially contained in a space). Bounding box geometric assessment indirectly captures the spatially-contained relationship between two object instances by assessing whether the bounding box of one instance is entirely inside the bounding box of the other instance, where bounding box refers to the geometric orthogonal box representation of an object instance. For example, Fig. 10 shows that the bounding box of an IfcLightFixture instance is inside the bounding box of an IfcSpace instance.

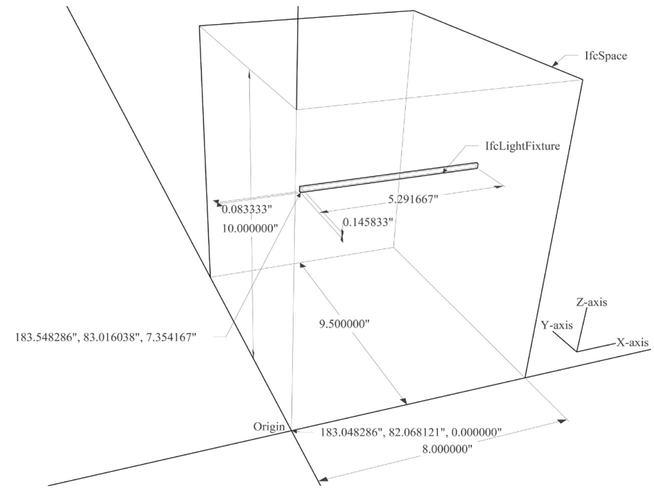


Fig. 10. The bounding box of an IfcLightFixture instance is inside the bounding box of an IfcSpace instance.

4.2.2. Unsupervised searching

Unsupervised searching aims to recognize, in an unsupervised manner (i.e., without any known or predefined relationships), the relationships that create instance pairs. A graph-based searching algorithm is used to automatically find the relationships. A relationship, in this case, is an instance of the IfcRelationship entity (including the instances of all its subentities except IfcRelAssociatesMaterial and IfcRelContainedInSpatialStructure, which are used in the supervised searching) that links object instances of IfcProduct (or IfcTypeProduct) and IfcSystem – using one-to-many association (i.e., one IfcRelationship instance may link one relating object instance to many related object instances). This research assumes that two object instances are linked if there exists at least one path between the two instances. For example, Fig. 11 shows that (1) an IfcRelAssignsToGroup (a subentity of the IfcRelationship) instance links an IfcSystem instance (a relating instance) to IfcValve and IfcPipeSegment instances (two related instances); and (2) an IfcValve instance and an IfcPipeSegment instance are linked by two paths: a longer path linked by instances of IfcRelConnectsPortToElement, IfcDistributionPort, and IfcRelConnectsPorts; and a shorter path linked by an instance of IfcRelAssignsToGroup. To automatically find the path(s) between two object instances, an IFC instance graph is built by parsing the instances of IfcRelationship, in which an edge represents an IfcRelationship instance and a node represents a relating or a related object instance. Then, a graph searching algorithm is used to search the IFC instance graph for a path between the two object instances. Because there may exist multiple paths between two instances, it was assumed that the shortest path is the one that determines if two instances are linked. For example, the IfcValve instance and the IfcPipeSegment instance (in Fig. 11) were linked as a pair by the IfcRelAssignsToGroup instance, because they occurred on the shortest path. The commonly-used Dijkstra searching algorithm [68] was used in this research for shortest path finding.

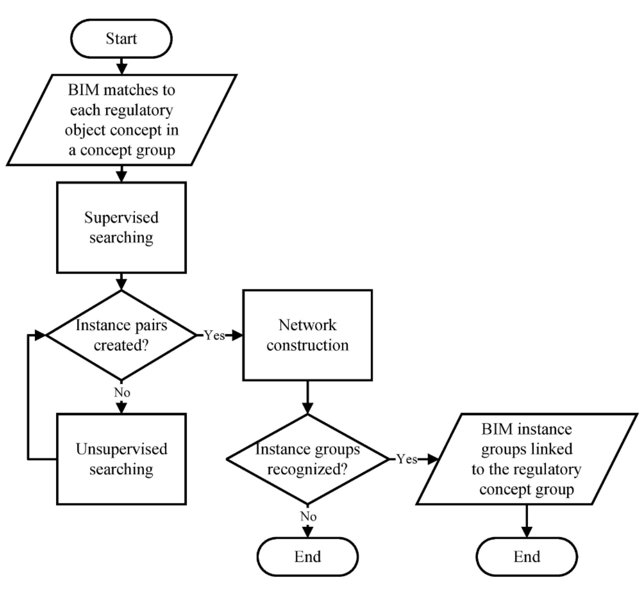


Fig. 9. Method for final complex alignment.

4.2.3. Network construction

Network construction aims to link concept groups and their associated instances into a set of networks. Instance pairs (identified in Section 4.2.1 and 4.2.2) are grouped and linked to their concept groups, where each concept group is related to a regulatory requirement. A set of instance pairs belongs to one instance group, if those instance pairs (1) have one-to-one correspondence to the concept pairs, and (2) can be linked in the same way as the concept pairs in the concept group. An example of final complex alignment is shown in Fig. 12, where eight instance pairs were identified using unsupervised searching, and two instance groups were accordingly formed and linked to a concept group using network construction.

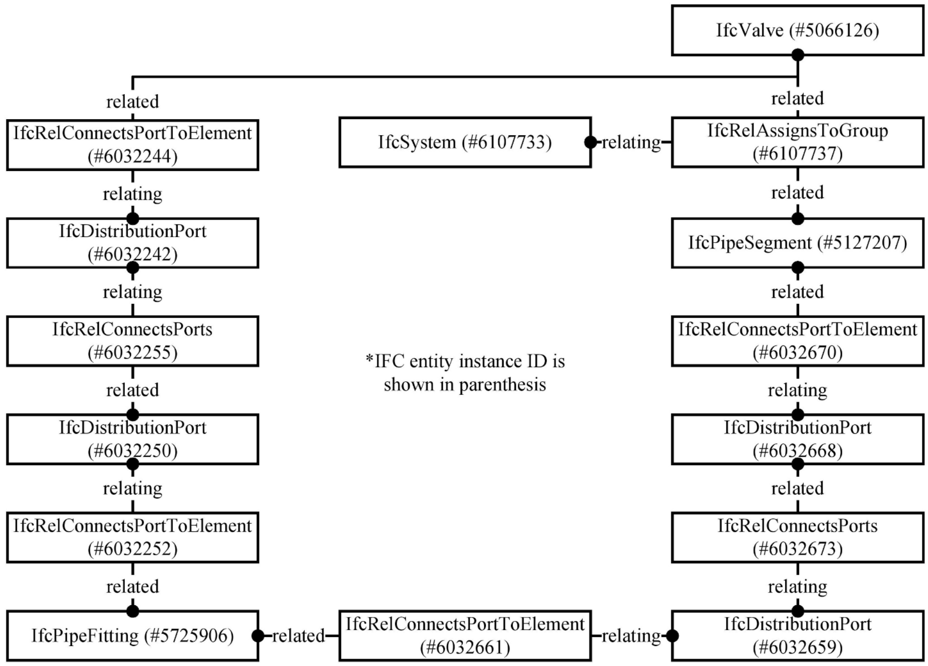


Fig. 11. Examples of relationships between IFC instances that were found using unsupervised searching.

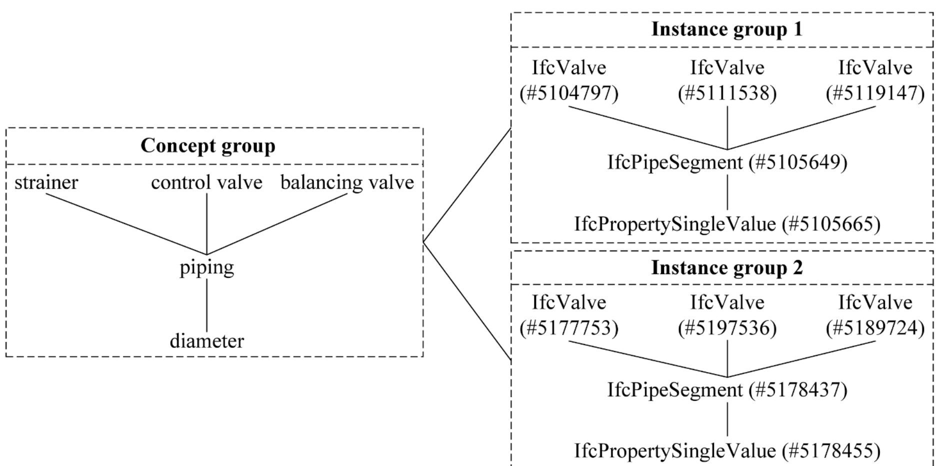


Fig. 12. Example of final complex alignment.

5. Implementation of the proposed method for testing and evaluation

The proposed information alignment method was implemented for testing and evaluation. A Revit model of an educational building project in Illinois, which was created using Autodesk Revit 2016 [69], was used for testing. Regulatory requirements were extracted from three energy regulatory documents – the 2012 International Energy Conservation Code [66], the 2013 Building Energy Efficiency Standards [67] (known as the California Energy Code), and the Ontario Building Code Supplementary Standard SB-10 [82], which represent energy regulatory documents developed by the international council, specific US states, and other countries, respectively. The scope of the testing was limited to commercial building thermal insulation requirements and lighting power requirements (i.e., two subtypes of energy requirements). The implementation included four steps: BIM information extraction, regulatory information extraction, semantic information alignment (using the proposed method, as per Section 4), and evaluation.

5.1. BIM information extraction

The BIM information was extracted from the .ifc file of the BIM model and processed into an intermediate representation for further alignment to the regulatory information. The BIM information was extracted using three steps: IFC export, information extraction from IFC data file, and post-processing of the extracted information.

5.1.1. IFC export

To ensure the platform independency of the implementation, the information in the Revit model was exported to an IFC data file (.ifc file) using the Revit IFC exporter. Since the current Revit IFC exporter [69] is limited in exporting commercial building energy efficiency design information (e.g., export of material thermal properties such as thermal conductivity is not supported), the exporter was extended in the C# programming language using the Microsoft Visual Studio 2015 [70]. The IFC4 specification [71] was selected because it is the most recent version supported by Revit, and has extended support for exporting

Table 3
List of IFC entities used in the BIM information extraction.

Relationship entities	Object-related entities	Material-related entities	Property-related entities	Unit-related entities	Geometry-related entities
IfcRelationship*	IfcProduct* IfcTypeProduct* IfcSystem	IfcMaterialList IfcMaterial IfcMaterialLayerSetUsage IfcMaterialProfileSetUsage IfcMaterialLayerSet IfcMaterialLayer	IfcPropertySet* IfcPreDefinedPropertySet* IfcSimpleProperty* IfcPhysicalSimpleQuantity* IfcQuantitySet* IfcMaterialProperties	IfcDerivedUnit IfcDerivedUnitElement IfcConversionBasedUnit IfcSIUnit IfcDimensionalExponents IfcMeasureWithUnit	IfcProductDefinitionShape IfcShapeRepresentation IfcStyledRepresentation IfcBoundingBox IfcCartesianPoint IfcLocalPlacement IfcAxis2Placement3D IfcDirection

*Abstract entity.

energy domain entities (e.g., lighting entities).

5.1.2. Information extraction from IFC data file

The instances of the entities were extracted from the .ifc file using an EXPRESS-based information extraction method (an adaptation of a method in the authors' previous work [72]). An EXPRESS-based data access method was used to extract all IFC entity instances into an intermediate representation: [IFC Entity Name, IFC Entity Instance ID, IFC Entity Attribute Names, IFC Entity Attribute Values]. Since not all IFC entities are relevant to regulatory compliance checking (e.g., the IfcProcess entity defines an individual activity or event, which may be used in construction scheduling), only instances of the IFC entities from the following six categories were extracted in this research: (1) relationship entities, (2) object-related entities, (3) material-related entities, (4) property-related entities, (5) unit-related entities, and (6) geometry-related entities. The specific IFC entities in each category are summarized in Table 3. If an IFC entity was an abstract entity (i.e., which is a non-instantiable entity), all instances of all its non-abstract subentities were extracted. Table 4 shows an example that includes two entity instances that were extracted in the intermediate representation, which are related to one regulatory requirement.

A special processing method was used in extracting instances of the property-related entities IfcPreDefinedPropertySet and IfcSimpleProperty. If the "Unit" attribute value of an instance was not available, the IfcValue measure type for the "NominalValue" attribute of that instance was used as a temporary unit. The measure type was used to deduce an explicit unit in the next post-processing step (see Section 5.1.3). For example, Table 4 shows that the measure type "IFCTHERMALRESISTANCEMEASURE" was used as a temporary unit for the property instance "Thermal Resistance", because the "Unit" attribute value for the IfcPropertySingleValue instance is "$" (i.e., not available).

5.1.3. Post-processing of extracted information

Post-processing was conducted to transform the intermediately-represented information to an alignment-ready representation, in which irrelevant design information (e.g., GlobalId, OwnerHistory) is filtered out to avoid unnecessary future processing efforts. After post-processing, the design information was represented in a number of hashmaps. These hashmaps were categorized into four groups according to their roles in information alignment: (1) the hashmaps "relationship", "product bounding box", and "spatially-contained product" were used for final complex alignment; (2) the hashmaps "material property", "object property", "product quantity", and "material layer-based property" contain the descriptors for property instances; (3) the hashmaps "object property", "object associated material information", "object entity information", "spatially-contained product", "material entity information", and "material category" contain the descriptors for object instances; and (4) the hashmaps "object ID-to-name", "object ID-to-entity name", "object entity name-to-ID", "relAggregates", "material ID-to-name", "material entity name-to-ID", and "material ID-to-object ID" are auxiliary hashmaps that were used to support the overall semantic information alignment. Table 5 provides examples to illustrate those hashmaps.

Special post-processing was conducted for measurement units, for the property instances in the following hashmaps: "material property", "object property", "product quantity", and "material layer-based property". Four post-processing methods were used to transform the IFC-represented units of the property instances into an intermediate representation for alignment to the units used in regulatory requirements. The intermediate representation: (1) is composed of one or multiple unit elements, (2) uses a 3-tuple representation – [prefix, name, exponent] – for each unit element, and (3) uses the International System of Units (SI) unit system, where each unit was represented using only the SI base unit (i.e., meter, kilogram, second, ampere, kelvin, candela, and mole). The four methods that were used to post-process the unit instances into the 3-tuple representation correspond to the four types of unit instances that were considered in this research: IfcSIUnit, IfcDerivedUnit, IfcConversionBaseUnit, and the temporary unit IfcValue measure type (see Section 5.1.2). The details of the intermediate representation and post-processing methods for measurement units are outside the scope of

Table 4
Examples of extracted entity instances in the intermediate representation.

IFC entity instance in .ifc file	IFC entity instance in the intermediate representation			
	IFC entity name	IFC entity instance ID	IFC entity attribute names	IFC entity attribute values
#4940907 = IFCDUCTSEGMENT ('3GeO_XJYnD0R2X24BqtWKm',#42, 'Oval Duct:Oval Duct – 2":9746348',$, 'Oval Duct:Oval Duct – 2":9746591', #4940833,#4940902, '9746348',$);	IFCDUCTSEGMENT	#4940907	GlobalId OwnerHistory Name Description ObjectType ObjectPlacement Representation Tag PredefinedType	3GeO_XJYnD0R2X24BqtWKm IFCOWNERHISTORY#42 Oval Duct:Oval Duct – 2":9746348 $(N/A) Oval Duct:Oval Duct – 2":9746591 IFCLOCALPLACEMENT#4940833 IFCPRODUCTDEFINITIONSHAPE#4940902 9746348 $(N/A)
#4940943 = IFCPROPERTYSINGLEVALUE('Thermal Resistance',$,IFCTHERMALRESISTANCEMEASURE (1.7611016132736),$);	IFCPROPERTYSINGLEVALUE	#4940943	Name Description NominalValue Unit	Thermal Resistance $(N/A) 1.761101613273600 IFCTHERMALRESISTANCEMEASURE

Table 5
Examples of post-processed BIM design information represented in hashmap format.

Hashmap name	Key	Value	Example ({Key : Value})	Note
Object property	IfcProduct/ IfcTypeProduct/ IfcSystem instance ID	List of properties, each property is represented in: (Property set name, Property instance ID, Property name, Value, Unit)	{#4975576:Analytical Properties(Type), #4975907, Heat Transfer Coefficient (U), 0.374015748031496, {{null, kelvin, '-1'}, {kilo, gram, '1'},{null, second, '-3'}}}	Properties of an IfcTypeObject instance are assigned to each occurrence IfcObject instance of that type, based on the assignment rules specified in the IfcRelDefinesByType entity.
Object entity information	IfcProduct/ IfcTypeProduct instance ID	The entity attribute values of the Key IfcProduct/IfcTypeProduct instance	{#4975576: Basic Roof:EPDM – 4 1/2" – lsf 3:9773671, Basic Roof:EPDM – 4 1/2" – lsf 3, 9773671, notdefined}	Exclude values of the following entity attributes: 1. Ownerhistory, Globalid. 2. Entity attribute value is null. 3. Entity attribute value is other entity instances.

this paper.

5.2. Regulatory information extraction

The regulatory information was extracted from the energy regulatory documents using the proposed ontology-based information extraction method in the authors' previous work [65]. The extracted regulatory information (i.e., a regulatory requirement or exception) was represented using nine semantic information elements (SIEs), including "subject", "subject restriction", "compliance checking attribute", "deontic operator indicator", "quantitative relation", "comparative relation", "quantity value", "quantity unit/reference", and "quantity restriction" [65]. "Subject" refers to the primary entity of a requirement or exception, which corresponds to an ontological concept. "Compliance checking attribute" refers to a property of a "subject", which corresponds to a concept. "Quantity value" refers to a quantitative measurement, which corresponds to a numeric value. "Quantity unit/reference" refers to either an explicit measurement unit or an implicit reference unit, which corresponds to SI unit symbols. "Subject restriction" and "quantity restriction" refer to constraints placed on a "subject" and "quantity value", respectively, which correspond to one or multiple concepts and relationships. For further details, the readers are referred to [65]. Table 6 shows examples of the SIEs for a regulatory requirement and a regulatory exception.

5.3. Semantic information alignment

The semantic information alignment was conducted to align the BIM information to the SIE-represented regulatory requirements and exceptions, following the method described in Section 4. The object concept interpretation and matching was conducted, as per Section 4.1.1.1, to recognize the candidate BIM object instances (candidate matches) for the regulatory object concepts. These object concepts refer to "subjects" and to objects referenced in the "subject restrictions" and "quantity restrictions", where object concepts that belong to one SIE tuple (i.e., SIEs that belong to one requirement or exception) form one concept group. The property recognition was conducted, as per Section 4.1.1.2, to recognize the candidate BIM property instances (candidate matches) for the regulatory property concepts. These property concepts refer to "compliance checking attributes" and to properties referenced in the "subject restrictions", "quantity units/references", and "quantity restrictions". The semantic similarity analysis was conducted, as per Section 4.1.2, to select the matches from the candidates. The instance pairs were then identified from the matches using the supervised and the unsupervised searching, and then grouped and linked to the requirements using the network construction, as per Sections 4.2.1 to 4.2.3. The measurement units in the "quantity units/references" were converted to the 3-tuple representation using a set of conversion rules.

The proposed semantic information alignment method was implemented in a Java-based platform. The platform used the following public APIs to accomplish some specific tasks. The JSDAI API [73] was used for EXPRESS-based BIM information extraction. The bSDD API [14] was used to search the bSDD for finding the matching IFC entities (or entity parents), synonyms of bSDD concepts, and superconcept information. The Protégé [74] was used to build the commercial building energy ontology, while the Apache Jena Ontology API [75] was used for parsing the ontology to find superconcepts and equivalent concepts. The deeplearning4j API [76], a Java implementation of Word2vec, was used for learning and computing all term-to-term similarities. The Stanford CoreNLP API [77] was used for morphological analysis.

5.4. Evaluation

A gold standard was manually built to test and evaluate the proposed semantic information alignment method. The gold standard includes 33 regulatory requirements and 10 regulatory exceptions (from the three regulatory documents), and 744 corresponding matches (i.e., BIM instances in the Revit model). The types of matches and their numbers are shown in Table 7.

Recall and precision were used to calculate the alignment performance. Recall was calculated as the total number of correctly aligned instances over the total number of instances in the gold standard. Precision was calculated as the total number of correctly aligned instances over the total number of instances aligned. To further test the statistical significance of the results, the confidence interval (p) was calculated for both recall and precision using the Wilson score without continuity correction [78,79], which is a computationally simple and satisfactory method for measuring confidence intervals [80]. Equation (10) [79] was used, where p_0 refers to the values of precision or recall, λ is the critical value for the confidence interval, n refers to either the total number of

Table 6
Examples of semantic information elements (SIEs) for regulatory requirement and exception.

Sentence type	Subject	Subject restriction	Compliance checking attribute	Deontic operator indicator	Quantitative relation	Comparative relation	Quantity value	Quantity unit/ reference	Quantity restriction
Regulatory requirement	supply and return air ducts	where located outside the building	R-value	shall	insulated	minimum	R-8	N/A	N/A
Regulatory exception	strainers, control valves, and balancing valves	associated with piping 1 in. or less in diameter	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 7
Performance of aligning BIM instances to semantic information elements (SIEs).

Total number of BIM instances	Subject	Subject restriction	Compliance checking attribute	Deontic operator indicator[a]	Quantitative relation[a]	Comparative relation[a]	Quantity value	Quantity unit/ reference	Quantity restriction	Total
In gold standard	176	133	159	N/A	N/A	N/A	138	138	0	744
Aligned	190	109	159	N/A	N/A	N/A	138	138	0	734
Correctly aligned	174	109	140	N/A	N/A	N/A	134	138	0	695
Precision	91.6%	100.0%	88.1%	N/A	N/A	N/A	97.1%	100.0%	N/A	94.7%
Recall	98.9%	82.0%	88.1%	N/A	N/A	N/A	97.1%	100.0%	N/A	93.4%

[a]SIE that is only used for representation of regulatory information.

instances aligned (for calculating p of precision) or the total number of instances in the gold standard (for calculating p of recall).

$$p = \frac{p_0 + t/2}{1+t} \pm \frac{\sqrt{p_0 q_0 t + t^2/4}}{1+t}, where\ q_0 = 1 - p_0, t = \lambda^2/n \tag{10}$$

6. Experimental results and analysis

6.1. Overall performance results

The overall performance results are summarized in Table 7. An overall performance of 93.4% recall [with confidence interval as (91.4%, 95.0%) at 95% confidence level] and 94.7% precision [with confidence interval as (92.8%, 96.1%) at 95% confidence level] was achieved. This indicates that the proposed semantic information alignment method is promising.

An example of the implementation of the proposed semantic information alignment method is shown in Fig. 13, where the BIM instance "IfcDuctSegment#4940907" is a match to the regulatory object concept "supply and return air duct" (an SIE "subject"), because the duct instance has the property value "supply air" (i.e., it is a supply air duct), "IfcBuilding#163" is a match to the object concept "building" (an object referenced in the SIE "subject restriction"), "Thermal Resistance" is a match to the property concept "R-value" (an SIE "compliance checking attribute"), and both the unit of "Thermal Resistance" and the unit referring to the SIE "quantity unit" are represented in the 3-tuple representation.

6.2. Error analysis for first-level simple alignment

Three main sources of errors (with approximate error percentages shown in parentheses) in first-level simple alignment were identified: ambiguity of regulatory concepts (15%), noise of BIM instances (20%), and errors in semantic similarity analysis (65%). Concepts may be ambiguously expressed in the text; they may not be easily mapped to the BIM instances [1], which may result in incorrect or missing recognition of matches. For example, (1) the property concept "diameter" was ambiguously stated in sentence S1, which resulted in recognizing both the property instances "outside diameter" and "inside diameter" as matches (only one is correct); and (2) the object concept "conditioned space" in S2 has a special definition in the energy regulatory documents, which resulted in failure of interpretation using only the concept descriptors to find the matches.

- S1: "…associated with piping 1 in. or less in diameter." [66]

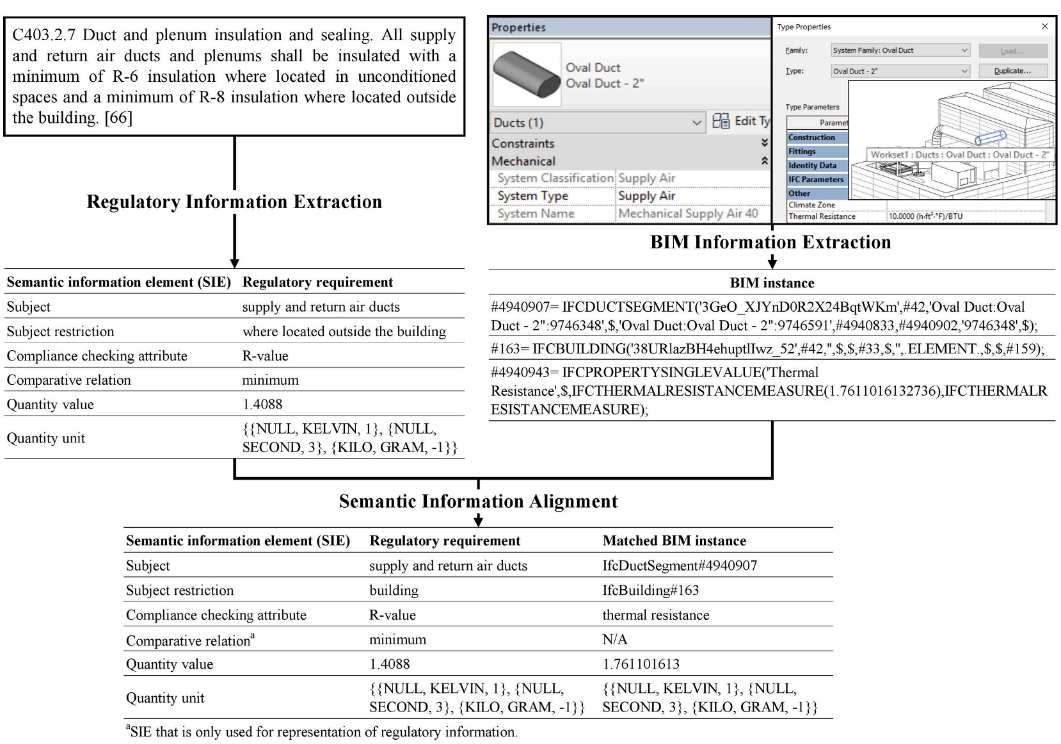


Fig. 13. Example of matched BIM information to a regulatory requirement.

- S2: “An area or room within a building being heated or cooled, containing uninsulated ducts, or with a fixed opening directly into an adjacent conditioned space.” [66]

The noise in BIM instances may result in incorrectly recognizing the matches for the regulatory concepts. For example, two BIM instances for the concept “balancing valve” were incorrectly recognized as matches for the concept “control valve”, because their instance descriptors – “Adjusting/Controlling Valves for Liquid Services” – share two common adjacent terms “controlling valves”, which resulted in a large-enough degree of similarity to the concept “control valve”.

The errors in semantic similarity analysis come from the semantic similarity scoring function and term position weighting function. First, the semantic similarity scoring function excludes the rare terms (i.e., term with a frequency less than five) in calculating the total instance-concept semantic similarity to indicate the matching, which may result in failing to distinguish incorrect matches. For example, both “aged solar reflectance index” and “solar reflectance index” were incorrectly recognized as matches to the concept “solar reflectance”, because the rare terms “aged” and “index” were excluded, which resulted in equivalent total instance-concept similarities to that of the correct match “solar reflectance”. This indicates that some rare terms may also carry meaningful information in the energy domain. Second, the term position weighting function overweights the rightmost terms in a concept, resulting in giving those terms excessive power to indicate the matching. For example, the BIM instance “Glass-Selux-Heat Tempered Convex Lens” was incorrectly recognized as a match to the concept “radiant heating”, because the common term “heat” is the rightmost term in the concept and was therefore given an overweighted common-term similarity to indicate the matching.

6.3. Error analysis for final complex alignment

Two main sources of errors in final complex alignment were identified. First, supervised searching may fail to identify the instance pairs created by the spatially-contained relationships that are captured using the bounding box geometric assessment, because some object instances may not be represented in bounding boxes. For example, two IfcSlab instances were not linked to two IfcSpace instances as instance pairs by the bounding box geometric assessment, because the two IfcSlab instances were represented in the swept solid geometric representation rather than bounding boxes.

Second, unsupervised searching may not recognize the correct instance pairs, because the relationships that are found by searching may not match the relationship that creates a concept pair. For example, the unsupervised searching incorrectly recognized the IfcDuctSegment (#4940907) and the IfcBuilding (#163) as an instance pair corresponding to the concept pair “supply and return air duct” and “building”. The relationship that creates the concept pair is “the duct is outside the building” (as per S3), but instance #4940907 is inside instance #163.

- S3: “C403.2.7 Duct and plenum insulation and sealing. All supply and return air ducts and plenums shall be insulated with a minimum of R-6 insulation where located in unconditioned spaces and a minimum of R-8 insulation where located outside the building.” [66]

7. Limitations and future work

Four limitations related to the proposed semantic information alignment method are acknowledged. First, the proposed concept interpretation and matching method may not be able to correctly interpret the meaning of all single regulatory concepts automatically. Typically, some regulatory concepts are ambiguously expressed in the regulations, which requires human experts to clarify such ambiguities and specify the mapping of the BIM instances to those ambiguous concepts [2,81]. Future research may further study such cases and explore the limit of artificial intelligence techniques in dealing with such ambiguities. Second, the proposed semantic similarity scoring method excluded the rare terms (i.e., terms with a frequency less than five) in assessing the term-to-term semantic similarity, because the deep learning technique was not able to learn accurate vector representations for rare terms. The experimental results showed that some rare terms may still carry important information in the energy domain. To assess the semantic similarity between rare terms, knowledge-based semantic similarity measures (which rely on ontology or WordNet, in contrast to the corpus-based measures used in this research) may be explored in future research. Third, the term position weighting function weights terms by only considering their positions in concepts, which may not be accurate. In future work, more factors (e.g., part-of-speech tags and term frequencies) may be considered, and unsupervised machine learning techniques may be explored to automatically learn the term position weighting function. Fourth, the unsupervised searching method made an oversimplified assumption that the shortest among all found paths determines if two BIM instances are linked as an instance pair. The experimental results showed that the relationships occurred on the shortest path may not correctly match the relationships between concepts. Therefore, further investigation may focus on how to select the one, among all found paths, that matches the relationship in a concept pair to create instance pairs.

Two limitations related to the implementation and testing are acknowledged. First, the implementation of the proposed information alignment method may be computationally expensive, especially in the post-processing of extracted information, and in the case of recognizing a large number of instance groups that are linked to a highly-complex concept group (e.g., the “subject restriction”, a complex SIE, may contain a large number of associated concepts, which may result in a complex concept group). For example, the computational expense may increase exponentially for constant bSDD online searching and unsupervised searching for recognizing a large number of complex instance groups. In future work, the bSDD could be entirely cached on local disks to avoid constant online searching, and techniques like parallel computing could be explored to improve the computational efficiency. Second, the proposed method was only tested on a limited number of energy regulatory requirements (related to two topics of energy requirements – thermal insulation and lighting power), because significant manual effort is needed for developing a gold standard. In future work, the authors plan to test the proposed method on more requirements in other energy topics (e.g., fenestration topic) from energy regulatory documents, and more energy requirements from other types of documents (e.g., contract specifications). The authors expect to achieve a similar level of performance, after some necessary adaptations. For example, the commercial building energy ontology may need extension to cover new concepts in other energy topics. The values of the proposed threshold types may also need adjustment for different types of documents. Such adaptation efforts should be significantly lower compared to the initial efforts. Third, the proposed method was only tested on quantitative requirements. Energy codes include both quantitative requirements and existential requirements (e.g., which require the existence of a building element). Existential requirements may contain different concepts and relationships. As a result, the alignment performance may vary. However, this variance is expected to be minimal, because the variability of concepts and relationships is relatively limited in energy codes. In future work, the authors plan to test the proposed method on such existential requirements and accordingly make necessary adaptations to the methods.

8. Conclusions and contributions

This paper presented a fully automated semantic information alignment method for aligning BIM information to regulatory information to support fully automated compliance checking. A first-level simple alignment method was proposed to align single BIM instances to single regulatory concepts, including concept interpretation and matching for interpreting the meaning of concepts to recognize the candidate matches, and semantic similarity analysis to select the matches. To recognize the instance groups that belong to one requirement or exception, a final complex alignment method was proposed, including supervised and unsupervised searching to identify the instance pairs, and network construction to group and link the instance pairs to the requirement or exception.

The proposed information alignment method was tested in aligning a set of BIM instances (extracted from an educational building model) to a number of commercial building energy efficiency requirements and exceptions (extracted from three energy regulatory documents). An overall performance of 93.4% recall and 94.7% precision was achieved. The experimental results indicate a number of conclusions. First, the proposed semantic information alignment method is promising. Second, errors in first-level simple alignment arise from failure to recognize the matches due to the ambiguity of regulatory concepts, noise of BIM instances, and errors in semantic similarity analysis. Third, errors in final complex alignment arise from some failures in recognizing spatially-related instance pairs or errors in the relationships found by searching.

This work contributes to the body of knowledge in three main ways. First, in comparison to existing information alignment efforts, this work offers a fully automated approach for alignment. Second, a novel method for first-level simple alignment is proposed to align single BIM instances to single regulatory concepts. The proposed method uses concept interpretation and matching to automatically recognize candidate matches to regulatory concepts. In the concept interpretation and matching, domain knowledge is captured in the form of ontology and bSDD and is used to support the interpretation of meaning of the regulatory concepts. The proposed method further uses semantic similarity analysis to select the matches to the regulatory concepts. In the semantic similarity analysis, a deep learning technique is used to explore the semantics behind words for enhanced assessment of semantic similarity, and an empirical way is used to analyze the patterns of semantic similarities for enhanced recognition of matches. Third, a novel method for final complex alignment method is proposed to recognize the associated regulatory concepts that belong to one requirement or exception, and group and link the matches (i.e., instance groups) to these associated regulatory concepts (i.e., concept group). The proposed method uses supervised and unsupervised searching to automatically search for the relationships that create the instance pairs, and uses network modeling to model a concept group and its associated instance groups as a network of linked concept pairs and instance pairs.

In future research – by the authors or the larger research community, the proposed method could also be used to support the development of computer-interpretable regulations. Developing such computable regulations requires the development and use of standardized regulatory concepts and relationships – for example in the form of a regulatory ontology. The proposed method could be used – along with other evaluation methods – to evaluate candidate regulatory concepts in the ontology in terms of their degree of match to the IFC concepts.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the National Science Foundation (NSF). This material is based upon work supported by NSF under Grant No. 1201170. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

References

[1] J.H. Garrett, M.E. Palmer, S. Demir, Delivering the infrastructure for digital building regulations, J. Comput. Civ. Eng. 28 (2) (2014) 167–169, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000369. ISSN 0887-3801.

[2] J. Dimyadi, P. Pauwels, R. Amor, Modelling and accessing regulatory knowledge for computer-assisted compliance audit, J. Inf. Technol. Constr. 21 (J. Inf. Technol. Constr. 21 (Special Issue CIB W78 2015 Special Track on Compliance Checking) (2016) 317–336, http://www.itcon.org/2016/21. ISSN 1874-4753.

[3] T.H. Beach, Y. Rezgui, H. Li, T. Kasim, A rule-based semantic approach for automated regulatory compliance in the construction sector, Expert Syst. Appl. 42 (12) (2015) 5219–5231, https://doi.org/10.1016/j.eswa.2015.02.029. ISSN 0957-4174.

[4] P. Pauwels, D. Van Deursen, R. Verstraeten, J. De Roo, R. De Meyer, R. Van De Walle, J. Van Campenhout, A semantic rule checking environment for building performance checking, Autom. Constr. 20 (5) (2011) 506–518, https://doi.org/10.1016/j.autcon.2010.11.017. ISSN 0926-5805.

[5] Y.C. Lee, C.M. Eastman, J.K. Lee, Automated rule-based checking for the validation of accessibility and visibility of a building information model, in: Proc. 2015 Intl. Workshop on Computing in Civil Engineering, ASCE, Reston, VA, 2015, pp. 572–579, 10.1061/9780784413616.067#sthash.gTzjtSUH.dpuf. ISBN 978-0-7844-7924-7.

[6] J. Zhang, N.M. El-Gohary, Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking, Autom. Constr. 73 (2017) 45–57, https://doi.org/10.1016/j.autcon.2016.08.027. ISSN 0926-5805.

[7] W. Solihin, C. Eastman, A knowledge representation approach in BIM rule requirement analysis using the conceptual graph, J. Inf. Technol. Constr. 21 (J. Inf. Technol. Constr. 21 (Special Issue CIB W78 2015 Special Track on Compliance Checking) (2016) 370–401, http://www.itcon.org/2016/24. ISSN 1874-4753.

[8] C. Eastman, Y. Jeong, R. Sacks, I. Kaner, Exchange model and exchange object concepts for implementation of national BIM standards, J. Comput. Civ. Eng. 24 (1) (2010) 25–34, https://doi.org/10.1061/(ASCE)0887-3801(2010)24:1(25). ISSN 0887-3801.

[9] C. Eastman, P. Teicholz, R. Sacks, K. Liston, BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors, John Wiley and Sons, Hoboken, NJ, 2011. ISBN 978-0-470-54137-1.

[10] buildingSMART, buildingSMART - International Home of openBIM, 2020, http://www.buildingsmart-tech.org/ (July 23, 2020).

[11] ISO (International Organization for Standardization), ISO 10303-11:1994, Industrial Automation Systems and Integration – Product Data Representation and Exchange – Part 11: Description Methods: the EXPRESS Language Reference Manual, 1994, http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=18348 (July 23, 2020).

[12] ISO (International Organization for Standardization), ISO 16739:2018, Industry Foundation Classes (IFC) for Data Sharing in the Construction and Facility Management Industries – Part 1: Data Schema, 2018, https://www.iso.org/standard/70303.html (July 23, 2020).

[13] T. Liebich, Y. Adachi, J. Forester, J. Hyvarinen, S. Richter, T. Chipman, M. Weise, J. Wix, Industry Foundation Classes Version 4.3 – RC1, 2020, https://standards.buildingsmart.org/IFC/DEV/IFC4_3/RC1/HTML/ (July 23, 2020).

[14] buildingSMART, buildingSMART Data Dictionary, 2016, http://bsdd.buildingsmart.org/#peregrine/about (July 23, 2020).

[15] buildingSMART, bSDD Content Guidelines, 2017, https://docs.google.com/document/d/1YUiR07A27lK0UB8ImYoaoLKCUvh1QFG1FfcvvLOYdP0/edit (July 23, 2020).

[16] T. Gerrish, K. Ruikar, M. Cook, M. Johnson, M. Phillip, C. Lowry, BIM application to building energy performance visualisation and management: challenges and potential, Energy Build. 144 (1) (2017) 218–228, https://doi.org/10.1016/j.enbuild.2017.03.032. ISSN 0378-7788.

[17] L.Y. Ding, B.T. Zhong, S. Wu, H.B. Luo, Construction risk knowledge management in BIM using ontology and semantic web technology, Saf. Sci. 87 (2016) 202–213, https://doi.org/10.1016/j.ssci.2016.04.008. ISSN 0925-7535.

[18] H. Liu, M. Lu, M. Al-Hussein, Ontology-based semantic approach for construction-oriented quantity take-off from BIM models in the light-frame building industry, Adv. Eng. Informatics. 30 (2) (2016) 190–207, https://doi.org/10.1016/j.aei.2016.03.001. ISSN 1474-0346.

[19] M.P. Nepal, S. Staub-French, R. Pottinger, J. Zhang, Ontology-based feature modeling for construction information extraction from a building information model, J. Comput. Civ. Eng. 27 (5) (2013) 555–569, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000230. ISSN 0887-3801.

[20] S. Sinha, A. Sawhney, A. Borrmann, F. Ritter, Extracting information from building information models for energy code compliance of building envelope, in: Proc. RICS COBRA Conf. 2013, CIB, Rotterdam, Netherlands, 2013, pp. 1–16, 10.1.1.727.4733.

[21] C. Preidel, A. Borrmann, Automated code compliance checking based on a visual language and building information modeling, in: Proc. 32nd Intl. Symp. Autom. Robot. Constr., Oulu, Finnland, 2015, pp. 256–263, 10.13140/RG.2.1.1542.2805. ISBN 978-1-5108-0924-6.

[22] L. Zhang, R. Issa, Ontology-based partial building information model extraction, J. Comput. Civ. Eng. 27 (6) (2013) 576–584, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000277. ISSN 0887-3801.

[23] M.K. Kim, Q. Wang, J.W. Park, J.C.P. Cheng, H. Sohn, C.C. Chang, Automated dimensional quality assurance of full-scale precast concrete elements using laser scanning and BIM, Autom. Constr. 72 (2016) 102–114, https://doi.org/10.1016/j.autcon.2016.08.035. ISSN 0926-5805.

[24] Y. Deng, J.C.P. Cheng, C. Anumba, Mapping between BIM and 3D GIS in different levels of detail using schema mediation and instance comparison, Autom. Constr. 67 (2016) 1–21, https://doi.org/10.1016/j.autcon.2016.03.006. ISSN 0926-5805.
[25] T.A. Teo, K.H. Cho, BIM-oriented indoor network model for indoor and outdoor combined route planning, Adv. Eng. Informatics 30 (3) (2016) 268–282, https://doi.org/10.1016/j.aei.2016.04.007. ISSN 1474-0346.
[26] M. Belsky, R. Sacks, I. Brilakis, Semantic enrichment for building information modeling, Comput. Civ. Infrastruct. Eng. 31 (4) (2016) 261–274, https://doi.org/10.1111/mice.12128. ISSN 1467-8667.
[27] Y.H. Lin, Y.S. Liu, G. Gao, X.G. Han, C.Y. Lai, M. Gu, The IFC-based path planning for 3D indoor spaces, Adv. Eng. Informatics 27 (2) (2013) 189–205, https://doi.org/10.1016/j.aei.2012.10.001. ISSN 1474-0346.
[28] G.N. Lilis, G.I. Giannakis, D.V. Rovas, Automatic generation of second-level space boundary topology from IFC geometry inputs, Autom. Constr. 76 (2016) 108–124, https://doi.org/10.1016/j.autcon.2016.08.044. ISSN 0926-5805.
[29] H. Kim, K. Anderson, Energy modeling system using building information modeling open standards, J. Comput. Civ. Eng. 27 (3) (2013) 203–211, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000215. ISSN 0887-3801.
[30] K. Kim, G. Kim, D. Yoo, J. Yu, Semantic material name matching system for building energy analysis, Autom. Constr. 30 (2013) 242–255, https://doi.org/10.1016/j.autcon.2012.11.011. ISSN 0926-5805.
[31] H. Kim, Z. Shen, I. Kim, K. Kim, A. Stumpf, J. Yu, BIM IFC information mapping to building energy analysis (BEA) model with manually extended material information, Autom. Constr. 68 (2016) 183–193, https://doi.org/10.1016/j.autcon.2016.04.002. ISSN 0926-5805.
[32] A. Cemesova, C.J. Hopfe, R.S. McLeod, PassivBIM: Enhancing interoperability between BIM and low energy design software, Autom. Constr. 57 (2015) 17–32, https://doi.org/10.1016/j.autcon.2015.04.014. ISSN 0926-5805.
[33] J. Won, G. Lee, C. Cho, No-schema algorithm for extracting a partial model from an IFC instance model, J. Comput. Civ. Eng. 27 (6) (2013) 585–592, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000320. ISSN 0887-3801.
[34] L. Zhang, X. Wu, L. Ding, M.J. Skibniewski, Y. Lu, BIM-based risk identification system in tunnel construction, J. Civ. Eng. Manag. 22 (4) (2016) 529–539, https://doi.org/10.3846/13923730.2015.1023348. ISSN 1392-3730.
[35] H. Kim, K. Anderson, S. Lee, J. Hildreth, Generating construction schedules through automatic data extraction using open BIM (building information modeling) technology, Autom. Constr. 35 (2013) 285–295, https://doi.org/10.1016/j.autcon.2013.05.020. ISSN 0926-5805.
[36] S.K. Lee, K.R. Kim, J.H. Yu, BIM and ontology-based approach for building cost estimation, Autom. Constr. 41 (2014) 96–105, https://doi.org/10.1016/j.autcon.2013.10.020. ISSN 0926-5805.
[37] I.H. Hijazi, M. Ehlers, S. Zlatanova, NIBU: A new approach to representing and analysing interior utility networks within 3D geo-information systems, Intl. J. Digit. Earth 5 (1) (2012) 22–42, https://doi.org/10.1080/17538947.2011.564661. ISSN 1753-8947.
[38] J. Choi, J. Choi, I. Kim, Development of BIM-based evacuation regulation checking system for high-rise and complex buildings, Autom. Constr. 46 (2014) 38–49, https://doi.org/10.1016/j.autcon.2013.12.005. ISSN 0926-5805.
[39] J. Dimyadi, C. Clifton, M. Spearpoint, R. Amor, Computerizing regulatory knowledge for building engineering design, J. Comput. Civ. Eng. 30 (5) (2016) C4016001, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000572. ISSN 1392-3730.
[40] Y. Bengio, Learning deep architectures for AI, Found. Trends® Mach. Learn. 2 (1) (2009) 1–127, 10.1561/2200000006. ISSN 1935-8237.
[41] J.C.P. Cheng, M. Das, A BIM-based web service framework for green building energy simulation and code checking, J. Inf. Technol. Constr. 19 (2014) 150–168, http://www.itcon.org/2014/8. ISSN 1874-4753.
[42] A. Yurchyshyna, C. Faron-Zucker, N.L. Thanh, A. Zarli, Towards an ontology-enabled approach for modeling the process of conformity checking in construction, in: Proc. CAiSE'08 Forum, RWTH Aachen University, Aachen, Germany, 2008, pp. 21–24, http://ceur-ws.org/Vol-344/paper6.pdf. ISSN 1613-0073.
[43] R.K. Dhillon, M. Jethwa, H.S. Rai, Extracting building data from BIM with IFC, Intl. J. Recent Trends Eng. Technol. 11(1) (2014) 202–210, 01.IJRTET.11.1.1500. ISSN 2158-5563.
[44] N. Nawari, Automating codes conformance, J. Archit. Eng. 18 (4) (2012) 315–323, https://doi.org/10.1061/(ASCE)AE.1943-5568.0000049. ISSN 1076-0431.
[45] X. Tan, A. Hammad, P. Fazio, Automated code compliance checking for building envelope design, J. Comput. Civ. Eng. 24 (2) (2010) 203–211, https://doi.org/10.1061/(ASCE)0887-3801(2010)24:2(203). ISSN 1392-3730.
[46] B.T. Zhong, L.Y. Ding, P.E.D. Love, H.B. Luo, An ontological approach for technical plan definition and verification in construction, Autom. Constr. 55 (2015) 47–57, https://doi.org/10.1016/j.autcon.2015.02.002. ISSN 0926-5805.
[47] K. Baumgärtel, M. Kadolsky, R.J. Scherer, An ontology framework for improving building energy performance by utilizing energy saving regulations, in: Proc. 10th European Conf. on Product & Process Modelling (ECPPM 2014), Taylor & Francis Group, London, 2015, pp. 519–526, 10.1201/b17396-86. ISBN 978-1-138-02710-7.
[48] B.T. Zhong, L.Y. Ding, H.B. Luo, Y. Zhou, Y.Z. Hu, H.M. Hu, Ontology-based semantic modeling of regulation constraint for automated construction quality compliance checking, Autom. Constr. 28 (2012) 58–70, https://doi.org/10.1016/j.autcon.2012.06.006. ISSN 0926-5805.
[49] J. Dimyadi, W. Solihin, D. Marchant, The design brief: requirements and compliance, J. Inf. Technol. Constr. 21 (J. Inf. Technol. Constr. 21 (Special Issue CIB W78 2015 Special Track on Compliance Checking) 337–353, http://www.itcon.org/2016/22. ISSN 1874-4753.
[50] J. Qi, R. Issa, S. Olbina, J. Hinze, Use of building information modeling in design to prevent construction worker falls., J. Comput. Civ. Eng. 28 (5) (2013) 1–10, 10.1061/%28ASCE%29CP.1943-5487.0000365. ISSN 1392-3730.
[51] A. Yurchyshyna, A. Zarli, An ontology-based approach for formalisation and semantic organisation of conformance requirements in construction, Autom. Constr. 18 (2009) 1084–1098, https://doi.org/10.1016/j.autcon.2009.07.008. ISSN 0926-5805.
[52] J. Melzner, S. Zhang, J. Teizer, H.J. Bargstädt, A case study on automated safety compliance checking to assist fall protection design and planning in building information models, Constr. Manag. Econ. 31 (6) (2013) 661–674, https://doi.org/10.1080/01446193.2013.780662. ISSN 0144-6193.
[53] S. Zhang, J. Teizer, J.K. Lee, C.M. Eastman, M. Venugopal, Building information modeling (BIM) and safety: automatic safety checking of construction models and schedules, Autom. Constr. 29 (2013) 183–195, https://doi.org/10.1016/j.autcon.2012.05.006. ISSN 0926-5805.
[54] T.H. Nguyen, J.L. Kim, Building code compliance checking using BIM technology, in: Proc. 2011 Winter Simul. Conf. IEEE, Washington, DC, 2011, pp. 3395–3400, 10.1109/WSC.2011.6148035. ISBN 978-1-4577-2108-3.
[55] H. Lee, J.K. Lee, S. Park, I. Kim, Translating building legislation into a computer-executable format for evaluating building permit requirements, Autom. Constr. 71 (Part 1) (2016) 49–61, 10.1016/j.autcon.2016.04.008. ISSN 0926-5805.
[56] E.A. Delis, A. Delis, Automatic fire-code checking using expert-system technology, J. Comput. Civ. Eng. 9 (2) (1995) 141–156, https://doi.org/10.1061/(ASCE)0887-3801(1995)9:2(141). ISSN 1392-3730.
[57] S.K. Goel, S.J. Fenves, Computer-Aided Processing of Structural Design Specifications, 1969, https://www.ideals.illinois.edu/bitstream/handle/2142/14777/SRS-348.pdf?sequence=2 (May 17, 2017).
[58] L. Ding, R. Drogemuller, M. Rosenman, D. Marchant, J. Gero, Automating code checking for building designs-DesignCheck, in: Proc. CRC for Construction, University of Wollongong, Wollongong, Australia, 2006, pp. 1–16, http://ro.uow.edu.au/engpapers/4842/.
[59] R. See, SMARTcodes: Enabling BIM Based Automated Code Compliance Checking, 2008, http://projects.buildingsmartalliance.org/files/?artifact_id=1535 (May 17, 2017).
[60] T. Liebich, J. Wix, J. Forester, Z. Qi, Speeding-up the building plan approval–the Singapore e-plan checking project offers automatic plan checking based on IFC, in: Proc. 4th Eur. Conf. Prod. Process Model., Swets & Zeitlinger, Lisse, 2002, pp. 467–471, ISBN 90-5809-507-X.
[61] SMC, Development of Solibri Model Checker (SMC) Accessibility Rules Set for checking Building Information Models, 2009, http://www.universell-utforming.miljo.no/file_upload/statsbygg_smc_accessibility.pdf (May 17, 2017).
[62] H. Luo, P. Gong, A BIM-based code compliance checking process of deep foundation construction plans, J. Intell. Robotic Syst. 79 (3–4) (2015) 549–576, https://doi.org/10.1007/s10846-014-0120-z. ISSN 0921-0296.
[63] C. Preidel, A. Borrmann, Towards code compliance checking on the basis of a visual programming language, 21 (J. Inf. Technol. Constr. 21 (Special Issue CIB W78 2015 Special Track on Compliance Checking) (2016) 402-421, http://www.itcon.org/2016/25. ISSN 1874-4753.
[64] C. Eastman, J. min Lee, Y. suk Jeong, J. kook Lee, Automatic rule-based checking of building designs, Autom. Constr. 18 (2009) 1011–1033. 10.1016/j.autcon.2009.07.002.
[65] P. Zhou, N. El-Gohary, Ontology-based automated information extraction from building energy conservation codes, Autom. Constr. 74 (2017) 103–117, https://doi.org/10.1016/j.autcon.2016.09.004. ISSN 0926-5805.
[66] ICC (International Code Council), 2012 International Energy Conservation Code, 2012, http://publicecodes.cyberregs.com/icod/iecc/2012/ (Mar. 21, 2015). ISBN 978-1-60983-058-8.
[67] California Energy Commission, 2013 Building Energy Efficiency Standards for Residential and Nonresidential Buildings, 2013, http://www.energy.ca.gov/2012publications/CEC-400-2012-004/CEC-400-2012-004-CMF-REV2.pdf (May 17, 2017).
[68] E.W. Dijkstra, A note on two problems in connexion with graphs, Numer. Math. 1 (1) (1959) 269–271, https://doi.org/10.1007/BF01386390. ISSN 0029-599X.
[69] Autodesk, Revit API Developers Guide, 2016, https://knowledge.autodesk.com/search-result/caas/CloudHelp/cloudhelp/2016/ENU/Revit-API/files/GUID-F0A122E0-E556-4D0D-9D0F-7E72A9315A42-htm.html (Aug. 3, 2016).
[70] Microsoft, Visual Studio IDE, 2016, https://www.visualstudio.com/ (May 10, 2017).
[71] T. Liebich, IFC4 – the new buildingSMART Standard, BuildingSMART, 2013, http://www.buildingsmart-tech.org/specifications/ifc-releases/ifc4-release/buildingSMART_IFC4_Whatisnew.pdf (May 17, 2017).
[72] J. Zhang, N. El-Gohary, Automated extraction of information from building information models into a semantic logic-based representation, in: Proc. 2015 Intl. Workshop on Comput. in Civ. Eng., ASCE, Reston, VA, 2015, pp. 173–180, 10.1061/9780784479247.022. ISBN 9780784479247.
[73] LKSoftWare GmbH, Java Standard Data Access Interface (JSDAI), 2016, http://www.jsdai.net/ (Nov. 7, 2016).
[74] M.A. Musen, The Protégé project: A look back and a look forward, AI Matters 1 (4) (2015) 4–12, https://doi.org/10.1145/2757001.2757003. ISSN 2372-3483.
[75] Apache Jena, Jena Ontology API, 2016, https://jena.apache.org/documentation/ontology/ (Nov. 7, 2016).
[76] Deeplearning4j Development Team, Deeplearning4j: Open-Source Distributed Deep Learning for the JVM, 2016, https://deeplearning4j.org/ (Nov. 7, 2016).
[77] C.D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S.J. Bethard, D. McClosky, The stanford CoreNLP natural language processing toolkit, in: Proc. 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations,

ACL, Stroudsburg, PA, 2014, pp. 55–60, 10.3115/v1/P14-5010. ISBN 978-1-941643-00-6.
[78] C. Goutte, E. Gaussier, A probabilistic interpretation of precision, recall and f-score, with implication for evaluation, in: Proc. European Conf. on Information Retrieval (ECIR'05), Springer, Berlin, 2005, pp. 345–359, 10.1007/978-3-540-31865-1_25.
[79] E. Wilson, Probable inference, the law of succession, and statistical inference, J. Am. Stat. Assoc. 22 (158) (1927) 209–212, https://doi.org/10.1080/01621459.1927.10502953. ISSN 0162-1459.
[80] R.G. Newcombe, Two-sided confidence intervals for the single proportion: comparison of seven methods, Stat. Med. 17 (8) (1998) 857–872, https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E. ISSN 1097-0258.
[81] W. Solihin, C. Eastman, Classification of rules for automated BIM rule checking development, Autom. Constr. 53 (2015) 69–82, https://doi.org/10.1016/j.autcon.2015.03.003. ISSN 0926-5805.
[82] Ontario Ministry of Municipal Affairs, Ontario Building Code Supplementary Standard SB-10, 2011, http://www.mah.gov.on.ca/AssetFactory.aspx?did=9227 (May 17, 2017).
[83] T.A. Teo, S.C. Yu, The extraction of indoor building information from BIM to OGC IndoorGML, in: Proc. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., Copernicus Publications, Göttingen, 2017, pp. 167–170, https://doi.org/10.5194/isprs-archives-XLII-4-W2-167-2017.
[84] N. Nawari, Generalized adaptive framework for computerizing the building design review process, J. Archit. Eng. 26 (1) (2020) 315–323, https://doi.org/10.1061/(asce)ae.1943-5568.0000382. ISSN 1076-0431.
[85] H. Kim, J.K. Lee, J. Shin, J. Choi, Visual language approach to representing KBimCode-based Korea building code sentences for automated rule checking, J. Comput. Des. Eng. 6 (2) (2019) 143–148, https://doi.org/10.1016/j.jcde.2018.08.002. ISSN 2288-5048.