BIM, NLP, and AI for Automated Compliance Checking

Ruichuan Zhang
PhD Candidate
University of Illinois at Urbana-Champaign
rzhang65@illinois.edu

Nora El-Gohary
Associate Professor
University of Illinois at Urbana-Champaign
gohary@illinois.edu

Abstract

The digital and integrated representation of the physical and functional characteristics of buildings enabled by building information modeling (BIM) provides a computational environment for automated compliance checking (ACC) of building designs. The integration of natural language processing (NLP) and artificial intelligence (AI) with BIM brings further opportunities for ACC – it can empower BIM with text analytics and AI capabilities, thereby injecting intelligence and automation in the compliance checking processes. This chapter highlights emerging approaches that aim to facilitate and harness the marriage of BIM, NLP, and AI to enable the next generation of automated compliance checking systems (ACC) systems. This chapter (1) reviews different types of BIM-based ACC systems that leverage NLP and AI techniques, (2) discusses how NLP and AI techniques are applied in regulatory text analytics tasks and BIM information analytics tasks in the context of ACC, and (3) discusses the future trends of BIM-based ACC systems.

Keywords: Automated compliance checking, Building information modeling, Natural language processing, Artificial intelligence.

Section 1 Introduction

Building designs must comply with a multitude of requirements from building codes, regulations, project specifications, etc. These requirements come from different authorities and cover a wide variety of topics such as energy, safety, and accessibility. Manually checking the compliance of a building design with all applicable requirements is costly, time consuming, and error prone. For example, in 2018, over $200 million were spent on, only, checking the compliance of the designs of new privately-owned housing units (US Census Bureau 2019). It takes 15 to 18 days to complete the review and checking of the design of a new residential building (City of Manassas 2019; Wisconsin Department of Safety and Professional Services 2019). For larger building projects (e.g., projects over $400,000 in construction valuation), the compliance checking process takes more time, with multiple checking cycles (City of Sacramento 2019). And, a study showed that about 29% of the manual checking has errors and inconsistencies (Fiatech 2012).

To reduce the time, cost, and errors of building code compliance checking, a number of automated code compliance checking (ACC) methods and systems have been developed and implemented, both in academia and industry. ACC systems are computational systems that process both the natural-language requirements and the digital building designs and analyze both to detect non-compliance instances (as illustrated in Fig. 1). Since the advent of building information modeling (BIM) technology, ACC systems have become BIM-based and have benefited from the more integrated and normalized design information representations brought by BIM. The goal of BIM-based ACC systems is to check the BIM-represented building design (e.g., the configuration and properties of building elements) for compliance with relevant requirements (e.g., requirements expressed in the building code).

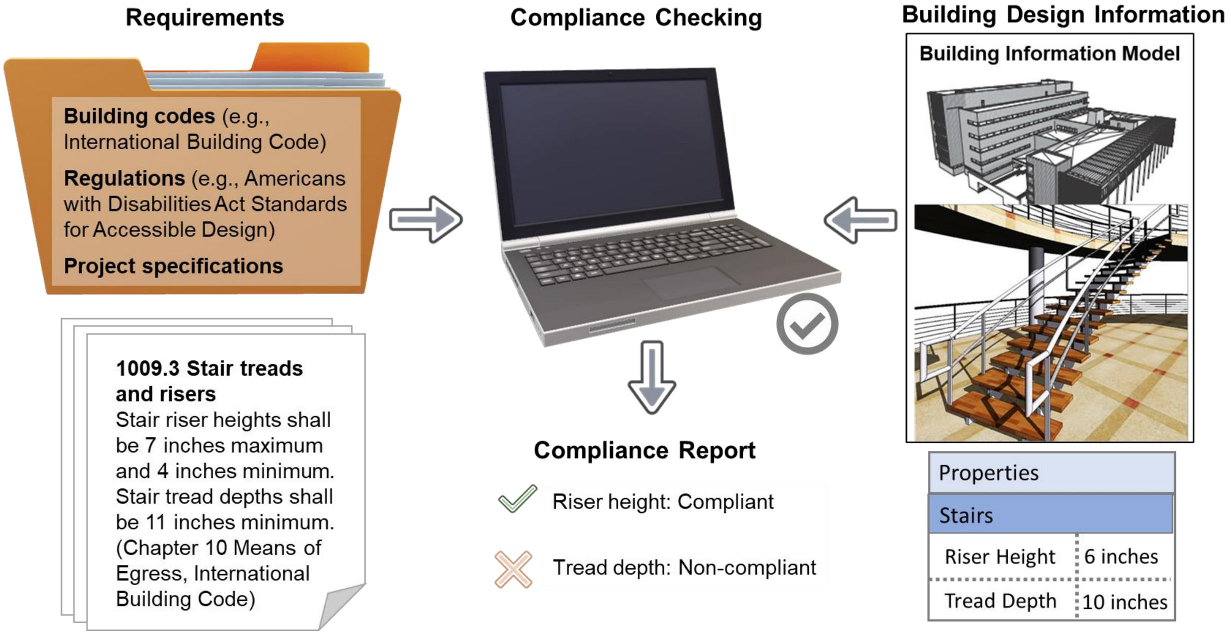

Requirements
Building codes (e.g., International Building Code)
Regulations (e.g., Americans with Disabilities Act Standards for Accessible Design)
Project specifications
1009.3 Stair treads and risers
Stair riser heights shall be 7 inches maximum and 4 inches minimum. Stair tread depths shall be 11 inches minimum. (Chapter 10 Means of Egress, International Building Code)
Compliance Checking
Compliance Report
Riser height: Compliant
Tread depth: Non-compliant
Building Design Information
Building Information Model
Properties
Stairs
Riser Height
6 inches
Tread Depth
10 inches

Fig. 1 BIM-based ACC systems.

Existing BIM-based ACC systems are, however, not entirely automated – still requiring several manual processes such as manual reading and interpretation of the building code and formalization of the requirements in a computer-processable form (e.g., rules). Fully automated ACC, instead, requires all processes to be automated, including the extraction of requirements (or rules) from the regulatory documents, the extraction of relevant design information from the BIM model, and the alignment of the semantic representations of both. In reality, however, achieving full automation is challenging, for three main reasons. First, regulatory documents are written in natural language, which are difficult for computer interpretation and can often be vague or ambiguous even to humans. For example, a requirement might be subject to more than one interpretation. Second, regulatory documents tend to have complex syntactic and semantic structures, such as hierarchically-complex clause and sentence structures including deeply nested syntactic and semantic structures, conjunctive and alternative obligations, and multiple exceptions. Third, the BIM models and the regulatory documents are largely speaking different languages, using different semantic representations and terminologies (Solihin and Eastman 2015; Zhou and El-Gohary 2017; Nawari 2019).

Despite the existence of these challenges, the rapidly advancing natural language processing (NLP) and artificial intelligence (AI) techniques are opening the doors for many new solutions. NLP is a field of linguistics and computer science that uses computational tools for computer systems to automatically process, analyze, and understand natural language data (e.g., text) (Goldberg 2017). AI is a field of computer science that develops computer systems capable to automatically interpret external data, learn from these data, and use the learnings to perform tasks that normally require human intelligence (Kaplan and Haenlein 2019). In ACC systems, NLP and AI techniques support text and BIM information analytics tasks including requirement classification, semantic annotation, regulatory information extraction, design information extraction, BIM semantic enrichment, BIM-regulatory information alignment, and compliance reasoning (as shown in Fig. 2). Requirement classification aims to classify natural-language requirements into predefined categories (e.g., relevant versus irrelevant requirements). Semantic annotation aims to annotate the requirements with markups that indicate the elements of the requirements (e.g., subject of compliance) and their meanings. Regulatory information extraction aims to extract the requirement information from the text (e.g., code) and represent the extracted information in a computer-processable form (e.g., rules). Design information extraction aims to extract relevant building design information from the BIM models. BIM semantic enrichment aims to add meaningful information to a BIM model. BIM-regulatory information alignment aims to align the meanings and representations of the requirements and the BIM models. And, compliance reasoning aims to analyze the aligned BIM-regulatory information to detect non-compliance instances.

Traditionally, NLP and AI were based on explicitly programmed rules and expert/knowledge systems, suffering from a lack of flexibility and adaptability. Recently, NLP and AI have been seeing an increasing adoption of machine learning (ML) techniques. ML techniques develop computational models that learn from data (i.e., training data) to make predictions/decisions, without the need for explicit programming (Alpaydin 2020). Researchers have already started – in the past several years – to explore how ML could help tackle the challenges in BIM-based ACC systems and further increase their flexibility and adaptability (e.g., Zhang and El-Gohary 2016; Ma et al. 2018; Xue and Zhang 2019; Bloch and Sacks 2020; Zhong et al. 2020).

This chapter aims to review different types of BIM-based ACC systems (Section 2), and how NLP and AI techniques have been applied in regulatory text analytics tasks (Section 3) and BIM information analytics tasks (Section 4) in the context of ACC, and ends with a discussion of the future of BIM-based ACC systems (Section 5) with an emphasis on the application of ML techniques. Overall, this chapter also illustrates how BIM, NLP, and AI can, together, drive innovation and convergence in the AEC domain.

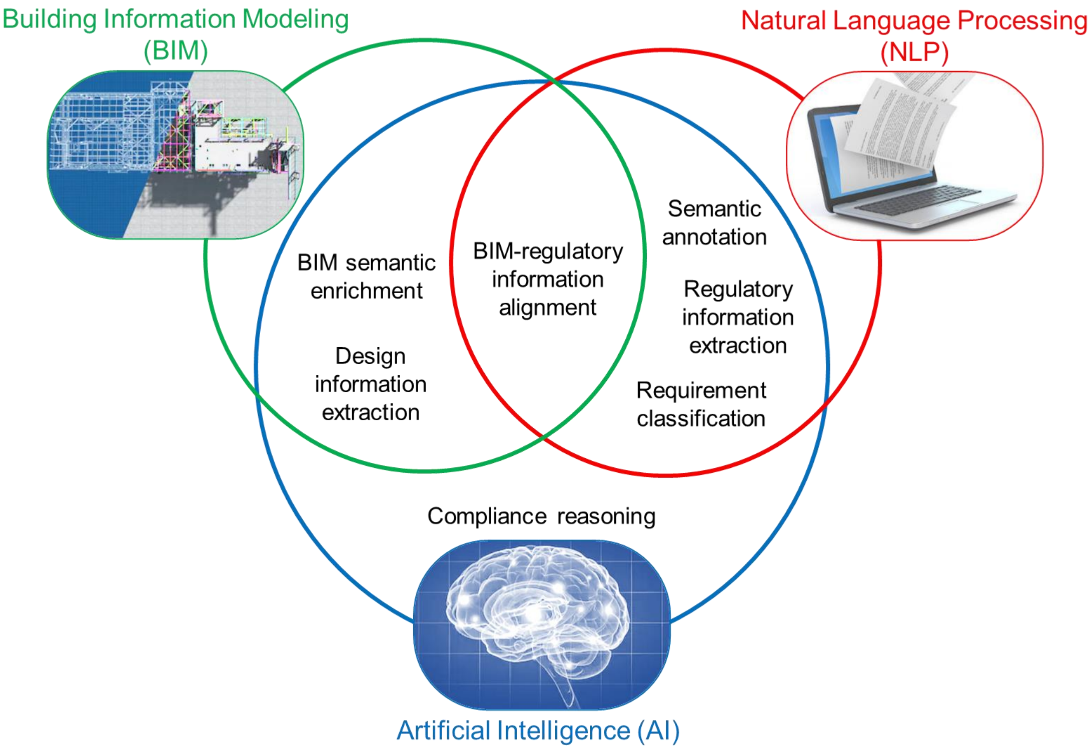

Building Information Modeling (BIM)
Natural Language Processing (NLP)
BIM semantic enrichment
Design information extraction
BIM-regulatory information alignment
Semantic annotation
Regulatory information extraction
Requirement classification
Compliance reasoning
Artificial Intelligence (AI)

Fig. 2 NLP and AI in BIM-based ACC systems.

Section 2 Existing BIM-based ACC Systems

2.1 *AI-assisted ACC Systems*

BIM-based ACC systems have been evolving from AI-assisted systems to AI-driven systems. AI-assisted systems predominately rely on human effort (e.g., extensive manual input and human interpretation of rules) for compliance checking, with a limited number of processes such as compliance reasoning being supported by AI techniques. Also, these systems mainly use symbolic AI techniques, where explicit representations composed of symbolic elements (e.g., constants, functions, and predicates) are manually designed by experts to represent implicit knowledge (Garnelo et al. 2016). The first symbolic AI techniques that were applied in (non-BIM-based) ACC systems are expert systems (e.g., Hayes-Roth et al. 1983). Since then, many research efforts on AI-assisted ACC systems have been undertaken, such as safety checking (e.g., Zhang et al. 2013, Choi et al. 2014), fire safety checking (e.g., Balaban et al. 2012, Kincelova et al. 2020), building accessibility checking (e.g., Lau and Law 2004, Lee et al. 2015), building performance checking (e.g., Pauwels et al. 2011), water distribution system checking (e.g., Martins and Monteiro 2013), and building design checking (e.g., Eastman et al. 2009, Tan et al. 2010). And many of the first commercialized or government-funded ACC systems are also AI-assisted such as CORENET ePlanCheck (AECBytes 2005), REScheck and COMcheck (US Department of Energy 2020), SMARTcodes (Government of Singapore 2016), and Solibri Model Checker (Solibri 2020). AI-assisted systems can be further classified into four main categories based on how the users interact with these systems, as shown in Fig. 3: White-box, semantic annotation-based, parametric template-based, and black-box systems.

White-box ACC systems. White-box systems require users (e.g., compliance checking professionals) to manually translate natural-language requirements into computer-processable forms using programming languages. For example, the Building Environment and Analysis Language (Lee 2011) was designed to represent building objects and their properties and relations contained in the requirements. The BIM Rule Language (Dimyadi et al. 2016) was designed to query BIM models using a Structured Query Language (SQL)-based syntax, and then rule checking algorithms are applied to the query results to perform compliance reasoning. Conceptual graphs (Solihin and Eastman 2016) and Visual Code Checking Language (Preidel and Borrmann 2016) were used to extract and represent the rules, constraints, building objects, and relationships between objects, and visualize the rules in graph-like structures. White-box ACC systems reveal most of the information representations to the users, but require users to know the syntax and vocabulary of the language used.

Semantic annotation-based ACC systems. Semantic annotation-based systems require manual annotation of the semantic concepts and relations that describe the requirements using semantic tags. For example, the users of the SMARTcodes software first read the building code, and then annotate the text with the requirement, applicability, selection, and exception (RASE) markups (Hjelseth and Nisbet 2010). Semantic annotation-based systems reveal partial information representations to the users – the semantic markups, which are typically simple to understand and use, but usually keep the post-processing processes of the annotated regulatory information and the compliance reasoning mechanisms hidden.

Parametric template-based ACC systems. Parametric template-based systems use predesigned rule templates, and require manual extraction of the regulatory requirements from the code by the users. For example, the users of Solibri Model Checker (Solibri 2020) need to read the natural-language requirements, identify the correct rule templates for the requirements, and obtain the values for the parameters of the templates from the requirements. Similar to semantic annotation-based ACC systems, parametric template-based systems only reveal partial information representations to the users.

Black-box ACC systems. Black-box systems are different from the previous three types of systems in that they are opaque to the users – the users have no control over the text and BIM information analytics processes, neither do they have knowledge of the representations nor the compliance reasoning mechanisms. Instead, the requirements have been already encoded using the chosen representation/language by ACC software developers; and users only have access to the input (i.e., the natural-language requirements and the BIM models) and the output (i.e., the compliance checking report) of the system. Most of the current commercialized or government-funded ACC systems are hardcoded black-box systems. The earliest ones among such systems are CORENET ePlanCheck (Government of Singapore 2016) and REScheck and COMcheck (US Department of Energy 2020), and the more recent ones include SMARTreview, UpCodes, Compliance Audit Systems Limited, Daima, and Invicara.

Even though AI-assisted ACC requires much less manual effort compared to manual compliance checking, the amount of manual effort needed is still significant. This manual effort is time consuming and could be a source of errors. For example, a complex requirement that includes multiple conjunctions and/or disjunctions and restrictions and/or exceptions could take 30 minutes or more to input into the rule templates of the Solibri Model Checker. Another example is REScheck – it takes one to three business days to generate a compliance checking report using this software (REScheck EZ 2020).

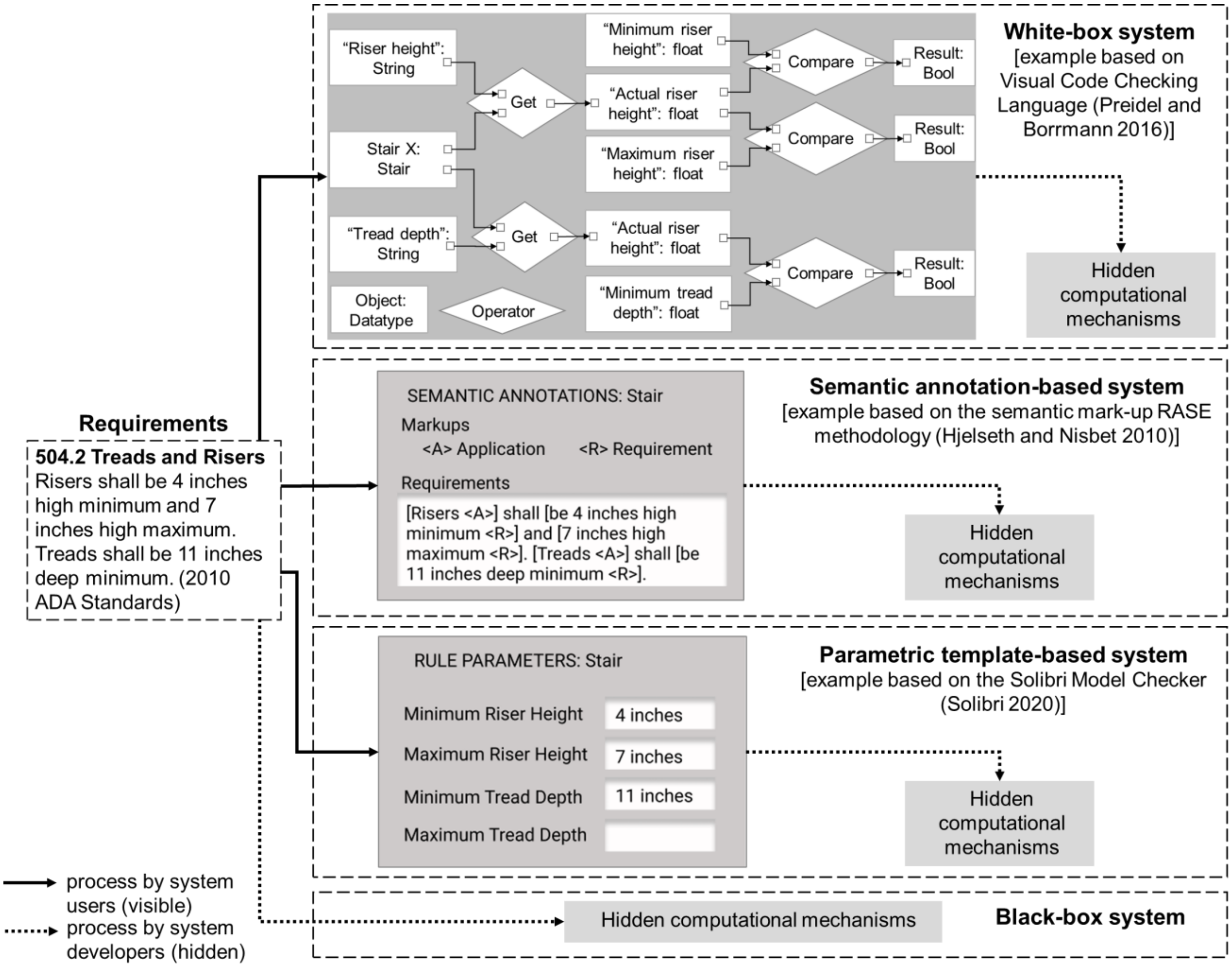

White-box system
[example based on Visual Code Checking Language (Preidel and Borrmann 2016)]
"Riser height": String
Stair X: Stair
"Tread depth": String
Object: Datatype
Get
Get
Operator
"Minimum riser height": float
"Actual riser height": float
"Maximum riser height": float
"Actual riser height": float
"Minimum tread depth": float
Compare
Compare
Compare
Result: Bool
Result: Bool
Result: Bool
Hidden computational mechanisms
Requirements
504.2 Treads and Risers
Risers shall be 4 inches high minimum and 7 inches high maximum. Treads shall be 11 inches deep minimum. (2010 ADA Standards)
Semantic annotation-based system
[example based on the semantic mark-up RASE methodology (Hjelseth and Nisbet 2010)]
SEMANTIC ANNOTATIONS: Stair
Markups
<A> Application
<R> Requirement
Requirements
[Risers <A>] shall [be 4 inches high minimum <R>] and [7 inches high maximum <R>]. [Treads <A>] shall [be 11 inches deep minimum <R>].
Hidden computational mechanisms
Parametric template-based system
[example based on the Solibri Model Checker (Solibri 2020)]
RULE PARAMETERS: Stair
Minimum Riser Height
4 inches
Maximum Riser Height
7 inches
Minimum Tread Depth
11 inches
Maximum Tread Depth
Hidden computational mechanisms
Black-box system
Hidden computational mechanisms
process by system users (visible)
process by system developers (hidden)

Fig. 3 Comparison of AI-assisted ACC systems.

2.2 AI-driven ACC Systems

AI-driven systems rely heavily on AI, with most of the complex ACC processes such as regulatory information extraction benefiting from AI techniques. Also, different from AI-assisted systems, they rely on NLP and ML techniques instead of symbolic AI. AI-driven systems are, thus, more flexible and scalable and have greater potential to be adapted to different types of regulatory documents in the AEC domain. They can be further classified into two main categories: rule-based systems and ML-based systems.

Rule-based ACC systems. Rule-based systems are built on expert-defined rules. For example, as shown in Fig. 4, the semantic NLP-based ACC (SNACC) system (Zhang and El-Gohary 2017a, 2017b) uses semantic modeling, NLP, and information extraction rules to automatically extract the regulatory information from the natural-language requirements and the design information from the BIM models, and formalizes both into the same form (first-order logic) for compliance reasoning. The rules define patterns of syntactic and semantic text features, and use pattern matching to identify the information to extract based on the recognized text patterns. Based on the aforementioned work, Zhou and El-Gohary (2017) and Li et al. (2016) further developed rule-based systems for energy compliance checking and utility spatial compliance checking, respectively. Compared to AI-assisted systems, rule-based systems are more flexible and scalable, offering a higher level of automation with minimal manual effort (e.g., the effort required by system developers to develop the regulatory information extraction rules). However, as all other types of AI-driven systems, rule-based ACC systems have their intelligence and automation limits. For example, these systems are limited in dealing with natural-language requirements that are ambiguous or require human judgment by nature.

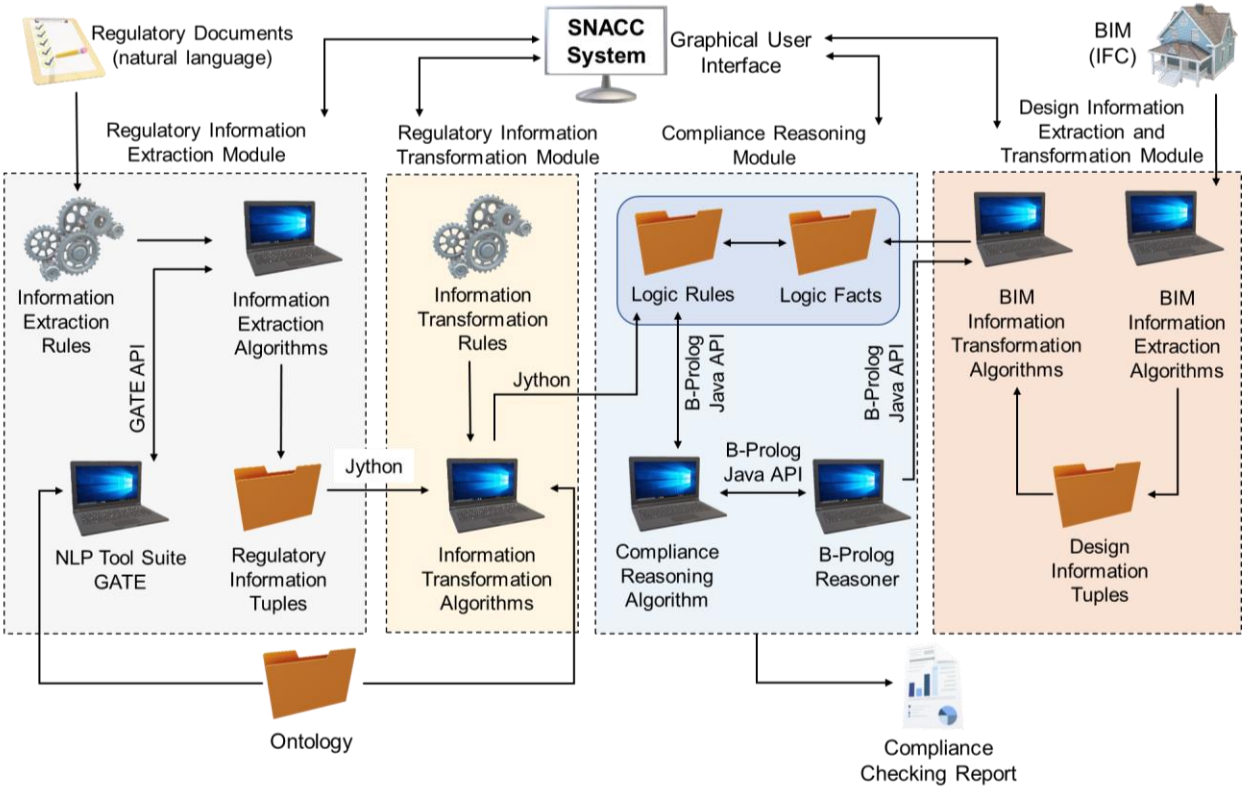

Regulatory Documents (natural language)
SNACC System
Graphical User Interface
BIM (IFC)
Regulatory Information Extraction Module
Regulatory Information Transformation Module
Compliance Reasoning Module
Design Information Extraction and Transformation Module
Information Extraction Rules
Information Extraction Algorithms
GATE API
NLP Tool Suite GATE
Regulatory Information Tuples
Jython
Information Transformation Rules
Jython
Information Transformation Algorithms
Logic Rules
Logic Facts
B-Prolog Java API
B-Prolog Java API
B-Prolog Java API
Compliance Reasoning Algorithm
B-Prolog Reasoner
BIM Information Transformation Algorithms
BIM Information Extraction Algorithms
Design Information Tuples
Ontology
Compliance Checking Report

Fig. 4 Example of a rule-based ACC system (Zhang and El-Gohary 2017b).

ML-based ACC systems. ML-based ACC systems use ML models to automatically learn the syntactic and semantic patterns in natural-language requirements and BIM models to guide the text and BIM analytics tasks such as information extraction. Such efforts have focused on applying ML and ML-based NLP techniques in different ACC processes and adapting these techniques, which originate from other domains such as computer science, to AEC domain-specific data and applications. For example, Zhang and El-Gohary (2019a) developed ML models to automatically extract and transform regulatory information from natural-language requirements into computer-processable forms. Wu and Zhang (2019) developed data-driven methods to automatically classify IFC objects for supporting the alignment of design information and regulatory information. Recent advances in deep learning methods and their application in AEC-domain tasks such as indoor localization using BIM image data (Ha et al. 2018) and design command prediction using BIM log data (Pan and Zhang 2020) have provided insights for leveraging deep learning in ACC systems. Deep learning methods use computational models such as deep neural networks to learn multi-layer abstract representations from raw data and thus can achieve great power and flexibility/adaptability (Goodfellow et al. 2016). Examples of the first research efforts to develop deep learning-based ACC systems include the recurrent neural network-based methods for regulatory information extraction and transformation developed by Zhang and El-Gohary (2019b, 2020). The application of ML and/or NLP techniques in AI-driven systems, particularly ML-based ACC systems, further pushes the levels of flexibility/adaptability and automation of ACC systems into a new optimal boundary (see Fig. 5). However, ML-based ACC systems are far less mature than AI-assisted ACC systems (see Fig. 5) – currently, there are no ACC systems that solely rely on ML techniques. In addition, none of the existing ML-based ACC approaches have been evaluated on a large number of testing cases while achieving performance competitive to the state-of-the-art AI-assisted or rule-based ACC systems.

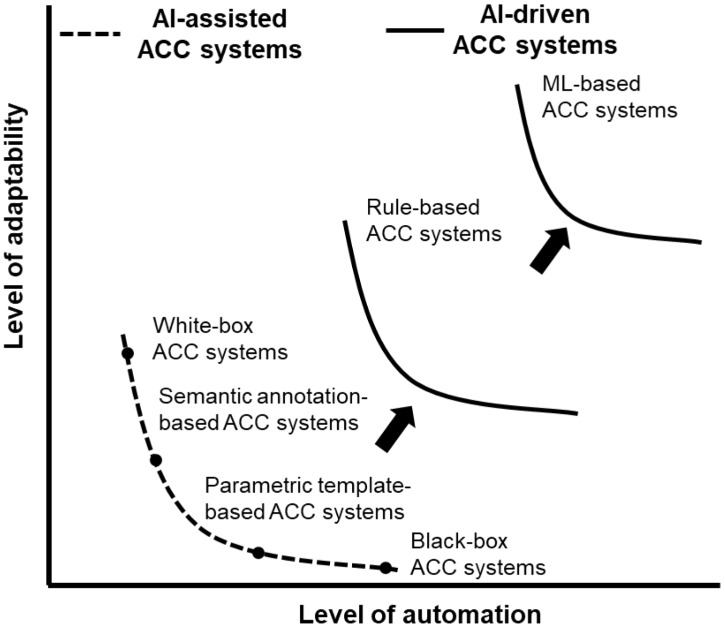


Fig. 5 Comparison of AI-assisted and AI-driven ACC systems.

Section 3: Regulatory Text Analytics in BIM-based ACC Systems

3.1 Natural-language Requirement Classification

Classifying natural-language requirements into predefined categories (e.g., binary classification of building code as checkable or not) prior to other ACC processes such as rule extraction and compliance reasoning, helps improve the performance of these processes by filtering out requirements that are either irrelevant or not checkable. Text classification is a foundational task in many NLP-based applications (Lai et al. 2015) and numerous research efforts have been undertaken in the computer science and computational linguistic domains. These efforts range from handcrafted syntactic and semantic rule-based methods [e.g., CONSTRUE/TIS by Hayes and

Weinstein (1990)] to methods using ML techniques, including supervised (Joachims 1999) and unsupervised learning algorithms (Turney 2002)]. Most recently, text classification efforts are predominantly using deep learning [e.g., character-level convolutional neural networks (Zhang et al. 2015), recurrent neural networks (Lai et al. 2015), and transfer learning with deep neural networks (Howard et al. 2018)]. Compared to domain-general text classification, classification of requirements for ACC is more challenging, mainly because there is a lack of predefined requirement types to guide the classification. Also, the regulatory text in the AEC domain has syntactic and semantic structures that are different from, and often more complex, than text in other domains such as social media (Zhou and El-Gohary 2016a; 2016b).

Manual and empirical classification. A few research efforts have manually classified natural-language requirements into types for supporting compliance checking purposes. For example, Malsane et al. (2015) classified building-code requirements based on whether they are checkable and interpretable by computers or not, and accordingly defined three requirement types: (1) declarative requirements: requirements having checkable information and thus are computer interpretable (e.g., simple geometrical rules); (2) informative requirements: requirements having information that needs human interpretation and thus are not directly interpretable by computers; and (3) remaining requirements: requirements that are not suitable for compliance checking. Solihin and Eastman (2015) classified building-code requirements based on whether they can be checked in some of the existing ACC systems (e.g., Solibri Model Checker), and if so, how they can be checked. Four requirement types were defined: (1) requirements that need explicit BIM data for compliance checking; (2) requirements that need attribute values derived from explicit BIM data for compliance checking; (3) requirements that need extended BIM data structures for compliance checking; and (4) requirements that need a “proof of solution” (e.g., an illustrative case and/or a manual compliance reasoning process). In many commercialized ACC applications (e.g., SMARTreview), developers manually classify requirements into two groups – those that need direct verification and those that do not.

NLP and ML-based classification. Recent research efforts used AI-driven methods that leverage NLP and ML techniques to automatically classify natural-language requirements and/or identify requirement types. Examples of supervised learning-based methods include Salama and El-Gohary (2016), who proposed a supervised learning-based method for classifying regulatory documents and contract clauses into predefined categories (e.g., environmental, safety, health) using a mixed set of text features (e.g., document frequency) and feature reweighting techniques. Zhou and El-Gohary (2016a) proposed a supervised learning-based method with word-level and document-level features (e.g., term frequency and inverse document frequency) to classify requirements (e.g., in the International Energy Conservation Code) according to environmental compliance checking topics. Le et al. (2019) developed a Naïve Bayes-based method to differentiate requirements from non-requirement text in contractual documents. Examples of unsupervised learning-based methods include Zhou and El-Gohary (2016b), who proposed an ontology- and clustering-based method to classify regulatory text according to an environmental compliance checking topic hierarchy. Zhang and El-Gohary (2018) proposed a hierarchical clustering-based method and syntactic and semantic features to classify requirements based on their syntactic and semantic features and their level of computability, which is defined as the ability of natural-language requirements to be automatically understood and processed by computers. NLP and ML-based requirement classification methods can greatly reduce manual effort and scale up to various types of regulatory documents.

3.2 Regulatory Information Extraction

In ACC systems, regulatory information (e.g., subject of compliance, quantity value, and quantity units) is extracted from natural-language requirements and transformed into computer-processable forms to support compliance reasoning. Numerous information extraction efforts have been undertaken in the computer science and computational linguistics domains for supporting different data analytics tasks, such as entity recognition (e.g., Bommarito II et al. 2018), event extraction (Chambers and Jurafsky 2011), and commonsense question answering (Fader et al. 2011). Regulatory information extraction in ACC systems is more challenging compared to information extraction for many other data analytics tasks because ACC requires deep or full information extraction – the entire meaning of the text must be captured for complete and correct extraction of the elements of the requirements, while in tasks such as entity recognition only partial information (e.g., companies and geopolitical entities) are extracted from the text.

Rule-based regulatory information extraction. Existing state-of-the-art information extraction methods rely on rules that are developed using NLP techniques and semantic analysis. For example, Zhang and El-Gohary (2013; 2015) and Zhou and El-Gohary (2017) developed semantic NLP-based methods, which use semantic and syntactic features and information extraction rules (as shown in Fig. 6) to extract semantic information elements from regulatory documents such as building codes, energy conservation codes, and specifications for supporting ACC. Li et al. (2016) used NLP techniques to translate the textual descriptions of spatial configurations into computer-processable spatial rules. Park and Lee (2016) developed NLP and logic-based rules to automatically translate natural-language requirements into queries. Despite the state-of-the-art performance levels many of them have achieved [e.g., nearly 100% recall reported by Zhang and El-Gohary (2013) and Zhou and El-Gohary (2017), with above 95 % precision], rule-based approaches are difficult to scale to a variety of documents due to the limited patterns that are used to develop the rules. In general, when the type of regulatory document or the characteristics of the text change, although some of the IE rules could be reused, most of these rules will require retesting and possibly modification or addition. Depending on the amount of retesting and adaptation involved, this could require significant effort.

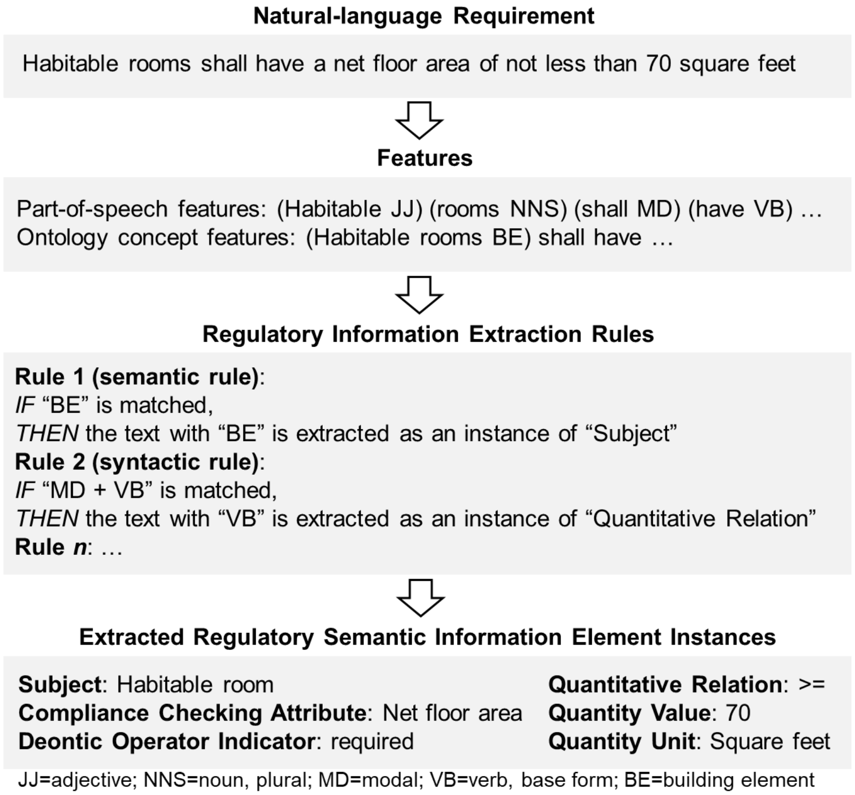
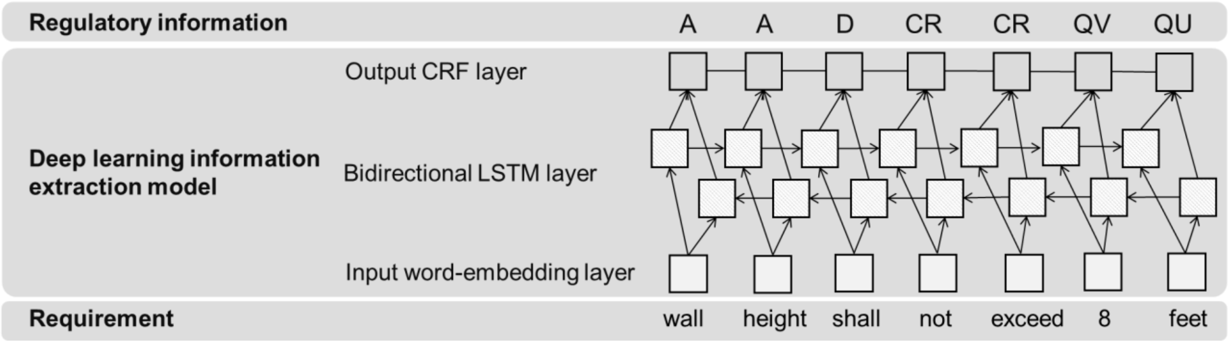

Natural-language Requirement
Habitable rooms shall have a net floor area of not less than 70 square feet
Features
Part-of-speech features: (Habitable JJ) (rooms NNS) (shall MD) (have VB) …
Ontology concept features: (Habitable rooms BE) shall have …
Regulatory Information Extraction Rules
Rule 1 (semantic rule):
IF "BE" is matched,
THEN the text with "BE" is extracted as an instance of "Subject"
Rule 2 (syntactic rule):
IF "MD + VB" is matched,
THEN the text with "VB" is extracted as an instance of "Quantitative Relation"
Rule n: …
Extracted Regulatory Semantic Information Element Instances
Subject: Habitable room
Compliance Checking Attribute: Net floor area
Deontic Operator Indicator: required
Quantitative Relation: >=
Quantity Value: 70
Quantity Unit: Square feet
JJ=adjective; NNS=noun, plural; MD=modal; VB=verb, base form; BE=building element

Fig. 6. Example of rule-based regulatory information extraction in ACC systems (Zhang and El-Gohary 2015).

ML-based regulatory information extraction. The most recent ACC research efforts have focused on leveraging ML-based NLP techniques in IE methods to automatically capture the syntactic and semantic patterns – which need to be explicitly identified by experts in rule-based IE methods. For example, Zhang and El-Gohary (2019a) proposed a conditional random field (CRF)-based method to extract semantic roles in the form of predicate-argument-modifier structures. Zhang and El-Gohary (2019b) proposed a recurrent neural network-based method to extract requirement hierarchies from building-code sentences, where each hierarchy consists of requirement units and dependencies between the units. Xu and Cai (2019) used a semantic frame-based information extraction method to support utility compliance checking. Xue and Zhang (2020) used ML-based NLP to improve part-of-speech tagging for supporting rule-based IE. Zhong et al. (2020) used bidirectional long short-term memory (LSTM) and CRF models to extract procedural constraints from construction regulations. Zhang and El-Gohary (2020) used deep learning models, consisting of LSTM and CRF, together with transfer learning strategies (as shown in Fig. 7) to extract information from the 2009 International Building Code.

Several of the aforementioned research efforts have succeeded to use ML models to reduce the amount of manual effort needed in the regulatory information extraction process, while achieving high performance. For example, Zhang and El-Gohary (2020) have achieved nearly 90% precision and recall for most information classes. However, the state-of-the-art rule-based methods, expectedly, still achieved higher levels of performance. For example, Zhang and El-Gohary (2013) and Zhou and El-Gohary (2017) have reached a recall of nearly 100%, with above 95% precision.

In terms of implementation, the aforementioned ML-based information extraction methods are also still in progress, requiring further integration in existing ACC systems and/or further testing on a wide range of regulatory documents.



A=compliance checking attribute; D=deontic operator indicator; CR=comparative relationship; QV=quantity value; QU=quantity unit; LSTM=long short term memory; CRF=conditional random field

Fig. 7. Example of machine learning-based regulatory information extraction in ACC systems (Zhang and El-Gohary 2020).

Section 4: BIM Information Analytics in BIM-based ACC Systems

4.1 BIM Semantic Enrichment

Current BIMs provide limited support for ACC because the information in the BIMs are typically incomplete and/or unnormalized, making the BIM representation unable to fully meet the needs of compliance checking (Sacks et al. 2020). Semantic enrichment of BIM models for supporting ACC aims to infer new meaningful information, which is required for or will facilitate ACC processes, and add the inferred information to the models (Belsky et al. 2016).

Rule-based semantic enrichment. Rule-based semantic enrichment methods apply expert-defined rules to add information to existing BIM models. The most recent, state-of-the-art semantic enrichment method in the context of ACC is the SeeBIM (Belsky et al. 2016, Sacks et al. 2017). SeeBIM, originally developed by Belsky et al. (2016), first parses the IFC file of the BIM model to extract the attributive information (e.g., geometry and location) of the objects, and then applies semantic enrichment rules to infer additional information about the objects, which is stored in an enriched IFC file that can be used in ACC systems. Sacks et al. (2017) further enhanced the SeeBIM by enabling the classification of BIM objects, and extending the semantic enrichment rules to facilitate the computing of complex geometry and processing of precise topological requirements.

ML-based semantic enrichment. ML-based semantic enrichment methods automatically supplement the BIM models with semantic information generated by ML models (e.g., classification and clustering models). For example, Zhang and El-Gohary (2016) developed a relation classification method that uses ML algorithms such as support vector machine (SVM) and k-nearest neighbors, and syntactic and semantic features, to classify the relationships between the regulatory and IFC concepts for semantically enriching existing BIMs with regulatory concepts. Koo et al. (2019) proposed a supervised learning-based method that uses SVM to supplement BIM models with mappings between BIM objects and IFC classes. Ma et al. (2018) proposed a similarity-based method to classify BIM objects for adding object classification information to the BIM models. Wu and Zhang (2019) proposed a pattern matching-based method to classify BIM objects based on geometric features into predefined categories and integrate these categories into the BIM models.

Compared to rule-based methods, ML-based methods eliminate the cost of developing semantic-enrichment rules. However, similar to other ML-based approaches, their performance could be limited. For example, the object misclassification detection method by Koo et al. (2019) achieved a range of 80.95% to 97.14% of accuracy for different classes (Koo et al. 2019). The object detection recall and precision ranged from 84.45% and 85.20% for common building element categories to 100% for detailed beam categories (Wu and Zhang 2019). On the other hand, rule-based methods could achieve perfect or nearly perfect performance. For example, 100% accuracy was achieved for 390 objects in the testing case of a bridge model (Ma et al. 2018).

4.2 BIM-Regulatory Information Alignment

BIMs and regulatory documents speak different languages – the information representations and terminology used in the BIMs are different from those used in the natural-language requirements. BIM-regulatory information alignment aims to align a concept or relationship in natural language to the corresponding BIM concept (e.g., an IFC entity, an enumeration type, etc.) or relationship by mapping or transforming one or both types of concepts/relationships.

Hardcoding or rule-based information alignment. Existing research efforts for BIM-regulatory information alignment are predominately based on hardcoding or predefined rules. They can be classified into three main groups based on how the two types of information are changed during the alignment: requirement translation, BIM-requirement mapping, and BIM adaptation.

In requirement translation, the concepts in the requirements are manually mapped to those in the BIM, and then translators are developed to automate the mapping of instances. These translators typically use modeling languages such as SPARQL protocol and Resource Description Framework (RDF) query language (SPARQL) (Yurchyshyna and Zarli 2009), visual code checking language (Preidel and Borrmann 2016), and building environment rule and analysis language (Lee 2011). In BIM-requirement mapping, concepts and relationships in the requirements are mapped to those in the BIM models using dictionaries (e.g., buildingSMART Data Dictionary), rules (e.g., Tan et al. 2010, Pauwels et al. 2011, Zhou and El-Gohary 2018), or ontologies (e.g., Yurchyshyna et al. 2009, Zhong et al. 2015, Beach et al. 2015). In BIM adaptation, the BIM models are modified to enable direct alignment between the representations of the requirements and the BIM models by adding concepts and relationships from the requirements (e.g., requirements in International Building Code) to the BIM schema (Zhang and El-Gohary 2016) or by modifying existing properties in the BIM model itself (Choi et al. 2014).

Despite the state-of-the-art performance achieved by the hardcoding or rule-based BIM-regulatory information alignment methods, they still require significant manual effort, making these methods time-consuming and costly. And, many of these methods lack flexibility/adaptability (e.g., due to the use of pre-defined mappings or mapping rules) and might not allow successful implementation across different BIM models (e.g., BIM models in different design stages), different types of regulations (e.g., building code versus energy code), and changes or updates to the BIM or the regulations (Garrett et al. 2014, Dimyadi et al. 2016).

ML-based information alignment. A few research efforts have explored the use of supervised learning models such as classification models for BIM-regulatory information alignment for supporting ACC. Zhang and El-Gohary (2016) developed a hybrid rule and ML-based method to extend the IFC schema. The hybrid method consists of three parts: (1) a regulatory concept extraction method, which consists of pattern-matching rules for extracting regulatory concepts from regulatory documents, (2) a similarity-based matching method, which assesses the similarity between concepts and selects the most related IFC concepts to the extracted regulatory concepts, and (3) a relation classification method, which uses an ML model to classify the relationship between the extracted regulatory concepts and their most related IFC concepts. In their later study, Zhang and El-Gohary (2017b) successfully integrated the hybrid rule and ML-based information alignment method with methods for regulatory information extraction and transformation and compliance reasoning, in their SNACC system, showing the potential of ML in solving challenging ACC problems. Based on this hybrid method, Zhang and El-Gohary (2019c) further explored similarity based on embeddings of concepts, and different types of supervised learning algorithms and features, in classifying the relationship between regulatory and IFC concepts.

Section 5 Next Generation BIM-based ACC Systems

Although ACC is challenging because of the complex and ambiguous nature of natural-language requirements and the discrepancy between the languages spoken by BIM and the requirements, the concept of ACC has been gaining strong industry and academia-wide support. With the fast-evolving NLP and AI techniques, and increasingly integrable and interoperable BIMs, an increasing number of research and commercialization efforts are being undertaken to develop BIM-based ACC systems with higher levels of performance, automation, and flexibility/adaptability. One example is a currently ongoing NSF Partnerships for Innovation (PFI) project (by the authors and other academic and industrial collaborators), which aims to develop and accelerate the commercialization of an advanced BIM-based ACC system, one that leverages NLP, AI, and interoperable BIM techniques to achieve high levels of performance, automation, and flexibility/adaptability (NSF 2020). As we continue to experience an increased technological shift from AI-assisted to AI-driven BIM-based ACC systems – a shift that increasingly uses ML and deep learning – the authors foresee four trends that will drive the AEC domain towards the next generation BIM-based ACC systems.

Procedural to end-to-end. BIM-based ACC systems will become more integrated and could reach an end-to-end status, with the boundaries between separate processes such as rule interpretation, rule representation, and building model preparation (Eastman et al. 2009) becoming blurred or even disappearing. For example, the aforementioned three processes in existing ACC systems could become a single process with the help of more advanced NLP and ML tools, where an ML model could automatically generate the compliance reasoning result (e.g., compliant, non-compliant, or irrelevant) given the natural-language requirement and a snippet of the IFC model.

However, this is challenging to achieve because of the complexity of the underlying semantic interpretations, mappings, and processing that would take place.

Empirical to data-driven. BIM-based ACC systems will become more data-driven, enabled by the use of advanced computational tools (e.g., deep neural networks) and fast-growing computational power (e.g., GPUs). By training ML models on large-scale, pattern-rich, and annotated training data from other domains using transfer learning techniques, the robustness and scalability of the ML-based methods could be improved, and the cost of preparing AEC domain-specific training data could be also reduced. Thus, the focus of developing ML-based methods for ACC systems will be on leveraging out-of-domain large datasets, fast creation of AEC-domain large datasets, and/or development of small-scale, highly discriminative datasets for fine-tuning trained ML models.

Predominance of ML-based NLP. ML-based NLP techniques will play an essential role in the future BIM-based ACC systems. By nature, BIM-based ACC systems require the alignment of BIM information, which is represented by IFC schemas, and regulatory information, which is represented in natural language. Research in ML-based NLP tasks such as semantic parsing, question answering, and machine translation could provide important insights into solving the alignment problem. For example, recurrent neural network-based methods have been proposed to automatically convert natural-language sentences to logic languages [e.g., the lambda calculus (Berant and Liang 2014, Yih et al. 2015)], query languages [e.g., SQL (Zhong et al. 2017) and SPARQL (Dubey et al. 2016)], or programming languages [e.g., Python (Yin and Neubig 2017)].

Emergence of explainable AI. Many ML models are "black-box", especially the ones that use complex computational tools (e.g., deep neural network), which risks reducing the "one-step" AI-driven ACC systems that use such models to be uninterpretable or unexplainable. Such "black-box" AI-driven ACC systems might not be trusted by the users, because of the difficulty for both the users and developers to evaluate the systems and analyze the errors. To achieve trustworthiness, and to better evaluate the systems and analyze the errors, explanations of AI decisions in the ML-based ACC systems are necessary – these explanations are expected to "provide insight into the rationale the AI uses to draw a conclusion" (Doran et al. 2017). Thus, these "black-box" AI-driven systems are expected to upgrade into "clear box", explainable AI-driven systems. Different from AI-assisted systems that mostly use explainable, symbolic AI techniques, where most of the work is carried out by humans, the majority of the work in explainable AI-driven systems would be done by ML models. Yet, users of the explainable AI-driven systems would be able to understand how the compliance results are generated, by possibly leveraging explainable AI techniques such as visualizing features and elucidating the neurons and layers in deep neural networks (Zhu et al. 2018) in the systems.

Section 6 Conclusion

In this chapter, the authors discussed the needs and challenges of BIM-based ACC systems, and reviewed existing BIM-based ACC efforts and systems that leverage NLP and AI techniques towards increasing performance and automation. The chapter covered both AI-assisted and AI-driven systems, and highlighted the technological shift towards AI-driven systems and ML approaches. The authors focused on four important processes needed in advanced BIM-based ACC systems: natural-language requirement classification, regulatory information extraction, semantic enrichment, and BIM-regulatory information alignment. Existing approaches and solutions were reviewed, with a focus on the state-of-the-art NLP- and AI-based methods. In the end, the authors identified four trends for the next generation BIM-based ACC systems.

This chapter mainly reviewed the current research progress in regulatory text analytics and BIM information analytics, in current BIM-based ACC systems. Yet new, more powerful NLP and AI technologies, such as deep learning, are evolving rapidly and are becoming the new standard in automatically processing, analyzing, and understanding digitalized information in numerous domains including the AEC domain. It is expected that the next generation of AI-driven ACC systems will rely on advanced technologies at the forefront of the NLP and AI domains, such as end-to-end ML and explainable AI, which would bring increased integration, automation, and adaptability to compliance checking as well as increased transparency and explainability to the analytics processes. These technologies will be the key to tackling the technological challenges in BIM and BIM-based ACC systems, and thus deserve the attention of researchers and entrepreneurs, both in academia and industry.

Acknowledgements

The authors would like to thank the National Science Foundation (NSF). This material is based on work supported by the NSF under Grant No. 1827733. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

Alpaydin, E., 2020. Introduction to machine learning. MIT press.

Balaban, Ö., Kilimci, E.S.Y. and Cagdas, G., 2012. Automated code compliance checking model for fire egress codes.

Beach, T.H., Rezgui, Y., Li, H. and Kasim, T., 2015. A rule-based semantic approach for automated regulatory compliance in the construction sector. Expert Systems with Applications, 42(12), pp.5219-5231.

Belsky, M., Sacks, R., and Brilakis, I., 2016. Semantic enrichment for building information modeling. Comput.-Aided Civ. Infrastruct. Eng., 31(4), 261–274.

Berant, J. and Liang, P., 2014, June. Semantic parsing via paraphrasing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1415-1425).

Bloch, T. and Sacks, R., 2020. Clustering Information Types for Semantic Enrichment of Building Information Models to Support Automated Code Compliance Checking. Journal of Computing in Civil Engineering, 34(6), p.04020040.

Bommarito II, M.J., Katz, D.M. and Detterman, E.M., 2018. LexNLP: Natural language processing and information extraction for legal and regulatory texts. arXiv preprint arXiv:1806.03688.

buildingSMART, 2020. buildingSMART Data Dictionary. http://bsdd.buildingsmart.org/#peregrine/about. (August 15, 2020).

Chambers, N. and Jurafsky, D., 2009, August. Unsupervised learning of narrative schemas and their participants. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP (pp. 602-610).

Choi, J., Choi, J. and Kim, I., 2014. Development of BIM-based evacuation regulation checking system for high-rise and complex buildings. Automation in Construction, 46, pp.38-49.

City of Manassas, 2019. Building Plan Review. https://www.manassascity.org/729/Building-Plan-Review. (October 10, 2019)

City of Sacramento, 2019. Plan Review Timelines. https://www.cityofsacramento.org/Community-Development/Building/Plan-Review/Plan-Review-Timelines (Oct. 10, 2019)

Dimyadi, J., Pauwels, P. and Amor, R., 2016. Modelling and accessing regulatory knowledge for computer-assisted compliance audit. Journal of Information Technology in Construction, 21, pp.317-336.

Doran, D., Schulz, S. and Besold, T.R., 2017. What does explainable AI really mean? A new conceptualization of perspectives. arXiv preprint arXiv:1710.00794.

Dubey, M., Dasgupta, S., Sharma, A., Höffner, K. and Lehmann, J., 2016, May. Asknow: A framework for natural language query formalization in sparql. In European Semantic Web Conference (pp. 300-316). Springer, Cham.

Eastman, C., Lee, J.M., Jeong, Y.S. and Lee, J.K., 2009. Automatic rule-based checking of building designs. Automation in construction, 18(8), pp.1011-1033.

Fader, A., Soderland, S. and Etzioni, O., 2011, July. Identifying relations for open information extraction. In Proceedings of the 2011 conference on empirical methods in natural language processing (pp. 1535-1545).

Fiatech, 2012. AutoCodes project: phase 1, proof-of-Concept final report. http://www.fiatech.org/images/stories/techprojects/project_deliverables/Updated_project_deliverables/AutoCodesPOCFINALREPORT.pdf. (October 10, 2019).

Garnelo, M., Arulkumaran, K. and Shanahan, M., 2016. Towards deep symbolic reinforcement learning. arXiv preprint arXiv:1609.05518.

Garrett Jr, J.H., Palmer, M.E. and Demir, S., 2014. Delivering the infrastructure for digital building regulations.

Goldberg, Y., 2017. Neural network methods for natural language processing. Synthesis Lectures on Human Language Technologies, 10(1), pp.1-309.

Goodfellow, I., Bengio, Y. and Courville, A., 2016. Deep learning. MIT press.

Government of Singapore, 2016. CORENET e-Submission System. https://www.corenet.gov.sg/general/corenet-e-submission-system.aspx. (December 01, 2020).

Ha, I., Kim, H., Park, S. and Kim, H., 2018. Image retrieval using BIM and features from pretrained VGG network for indoor localization. Building and Environment, 140, pp.23-31.

Hayes, P.J. and Weinstein, S.P., 1990. CONSTRUE/TIS: A System for Content-Based Indexing of a Database of News Stories. In IAAI (Vol. 90), pp. 49-64.

Hayes-Roth, F., Waterman, D.A. and Lenat, D.B., 1983. Building expert system.

Hjelseth, E., and Nisbet, N., 2010. Exploring semantic based model checking. Proc. 27th CIB W78 Int. Conf. http://itc.scix.net/data/works/att/w78-2010-54.pdf. (May 15, 2018).

Howard, J. and Ruder, S., 2018. Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146.

Joachims, T., 1999. Transductive inference for text classification using support vector machines. In Icml (Vol. 99), pp. 200-209.

Kaplan, A. and Haenlein, M., 2019. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons, 62(1), pp.15-25.

Kincelova, K., Boton, C., Blanchet, P. and Dagenais, C., 2020. Fire safety in tall timber building: A BIM-based automated code-checking approach. Buildings, 10(7), p.121.

Koo, B., La, S., Cho, N.W. and Yu, Y., 2019. Using support vector machines to classify building elements for checking the semantic integrity of building information models. Automation in Construction, 98, pp.183-194.

Lai, S., Xu, L., Liu, K. and Zhao, J., 2015. Recurrent convolutional neural networks for text classification. In Twenty-ninth AAAI conference on artificial intelligence.

Lau, G. and Law, K., 2004. An information infrastructure for comparing accessibility regulations and related information from multiple sources (p. 11). Professur Informatik im Bauwesen.

Le, T., Le, C., Jeong, H.D., Gilbert, S.B. and Chukharev-Hudilainen, E., 2019. Requirement text detection from contract packages to support project definition determination. In Advances in informatics and computing in civil and construction engineering (pp. 569-576). Springer, Cham.

Lee, J.K., 2011. Building environment rule and analysis (BERA) language and its application for evaluating building circulation and spatial program (Doctoral dissertation, Georgia Institute of Technology).

Lee, Y.C., Eastman, C.M. and Lee, J.K., 2015. Automated rule-based checking for the validation of accessibility and visibility of a building information model. In Computing in Civil Engineering 2015, pp. 572-579.

Li, S., Cai, H. and Kamat, V.R., 2016. Integrating natural language processing and spatial reasoning for utility compliance checking. Journal of Construction Engineering and Management, 142(12), p.04016074.

Ma, L., R. Sacks, U. Kattel, and T. Bloch. 2018. 3D object classification using geometric features and pairwise relationships. Comput. Aided Civ. Infrastruct. Eng. 33 (2): 152–164. https://doi.org/10.1111/mice.12336.

Malsane, S., Matthews, J., Lockley, S., Love, P. E., and Greenwood, D., 2015. Development of an object model for automated compliance checking. Autom. Construct., 49, 51-58.

Martins, J.P. and Monteiro, A., 2013. LicA: A BIM based automated code-checking application for water distribution systems. Automation in Construction, 29, pp.12-23.

Nawari, N.O., 2012. Automating codes conformance. Journal of architectural engineering, 18(4), pp.315-323.

Nawari, N.O., 2019. A Generalized Adaptive Framework (GAF) for Automating Code Compliance Checking. Buildings, 9(4), p.86.

NSF (Natural Science Foundation), 2020. PFI-RP: Automating building code compliance checking and modular construction through interoperable building information modeling technology. https://www.nsf.gov/awardsearch/showAward?AWD_ID=1827733&HistoricalAwards=false. (August 15, 2020)

Pan, Y. and Zhang, L., 2020. BIM log mining: Learning and predicting design commands. Automat. Constr. 112, p.103107.

Park, S. and Lee, J.K., 2016. KBimCode-based applications for the representation, definition and evaluation of building permit rules. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 33, p. 1). IAARC Publications.

Pauwels, P., Van Deursen, D., Verstraeten, R., De Roo, J., De Meyer, R., Van de Walle, R. and Van Campenhout, J., 2011. A semantic rule checking environment for building performance checking. Automation in construction, 20(5), pp.506-518.

Preidel, C. and Borrmann, A., 2016. Towards code compliance checking on the basis of a visual programming language. Journal of Information Technology in Construction (ITcon), 21(25), pp.402-421.

REScheck EZ, 2020. REScheck energy compliance checking reports. http://rescheckez.com/rescheckcertificates/rescheckorder.htm. (August 15, 2020)

Sacks, R., Girolami, M. and Brilakis, I., 2020. Building Information Modelling, Artificial Intelligence and Construction Tech. Developments in the Built Environment, p.100011.

Sacks, R., Ma, L., Yosef, R., Borrmann, A., Daum, S. and Kattel, U., 2017. Semantic enrichment for building information modeling: Procedure for compiling inference rules and operators for complex geometry. Journal of Computing in Civil Engineering, 31(6), p.04017062.

Salama, D.M. and El-Gohary, N.M., 2016. Semantic text classification for supporting automated compliance checking in construction. Journal of Computing in Civil Engineering, 30(1), p.04014106.

Solibri, 2020. Solibri Model Checker. https://www.solibri.com/products/solibri-model-checker. (August 15, 2020)

Solihin, W., and Eastman, C., 2015. Classification of rules for automated BIM rule checking development. Autom. Construct., 53, 69-82.

Solihin, W. and Eastman, C.M., 2016. A knowledge representation approach in BIM rule requirement analysis using the conceptual graph. ITcon, 21, pp.370-401.

Tan, X., Hammad, A. and Fazio, P., 2010. Automated code compliance checking for building envelope design. Journal of Computing in Civil Engineering, 24(2), pp.203-211.

Turney, P.D., 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. arXiv preprint cs/0212032.

US Census Bureau, 2019. Building Permits Survey. https://www.census.gov/construction/bps/ (October 10, 2019)

US Department of Energy, 2020. Building Energy Codes Program. https://www.energycodes.gov/compliance. (August 15, 2020)

Wisconsin Department of Safety and Professional Services, 2019. Building Plan. Review & Inspection. https://dsps.wi.gov/Documents/Programs/CommercialBuildings/BuildingPlanReviewInspection.pdf (October 10, 2019).

Wu, J. and Zhang, J., 2019. New automated BIM object classification method to support BIM interoperability. Journal of Computing in Civil Engineering, 33(5), p.04019033.

Xu, X. and Cai, H., 2019. Semantic frame-based information extraction from utility regulatory documents to support compliance checking. In Advances in Informatics and Computing in Civil and Construction Engineering (pp. 223-230). Springer, Cham.

Xue, X. and Zhang, J., 2020. Building Codes Part-of-Speech Tagging Performance Improvement by Error-Driven Transformational Rules. Journal of Computing in Civil Engineering, 34(5), p.04020035.

Yih, S.W.T., Chang, M.W., He, X. and Gao, J., 2015. Semantic parsing via staged query graph generation: Question answering with knowledge base.

Yin, P. and Neubig, G., 2017. A syntactic neural model for general-purpose code generation. arXiv preprint arXiv:1704.01696.

Yurchyshyna, A. and Zarli, A., 2009. An ontology-based approach for formalisation and semantic organisation of conformance requirements in construction. Automation in Construction, 18(8), pp.1084-1098.

Zhang, J. and El-Gohary, N., 2013. Semantic NLP-based information extraction from construction regulatory documents for automated compliance checking. Journal of Computing in Civil Engineering, 30(2), p.04015014.

Zhang, J. and El-Gohary, N., 2015. Automated information transformation for automated regulatory compliance checking in construction. Journal of Computing in Civil Engineering, 29(4), p.B4015001.

Zhang, J., and El-Gohary, N., 2016. Extending building information models semiautomatically using semantic natural language processing techniques. J. Comput. Civ. Eng., 10.1061/(ASCE)CP.1943-5487.0000536, C4016004.

Zhang, J., and El-Gohary, N., 2017a. Semantic-based logic representation and reasoning for automated regulatory compliance checking. J. Comput. Civ. Eng., 31(1), 10.1061/(ASCE)CP.1943-5487.0000583.

Zhang, J., and El-Gohary, N., 2017b. Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking. Autom. Construct., 73, 45-57.

Zhang, R. and El-Gohary, N.M., 2018. A clustering approach for analyzing the computability of building code requirements. In Construction Research Congress 2018 (pp. 86-95).

Zhang, R. and El-Gohary, N., 2019a. A machine learning approach for compliance checking-specific semantic role labeling of building code sentences. In Advances in informatics and computing in civil and construction engineering (pp. 561-568). Springer, Cham.

Zhang, R. and El-Gohary, N., 2019b. A machine learning-based method for building code requirement hierarchy extraction. In 2019 Canadian Society for Civil Engineering Annual Conference, CSCE 2019.

Zhang, R. and El-Gohary, N., 2019c. A Machine-Learning Approach for Semantic Matching of Building Codes and Building Information Models (BIMs) for Supporting Automated Code Checking. In International Congress and Exhibition" Sustainable Civil Infrastructures" (pp. 64-73). Springer, Cham.

Zhang, R. and El-Gohary, N., 2020, November. A Machine-Learning Approach for Semantically-Enriched Building-Code Sentence Generation for Automatic Semantic Analysis. In Construction Research Congress 2020: Computer Applications (pp. 1261-1270). Reston, VA: American Society of Civil Engineers.

Zhang, S., Teizer, J., Lee, J.K., Eastman, C.M. and Venugopal, M., 2013. Building information modeling (BIM) and safety: Automatic safety checking of construction models and schedules. Automation in construction, 29, pp.183-195.

Zhang, X., Zhao, J. and LeCun, Y., 2015. Character-level convolutional networks for text classification. In Advances in neural information processing systems (pp. 649-657).

Zhong, B., Xing, X., Luo, H., Zhou, Q., Li, H., Rose, T. and Fang, W., 2020. Deep learning-based extraction of construction procedural constraints from construction regulations. Advanced Engineering Informatics, 43, p.101003.

Zhong, B.T., Ding, L.Y., Love, P.E. and Luo, H.B., 2015. An ontological approach for technical plan definition and verification in construction. Automation in Construction, 55, pp.47-57.

Zhong, V., Xiong, C. and Socher, R., 2017. Seq2sql: Generating structured queries from natural language using reinforcement learning. arXiv preprint arXiv:1709.00103.

Zhou, P. and El-Gohary, N., 2016a. Domain-specific hierarchical text classification for supporting automated environmental compliance checking. Journal of Computing in Civil Engineering, 30(4), p.04015057.

Zhou, P. and El-Gohary, N., 2016b. Ontology-based multilabel text classification of construction regulatory documents. Journal of Computing in Civil Engineering, 30(4), p.04015058.

Zhou, P. and El-Gohary, N., 2017. Ontology-based automated information extraction from building energy conservation codes. Automation in Construction, 74, pp.103-117.

Zhou, P. and El-Gohary, N., 2018, January. Automated matching of design information in BIM to regulatory information in energy codes. In Construction Research Congress 2018 (pp. 75-85).

Zhu, J., Liapis, A., Risi, S., Bidarra, R. and Youngblood, G.M., 2018, August. Explainable AI for designers: A human-centered perspective on mixed-initiative co-creation. In 2018 IEEE Conference on Computational Intelligence and Games (CIG) (pp. 1-8). IEEE.