frontiers | Frontiers in Built Environment

TYPE Original Research
PUBLISHED 17 June 2025
DOI 10.3389/fbuil.2025.1599787

OPEN ACCESS

EDITED BY
Umberto Berardi,
Toronto Metropolitan University, Canada

REVIEWED BY
Da Xu,
China University of Geosciences
Wuhan, China
Feifan Shen,
Hunan University, China

*CORRESPONDENCE
Botong Li,
botongli011@gmail.com

RECEIVED 25 March 2025
ACCEPTED 29 May 2025
PUBLISHED 17 June 2025

CITATION
Li B, Jia H, Yu H and Fu C (2025) An integrated approach of knowledge extraction and ontology-based reasoning for green building evaluation and electricity efficiency.
Front. Built Environ. 11:1599787.
doi: 10.3389/fbuil.2025.1599787

An integrated approach of knowledge extraction and ontology-based reasoning for green building evaluation and electricity efficiency

Botong Li[1,2]*, Hongjie Jia[1], Hongwei Yu[3] and Cong Fu[3]

[1]School of Electrical and Information Engineering, Tianjin University, Tianjin, China, [2]State Grid Tianjin Electric Power Company, Tianjin, China, [3]Beijing Guodian Futong Science and Technology Development Co., Ltd., Beijing, China

Introduction: Promoting green building practices is essential in addressing climate change and achieving sustainable development goals. Green building evaluation plays a critical role in assessing building performance across multiple criteria, including electricity efficiency, environmental protection, and occupant wellbeing. However, existing evaluation methods are often manual, subjective, and heavily reliant on expert judgment.

Methods: This study proposes an intelligent and automated approach to green building evaluation by integrating knowledge extraction and ontology development. Using advanced natural language processing (NLP) and machine learning techniques, relevant knowledge is extracted from diverse sources, including regulatory documents, building standards, and academic literature. The structured knowledge is then formalized into an ontology using Protégé, enabling the application of Semantic Web Rule Language (SWRL) rules for comprehensive evaluation.

Results: The proposed method enables the systematic and automated assessment of green building performance with a focus on electricity efficiency. It significantly improves the objectivity, accuracy, and scalability of the evaluation process compared to traditional expert-driven methods.

Discussion: This research demonstrates the potential of combining semantic technologies and machine learning for sustainable building assessment. The framework supports more consistent and efficient evaluations, providing a scalable tool for policymakers, developers, and sustainability assessors. Future work may extend the ontology to include dynamic sensor data and real-time monitoring.

KEYWORDS

green building evaluation, knowledge extraction, ontology, natural language processing, electricity efficiency

1 Introduction

Due to the pressing threats posed by climate change, global warming, and resource depletion, as well as the increasing demand for electricity, the concept of sustainable

development has gained widespread attention. Buildings, encompassing both residential and commercial sectors, are significant contributors to global energy demand and, consequently, to environmental impacts (Zhang et al., 2019). Statistics show that energy consumption from buildings can account for up to 40% of total energy use, primarily for heating, cooling, and lighting, contributing significantly to carbon emissions (Pérez-Lombard et al., 2008). In this context, green buildings have emerged as a vital solution, offering the potential to drastically reduce energy consumption and emissions, while also providing healthier and more comfortable environments for occupants (Darko and Chan, 2016). Promoting green building practices is, therefore, essential in the broader effort to combat climate change and achieve sustainable development.

According to the core principles of sustainable development in China, a green building is one that conserves resources such as energy, land, water, and materials to the greatest extent possible (Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2006). It aims to minimize environmental impact and pollution throughout its entire lifecycle, providing occupants with healthy, comfortable, and efficient spaces that coexist harmoniously with nature. Given this definition of green building, it becomes crucial to have reliable methods for evaluating whether buildings truly meet these sustainability standards. Hence, green building evaluation plays a vital role in assessing a structure's performance across various criteria, ensuring that it aligns with the principles of resource conservation, environmental protection, and occupant wellbeing (Tang et al., 2019; Ding et al., 2018). To this end, many countries have implemented policies and developed green building evaluation or rating systems to evaluate the performance of buildings. The Building Research Establishment's Environmental Assessment Method, commonly known as BREEAM, was first introduced in 1990 by the Building Research Establishment (BRE) in the United Kingdom, making it one of the earliest evaluation methods for green buildings (Mitchell, 2010; Lee, 2013; Alyami and Rezgui, 2012). Other notable examples include the Leadership in Energy and Environmental Design (LEED) created by the United States (The U.S. Green Building Council, 2008) and Evaluation Standard for Green Building established by China (Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2006). It should be noted that the evaluation process for green buildings varies depending on the specific rating systems used. For instance, in BREEAM, there is an initial pre-assessment stage where a pre-assessment estimator is utilized to predict a potential score. Although this score is neither certified nor final, it serves as a helpful guide to understanding the achievable outcomes for the building's sustainability performance (Ding et al., 2018).

Based on the development of national green building evaluation systems, an increasing number of scholars have begun to focus on green building evaluation in their research (Doan et al., 2017; Hong et al., 2019; Shen et al., 2017; He et al., 2024). This growing body of work aims to refine and expand the methods used to assess the performance of green buildings, providing valuable insights and advancements in the field. For instance, Wu and Chang (Wu and Chang, 2013) developed a visual Req calculation tool based on the 3D environment to evaluate whether a building is energy-efficient, which can reduce computation time and errors compared to manual methods. The Building Information Model (BIM), which integrates multidisciplinary data across the entire building lifecycle, offers significant convenience by providing easy access to the information needed for green building evaluation, so many research efforts have also been allocated to BIM-enabled green building evaluation. Azhar et al. (2011) integrated BIM-based sustainability analyses with the LEED certification process, which have been proved to streamline the certification process. Motawa and Carter, (2013) integrated BIM-based models with sensor technologies to enhance the assessment of buildings' energy performance and carbon emissions. Though BIM can provide multidisciplinary information of the whole life cycle, lack of analytical solutions for green building evaluation further limits the use of BIM in evaluating green building (Lu et al., 2017).

Actually, the evaluation of green buildings is a typical knowledge-intensive task. This has been verified by the study conducted by Ali and Al Nsairat, (2009), which used the questionnaire with domain experts of sustainable development to gain perspectives for green building evaluation. Green building evaluation requires integrating information from various sources, including design specifications, operational data, and regulatory standards. An ontology provides a structured framework for representing and organizing complex domains of knowledge (Pauwels et al., 2017). Hence, increasingly, research in green building evaluation is adopting knowledge-based and ontology-enabled approaches to enhance the evaluation process. In this regard, Zhang et al. (2019) developed an ontology manually and integrated ontology with BIM to conduct green building evaluation from a semantic and social approach. Similarly, Baumgärtel and Scherer, (2016) used ontology-based workflow to achieve automatic check of green building design, where the knowledge in ontology is transferred from BIM model and several rules are established in ontology for automatic check. Jiang et al. (2018) also integrated BIM with ontology to facilitate the process of green building evaluation, where the ontology was established manually and knowledge in the ontology was mainly from the standards. It can be seen that in the context of green building evaluation, ontology helps to define relationships between different concepts, enabling a more coherent and systematic analysis and leading to more accurate and comprehensive evaluations. However, current approaches to building ontologies for green building evaluation are largely manual, relying heavily on expert input. This process introduces a significant degree of subjectivity, which can affect the consistency and reliability of the evaluations. There is an urgent need for automated knowledge extraction methods to construct ontologies more objectively. To the best of our knowledge, limited research has addressed this gap by integrating deep learning-based information extraction with ontology-driven reasoning, which forms the key novelty of our proposed approach.

Hence, motivated by such need, this paper proposes the use of automated knowledge extraction techniques to construct ontologies for green building evaluation. Specifically, we leverage a BiLSTM-CRF model for entity recognition and relation extraction from green building texts, which are then structured into a domain ontology for semantic reasoning. This automation significantly differentiates our approach from previous manually constructed ontology-based systems. By automating the ontology-building process, this paper aims to reduce subjectivity and enhance the

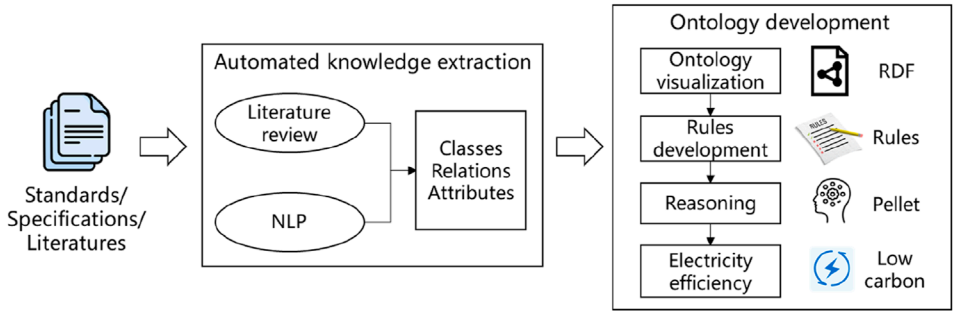

FIGURE 1
The overall methodology of integrated approach of knowledge extraction and ontology-based reasoning for green building evaluation.

consistency of the evaluation, enabling a more comprehensive and data-driven approach to sustainability evaluation. The following of this paper is structured as follows. Section 2 elucidates the overall methodology of this study. Section 3 provides the implementation of the methodology. Section 4 demonstrates the automated knowledge extraction results and the development of ontology for green building evaluation, as well as the evaluation results. Section 5 discusses the limitations of this study. Section 6 provides findings and concluding remarks.

2 Methodology

The overall methodology of this study, which uses integrated approach of knowledge extraction and ontology-based reasoning for green building evaluation, is shown in Figure 1. Firstly, machine learning and NLP techniques are employed to automatically extract relevant knowledge from textual sources such as standards and regulations. This automated knowledge extraction process enables the identification and organization of key concepts, relationships, and criteria that are essential for green building evaluation. In the second part, the extracted knowledge is utilized to construct a detailed ontology, which serves as a structured representation of the domain. Building on this ontology, specific evaluation rules are then developed which can be applied to evaluate green buildings more comprehensively and accurately. This integrated approach not only streamlines the evaluation process but also ensures that it is grounded in a robust, objective knowledge base, ultimately leading to more reliable and effective green building evaluation. Finally, the green building evaluation is expected to improve the electricity efficiency, thus contributing to a more low-carbon building.

2.1 Automated knowledge extraction

In this study, the primary data sources for knowledge extraction are textual documents, including standards, regulations, and relevant academic literature. Given the nature of these sources, NLP is employed as the main technique for extracting knowledge. NLP is particularly well-suited for this task due to its ability to analyze and interpret human language, allowing for the automated identification of key concepts, relationships, and patterns within large volumes of text (Saka et al., 2023). One of the key advantages of NLP is its efficiency in processing and extracting meaningful information from unstructured data, which significantly reduces the time and effort required for manual analysis. This capability is crucial for ensuring that the extracted knowledge is both comprehensive and relevant for building an ontology that supports green building evaluation. Figure 2 presents the workflow of the automated knowledge extraction, including data collection, data pre-processing, NLP model setup and finally, automated knowledge extraction.

2.1.1 Data collection

In the data collection phase, the primary focus is on gathering relevant textual materials, which are predominantly in Chinese. These texts are sourced from various platforms to ensure a comprehensive coverage of the domain. Key sources include official government websites that publish standards and regulations related to green building (such as Evaluation Standard for Green Building (Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2006)), authoritative academic databases such as China National Knowledge Infrastructure (CNKI), and publicly available reports from reputable sources on the internet. These diverse sources provide the foundational data required for the knowledge extraction process, ensuring that the extracted information is both accurate and aligned with the latest standards and academic research in the field.

2.1.2 Data pre-processing

The second step involves data preprocessing to prepare the collected documents for knowledge extraction. Since the data comes in various formats (such as.pdf and.caj), the first task is to convert all files into a uniform.txt format. After conversion, natural language processing techniques are applied, starting with tokenization, which breaks the text into meaningful units or words (Shen et al., 2017).

Following this, stop words (commonly used words that do not contribute significant meaning in Chinese) are removed to focus on more relevant terms. Finally, the texts are annotated using the BIO tagging method, where "B" represents the beginning of an entity, "I" indicates that the entity is continuing, and "O" signifies tokens outside of any entity. BIO tagging method is a widely used

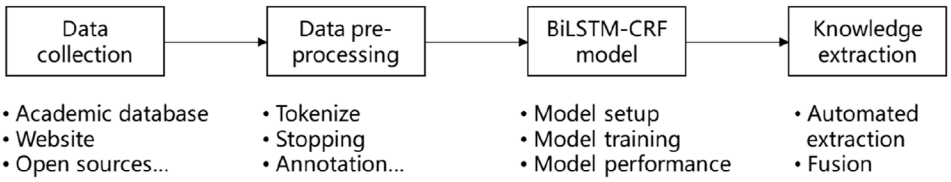

FIGURE 2
The workflow of the automated knowledge extraction.

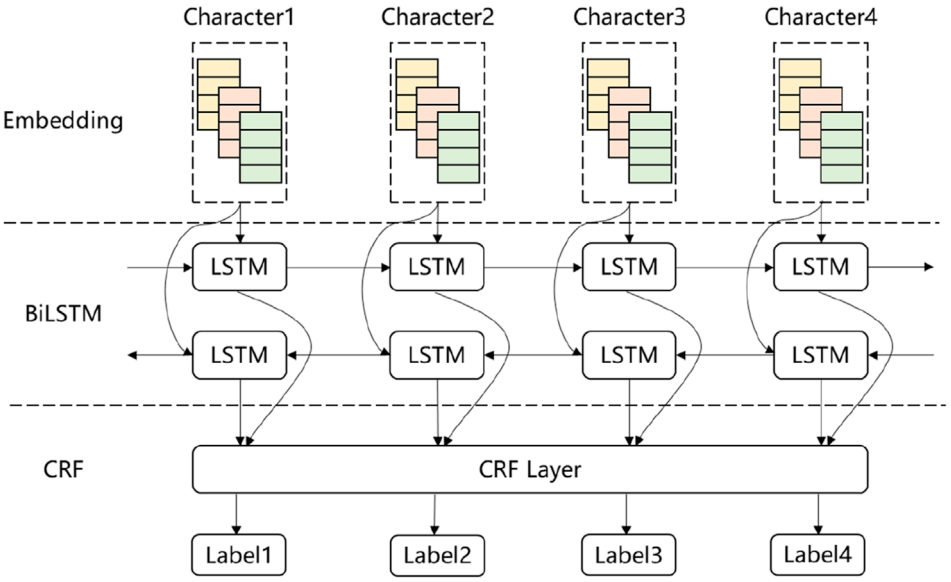

FIGURE 3
The structure of BiLSTM-CRF model.

annotation scheme in NLP for sequence labeling tasks, particularly in knowledge extraction (Ling et al., 2024). The use of the BIO tagging method is essential for the subsequent automated knowledge extraction process. By clearly marking the boundaries and structure of entities within the text, it enables machine learning models to more accurately identify and extract relevant knowledge.

2.1.3 BiLSTM-CRF model

The BiLSTM (bidirectional Long Short-Term Memory) – CRF (Conditional Random Field) model is a combination of two powerful techniques commonly used in knowledge extraction (Shao et al., 2024; Meng et al., 2022; Feng and Chen, 2021; Li et al., 2018; Wu et al., 2023). This model integrates a (BiLSTM) network with a (CRF) layer, combining the strengths of both approaches to achieve more accurate and robust predictions. BiLSTM captures contextual dependencies from both directions, while the CRF layer ensures optimal label sequence prediction by considering tag dependencies. Compared to more complex models like BERT, BiLSTM-CRF offers a good balance between performance and computational efficiency, making it suitable for domain-specific applications where training data is relatively constrained. The structure of the BiLSTM-CRF model is given in Figure 3.

As shown in Figure 3, the input to the model consists of individual characters, which are first transformed into vectors through an embedding process. In this study, one-hot encoding is used as the embedding method. One-hot encoding is a simple yet effective technique where each character is represented as a binary vector with a dimension equal to the size of the character set. In this vector, only one element is set to one (indicating the presence of a specific character), while all other elements are set to 0. This approach allows the model to process and distinguish between different characters effectively.

Then, the BiLSTM improves on traditional LSTMs by processing the input character in both forward and backward directions, enabling the model to capture context from both the past and the future. This is particularly useful in the task of knowledge extraction, where the meaning of a word depends on the surrounding context. At each time step t, an LSTM cell has three gates: the forget gate, input gate, and output gate (Ma et al., 2022). Forget gate f_t determines how much of the previous cell state to retain, input gate i_t decides how much of the new input should be added to the cell state, and output gate o_t determines the output of the cell. In BiLSTM, there are two LSTMs (one forward and one backward), and their

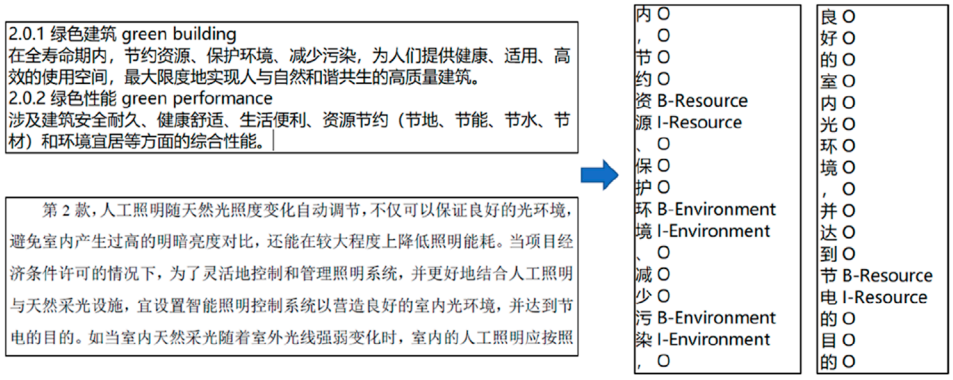

FIGURE 4
BIO annotation method for texts.

TABLE 1 Model parameters.

No.	Hyper parameter	Value
1	Hidden layer size	2*128
2	Embedding size	100
3	Epochs	50
4	Batch size	32
5	Learning rate	0.01
6	Dropout	0.5

outputs are concatenated. The calculation of BiLSTM is given from Equations 1–7.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{2}$$

$$C'_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_c\right) \tag{3}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C'_t \tag{4}$$

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{5}$$

$$h_t = o_t \cdot \tanh\left(C_t\right) \tag{6}$$

$$h_t^{BiLSTM} = \left[h_t^{forward}, h_t^{backward}\right] \tag{7}$$

Where x_t is the input vector at time step t, h_{t-1} is the hidden state vector from the previous time step t-1, σ is the sigmoid activation function, h_t is the hidden state vector at time step t, and W and b refer to the weight matrices and bias vectors, respectively, allowing the model to learn how much information to keep, forget, or output at each step.

The CRF layer is added on top of the BiLSTM to improve the model's ability to make coherent predictions at the sequence level. While BiLSTM predicts each token independently, CRF ensures that the predicted labels are globally optimized, taking into account the dependencies between labels (Chen et al., 2017). Let $x = (x_1, x_2, \ldots, x_n)$ be the input sequence, and $y = (y_1, y_2, \ldots, y_n)$ be the predicted label sequence. The score of a label sequence y for an input x in defined in Equation 8:

$$score(x, y) = \sum_{t=1}^{n} W_{yt} \cdot x_t + \sum_{t=1}^{n-1} T_{y_t, y_{t+1}} \tag{8}$$

Where W_{yt} is the score of assigning label y_t to token x_t, and $T_{y_t, y_{t+1}}$ is the transition score from label y_t to y_{t+1}.

To obtain the best label sequence, the model selects the sequence y that maximized the score, which in calculated in Equation 9:

$$y^* = \arg\max_{y} score(x, y) \tag{9}$$

2.1.4 Knowledge extraction

The final step in the process is knowledge extraction. For this, the trained BiLSTM-CRF model is used to directly input the text and predict the label for each character within the text. These labels, which have been learned during training, indicate the roles of characters in terms of the knowledge they represent—whether they are the beginning, inside, or outside of an entity, and what kind of knowledge they belong to in the context of green building context.

By applying the model to the input text, relevant knowledge embedded within the document can be systematically identified and extracted. The model's ability to accurately label each character allows for precise extraction of key concepts, relationships, and entities from the text. This automated extraction process not only streamlines the identification of critical information but also ensures a higher level of accuracy and consistency compared to manual extraction methods, making it an essential component of the overall green building evaluation framework.

2.2 Ontology development

After extracting the knowledge from texts, this study adopts Stanford's well-known Seven-Step Methodology for ontology development (Noy and McGuinness, 2001). This method provides a structured approach to building comprehensive and flexible

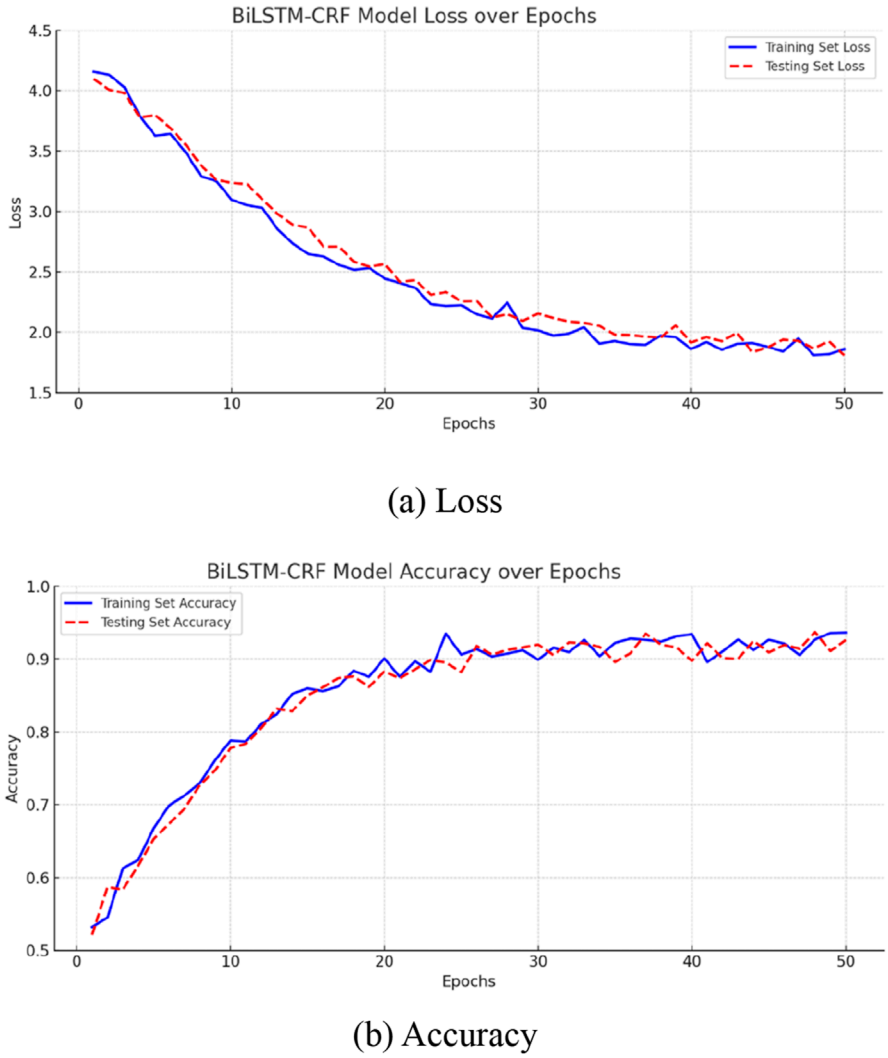

FIGURE 5
The variation of loss and accuracy of the model during the training process. **(a)** Loss. **(b)** Accuracy.

TABLE 2 Evaluation metrics for the model performance.

Class	Accuracy	Precision	Recall
Safety and durability	0.92	0.85	0.82
Health and comfort	0.88	0.87	0.81
Resource savings	0.84	0.80	0.74
Environmental livability	0.75	0.65	0.67
Living convenience	0.67	0.66	0.60

ontologies that can capture domain knowledge efficiently. In this study, the knowledge used in the ontology are derived from the automatic knowledge extraction process based on NLP techniques. The specific steps include.

2.2.1 Determine the domain and scope of the ontology

The first step involves identifying the domain of interest and the scope of the ontology. For this study, the ontology focuses on green building evaluation, including relevant standards, guidelines, building components, sustainability metrics, and environmental factors. The scope is centered around automating the evaluation process by linking these concepts and formalizing relationships based on the extracted knowledge.

2.2.2 Consider reuse of existing ontologies

In the domain of green building evaluation, available ontologies are relatively scarce, with many focusing on adjacent areas that are not closely related to the evaluation process itself (e.g., green building material ontology (Hong et al., 2019)). Some existing ontologies, while relevant, are proprietary or not publicly accessible, limiting their reuse. As a result, the potential for directly reusing existing ontological frameworks is limited. Therefore, in turn, it

TABLE 3 The extracted entities based on BiLSTM-CRF model.

Class	Count	Examples
Safety and durability	120	Fire resistance, earthquake protection, load-bearing capacity
Health and comfort	93	Indoor air quality, noise reduction, lighting comfort
Resource savings	265	Energy efficiency, water conservation, sustainable materials, renewable energy sources
Environmental livability	187	Green spaces, urban heat island mitigation, ecological footprint
Living convenience	68	Smart home technologies, public transportation access

proves the necessity of establishing a custom ontology tailored specifically to the requirements of green building evaluation, leveraging knowledge extracted from domain-specific standards, guidelines, and literature.

2.2.3 Enumerate important terms in the ontology

Key terms for the ontology, such as "energy efficiency", "thermal insulation", and "building lifecycle", are primarily derived from the knowledge extraction process. The automatic extraction identifies terms that are critical for evaluating green buildings. These terms are compiled into a comprehensive list, ensuring that the ontology covers all relevant concepts required for accurate and detailed assessments.

2.2.4 Define the classes and class hierarchy

Based on the extracted knowledge, the main classes are defined in the ontology with their hierarchical relationships. The top-level classes are further subdivided into subclasses such as materials (e.g., "concrete", "wood") and Emission Metrics (e.g., "CO2 emissions"). The hierarchy is developed to organize these concepts in a logical manner, supporting the efficient retrieval of information during the evaluation process.

2.2.5 Define the properties of classes

For each class, properties are defined to capture their attributes and relationships. These properties are also derived from the knowledge extraction phase, where textual descriptions are analyzed to identify key features and characteristics associated with different green building concepts.

2.2.6 Define the facets of the properties

Once the properties are defined, facets of these properties—such as data types, value constraints, and cardinality—are specified. For instance, the property "CO2 emissions" might be defined to accept numerical values within a specific range, reflecting realistic emission data. These constraints help ensure that the ontology can effectively model the domain and support accurate green building evaluations.

2.2.7 Create instances of classes

Finally, instances of the defined classes are created. These instances represent specific entities such as particular building projects, materials, or energy performance data.

3 Implementation of the methodology

3.1 Data collection and pre-processing

Through manual screening, a total of ten relevant local and national standards were collected, along with 145 related academic papers. These documents serve as the core data for the knowledge extraction process, providing a comprehensive set of sources that reflect both regulatory frameworks and scholarly perspectives on green building evaluation.

For the collected documents, after converting them into.txt format, the next step involves annotating the text using the BIO tagging method, as described in Section 2.1.2. This method is effective for labeling sequences of words in the text based on predefined categories. Drawing on relevant standards (Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2006; L. Shanghai Research Institute of Building Sciences Co, 2019), the annotation focuses on five main themes: safety and durability, health and comfort, resource savings, environmental livability, and living convenience. These categories are commonly referenced in national standards and prior research, forming the foundation for green building assessment. Each category was assigned two tags—B- (beginning of the entity) and I- (inside the entity)—while tokens not belonging to any category were labeled as O. This resulted in a total of 11 tags.

After this preprocessing step, the result is a fully annotated file where each line contains a character from the text along with its corresponding label, as shown in Figure 4. Each character is tagged according to the BIO annotation scheme, linking it to one of the 11 predefined tags. This format allows for a clear, line-by-line mapping of each character to its respective annotation, making it easy to process the data for further BiLSTM-CRF model training. In total, the annotated file contains 80,473 lines. The data were divided into training set and testing set with a ratio of 7:3.

3.2 Model parameters setup

For the BiLSTM-CRF model used in this project, several parameter settings were applied, as presented in Table.1. The parameter settings for the BiLSTM model are consistent with those used in similar research studies (Miwa and Bansal, 2016). Specifically, the hidden layer size of 128 units, the embedding dimension of 100, and the use of 50 epochs align with

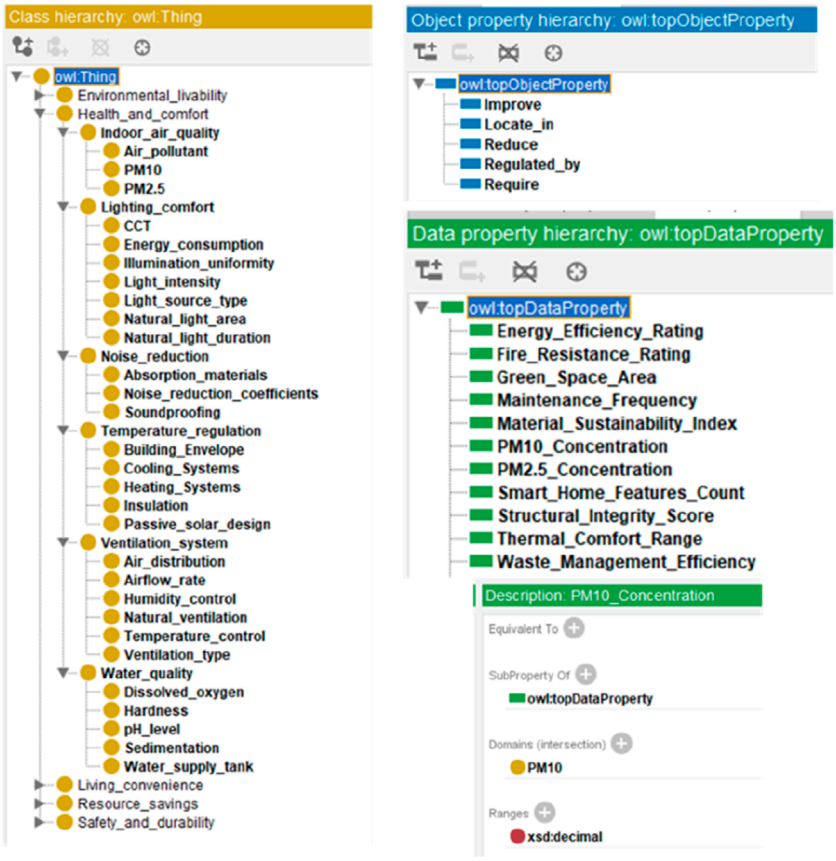

FIGURE 6
The established ontology in Protégé.

commonly adopted practices in text processing tasks involving sequence labeling.

3.3 Knowledge extraction results

The variation of the loss and accuracy during the training process is shown in Figures 5a,b, respectively. The BiLSTM-CRF model shows promising performance based on the training and testing results. The loss steadily decreases from 4.3 to around 1.8 over the course of 50 epochs, demonstrating effective model learning without overfitting. Similarly, the accuracy for both the training and testing sets improves consistently and stabilizes after about 30 epochs, reaching approximately 0.92 by the end of the training period.

To further evaluate the performance of the BiLSTM-CRF model, three key metrics will be used, namely, Accuracy, Precision, and Recall. Accuracy reflects the overall correctness of the model by measuring the proportion of correctly predicted labels out of all predictions, Precision assesses the quality of the positive predictions by calculating the proportion of true positives out of all positive predictions made by the model, and Recall evaluates the model's ability to identify all relevant instances by measuring the proportion of true positives out of the total actual positives. The overall performance of the model is shown in Table 2, which indicate the model can be used to extract knowledge from the large volume of unannotated texts.

It can be seen from Table 2 that for certain categories, such as living convenience, exhibited relatively lower recall compared to others. This can be attributed to the fact that expressions related to living convenience are less frequently and less explicitly mentioned in standard documents and technical texts, leading to fewer annotated examples in the training dataset. Moreover, the linguistic patterns associated with this category are often more diverse and context-dependent, making them harder for the model to learn effectively. While this limitation exists, it does not affect the core objective of this study, which is to demonstrate the feasibility of automated entity extraction for green building evaluation. In future work, this issue can be addressed by expanding the annotated corpus for underrepresented categories and exploring more advanced models to improve recognition performance where necessary.

The trained BiLSTM-CRF model can be utilized for knowledge extraction, with the extracted knowledge categorized under five main themes: safety and durability, health and comfort, resource savings, environmental livability, and living convenience. After manual check, a total of 733 entities were extracted. The extraction results of entities are shown in Table 3. In addition, the relationships linking the entities were also extracted and summarized. "Improve", "reduce", "locate-in", "require", and "regulated-by" were extracted as the main relationships between entities.

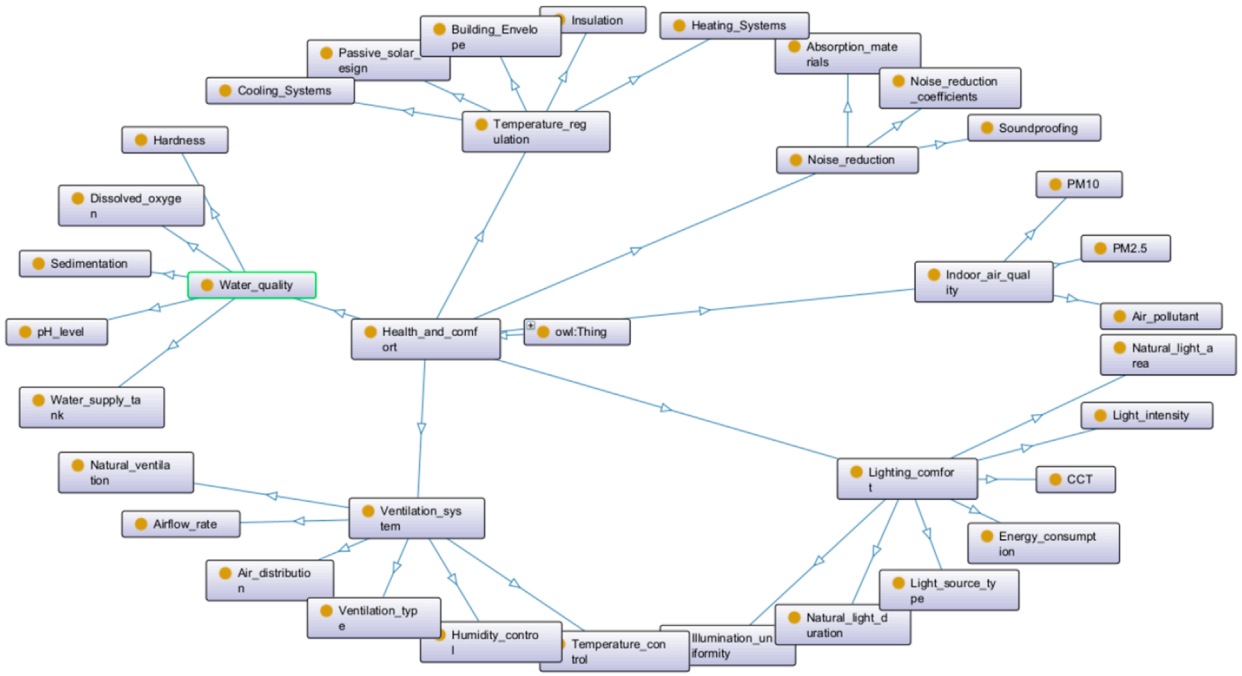

FIGURE 7
Overall structure of the ontology (example of entity "Health and comfort").

TABLE 4 Examples of SWRL rules for green building evaluation.

Rule	Description
1	Building (?b) ^hasSystem (?b, VentilationSystem) ^VentilationSystem (?vs) ^improve (?vs, IndoorAirQuality) ^IndoorAirQuality (?iaq, ?level) ^lessThan (?level, 50) - > CompliantAirQuality (?b)
2	Building (?b) ^hasSystem (?b, WaterSavingSystem) ^WaterSavingSystem (?ws) ^reduce (?ws, WaterConsumption) ^WaterConsumption (?wc, ?amount) ^lessThan (?amount, 200) - > CompliantWaterSavings (?b)
3	Building (?b) ^hasSystem (?b, WasteManagementSystem) ^WasteManagementSystem (?wms) ^reduce (?wms, WasteProduction) ^WasteProduction (?wp, ?volume) ^lessThan (?volume, 100) - > CompliantWasteManagement (?b)
4	Building (?b) ^locate-in (?b, ProximityToPublicTransport) ^ProximityToPublicTransport (?pt, ?distance) ^lessThan (?distance, 500) - > CompliantConvenience (?b)
5	Building (?b) ^hasScore (?b, SafetyDurability, ?sdScore) ^hasScore (?b, HealthComfort, ?hcScore) ^hasScore (?b, ResourceSavings, ?rsScore) ^hasScore (?b, EnvironmentalLivability, ?elScore) ^hasScore (?b, LivingConvenience, ?lcScore) ^swrlb:add (?totalScore, ?sdScore, ?hcScore, ?rsScore, ?elScore, ?lcScore) ^swrlb:divide (?avgScore, ?totalScore, 5) - > hasAverageScore (?b, ?avgScore)
6	Building (?b) ^hasAverageScore (?b, ?avgScore) ^greaterThanOrEqual (?avgScore, 60) ^lessThan (?avgScore, 70) - > OneStarRating (?b)
7	Building (?b) ^hasAverageScore (?b, ?avgScore) ^greaterThanOrEqual (?avgScore, 70) ^lessThan (?avgScore, 85) - > TwoStarRating (?b)
8	Building (?b) ^hasAverageScore (?b, ?avgScore) ^greaterThanOrEqual (?avgScore, 85) - > ThreeStarRating (?b)

3.4 Ontology development

Based on the extracted entities and relationships, an ontology was constructed to organize and represent the knowledge systematically. The ontology was developed using the Protégé software, which provides a flexible and powerful environment for ontology modeling (Benomrane et al., 2016). Figure 6 illustrates the ontology created in Protégé, where previously extracted entities are represented as classes, and attributes are recorded as data properties and object properties. Figure 7 displays the ontology interface created with OntoGraf, using "Health and Comfort" as an example to illustrate the relationships between entities.

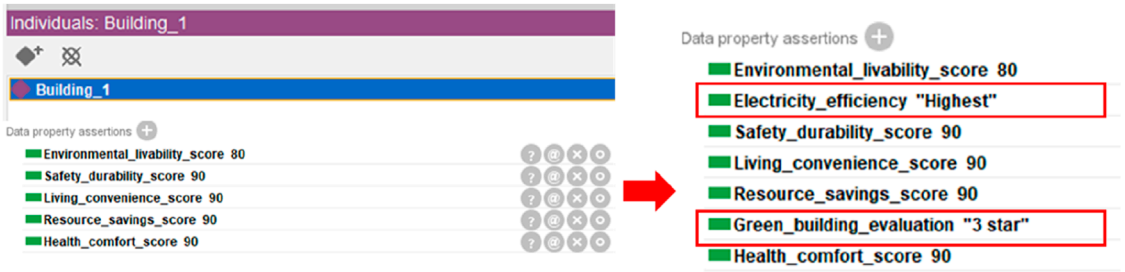

FIGURE 8
Example of green building evaluation based on SWRL.

4 Knowledge-driven green building evaluation

Based on the knowledge extracted through the automated knowledge extraction process and the ontology developed from this knowledge, a rule-based approach will be applied to evaluate green buildings. The ontology serves as a structured representation of the critical concepts, entities, and relationships in the green building domain, providing a foundation for defining evaluation criteria. By leveraging Semantic Web Rule Language (SWRL) reasoning mechanisms, it becomes possible to apply logical rules that capture the relationships between different elements in the ontology and assess compliance with green building standards.

In this study, to implement the rule-based evaluation for green buildings, the JESS (Java Expert System Shell) inference engine will be used in conjunction with the developed ontology within Protégé. JESS is a rule-based inference engine that allows for the execution of logical rules defined in SWRL, and by using JESS, the system can infer new knowledge based on the predefined rules and the relationships captured in the ontology. For instance, it can automatically deduce whether a building meets certain sustainability criteria by checking its compliance with energy efficiency, resource usage, or environmental impact regulations.

Table 4 presents some examples of SWRL rules for green building evaluation. Rules 1-4 are used to evaluate whether the design of an existing building meets the standards for Health and Comfort, Resource Savings, Environmental Livability, and Living Convenience, respectively. These rules help in determining the compliance of building features with each of these key green building criteria. Rule five is designed to aggregate the scores from the five major categories, including Safety and Durability, and calculate an overall score for the building's performance in terms of green building standards (L. Shanghai Research Institute of Building Sciences Co, 2019). This process ensures a balanced assessment by considering all critical aspects of sustainability. Rules 6-8 are used to evaluate the overall score, assigning a green building rating based on the established thresholds. Buildings that meet the required score receive ratings that range from one star (basic compliance) to three stars (high performance), providing a clear and standardized measure of their environmental and sustainability performance.

Figure 8 illustrates how green building evaluation is conducted in Protégé using SWRL rules. After creating a new individual representing the building, relevant attributes are input, such as scores for health, comfort, and resource savings. Once these attributes are defined, the SWRL rules are executed using Protégé reasoning engine. The rules automatically process the input data and generate the final evaluation results, determining the building's green building rating based on the predefined criteria. Upon completion of the green building evaluation, the system also assesses electricity efficiency. As depicted in Figure 8, a higher green building rating is associated with improved electricity efficiency, reflecting the building's overall sustainability and its effectiveness in reducing energy consumption.

5 Discussions

Compared with existing intelligent evaluation systems for green building certification, our proposed approach offers distinct advantages in knowledge acquisition and scalability. Traditional systems often rely on expert-defined rules or BIM-based workflows, which require manual encoding of sustainability criteria and are difficult to update when new standards emerge. In contrast, our method employs a BiLSTM-CRF model to automate the extraction of evaluation-relevant knowledge from domain texts, reducing subjectivity and enabling the construction of a dynamic, ontology-based reasoning framework. This automation enhances adaptability and provides a more robust foundation for continuous improvement and extension of evaluation systems.

Similarly, while electricity efficiency assessment has been extensively studied using data-driven methods such as simulation models, real-time sensor data analysis, and statistical forecasting, these approaches often focus solely on energy metrics and are limited to the operational phase of a building. Our framework complements these efforts by introducing a knowledge-driven evaluation perspective, where electricity efficiency is assessed alongside other criteria—such as health, comfort, and resource conservation—within a unified ontology. This holistic integration ensures that the evaluation process aligns more closely with the multi-dimensional nature of sustainability in green building practices.

The proposed framework is designed with flexibility to support different green building rating systems such as LEED, BREEAM, and China's GB/T 50,378. To accommodate variations in criteria hierarchies across standards, the ontology structure can be modularly extended or restructured, with each evaluation category and sub-criterion represented as separate classes or properties. This

allows for the integration of standard-specific requirements without changing the core model. The SWRL rules are also designed to be easily adjustable—regional thresholds can be modified directly within the rules, requiring only minimal expert input. However, we acknowledge that some fundamental differences in evaluation logic, such as the weighting of certain criteria or context-specific priorities, may not be fully harmonized within a single unified framework. Addressing these deeper incompatibilities remains a direction for future work, where we aim to explore systematic strategies for multi-standard alignment and cross-regional comparability.

The proposed framework demonstrates strong potential for generalization beyond electricity efficiency evaluation. Since the ontology structure and knowledge extraction pipeline are not limited to a single sustainability criterion, the method can be adapted to assess other key aspects of green building performance, such as water efficiency, indoor environmental quality, and material sustainability. By updating the training corpus and refining entity-relation definitions, the BiLSTM-CRF model can extract relevant domain knowledge from new textual sources, which can then be structured into an extended ontology. This flexibility suggests that the framework is not only scalable but also transferable to broader sustainability assessment scenarios, supporting a more comprehensive and modular green building evaluation system.

In addition, this study is primarily based on Chinese green building standards and regulations, which may limit the global applicability of the proposed framework. In future work, we plan to extend the system to support multilingual knowledge extraction and incorporate international standards such as LEED or BREEAM. Additionally, the current ontology is constructed using static textual data. To enable real-time and adaptive evaluation, future developments will explore the integration of dynamic IoT data, such as real-time energy consumption and indoor environmental metrics.

6 Conclusions and limitations

In this study, a comprehensive approach to green building evaluation has been proposed, utilizing automated knowledge extraction and ontology development to streamline the assessment process. Through the application of machine learning techniques—specifically a BiLSTM-CRF model—for extracting domain knowledge from standards and literature, and the use of Protégé with SWRL rules for structured evaluation, the study provides a more objective, consistent, and scalable framework for assessing building sustainability. The proposed method enables systematic evaluation of key green building criteria, including health, comfort, resource efficiency, environmental livability, and user convenience, ultimately producing a standardized and explainable rating. By reducing reliance on manual ontology construction, this approach not only improves the accuracy and transparency of evaluations but also addresses a critical gap in current research—namely, the lack of automated, data-driven tools for ontology development in the context of green building assessment.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

BL: Writing – original draft, Investigation, Data curation, Methodology, Conceptualization, Formal Analysis. HJ: Supervision, Writing – review and editing, Project administration. HY: Resources, Writing – review and editing. CF: Writing – review and editing, Validation.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

Author BL was employed by State Grid Tianjin Electric Power Company. Authors HY and CF were employed by Beijing Guodian Futong Science and Technology Development Co., Ltd.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Ali, H. H., and Al Nsairat, S. F. (2009). Developing a green building assessment tool for developing countries - case of Jordan. *Build. Environ.* 44, 1053–1064. doi:10.1016/j.buildenv.2008.07.015

Alyami, S. H., and Rezgui, Y. (2012). Sustainable building assessment tool development approach. *Sustain. Cities Soc.* 5, 52–62. doi:10.1016/j.scs.2012.05.004

Azhar, S., Carlton, W. A., Olsen, D., and Ahmad, I. (2011). Building information modeling for sustainable design and LEED® rating analysis. *Automation Constr.* 20, 217–224. doi:10.1016/j.autcon.2010.09.019

Baumgärtel, K., and Scherer, R. J. (2016). "Automatic ontology-based green building design parameter variation and evaluation in thermal energy building performance analyses," in *eWork and eBusiness in Architecture, Engineering and Construction: ECPPM*. CRC Press, 667–672.

Benomrane, S., Sellami, Z., and Ben Ayed, M. (2016). An ontologist feedback driven ontology evolution with an adaptive multi-agent system. *Adv. Eng. Inf.* 30, 337–353. doi:10.1016/j.aei.2016.05.002

Chen, T., Xu, R. F., He, Y. L., and Wang, X. (2017). Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Syst. Appl.* 72, 221–230. doi:10.1016/j.eswa.2016.10.065

Darko, A., and Chan, A. P. C. (2016). Critical analysis of green building research trend in construction journals. *Habitat Int.* 57, 53–63. doi:10.1016/j.habitatint.2016.07.001

Ding, Z. K., Fan, Z., Tam, V. W. Y., Bian, Y., Li, S. H., Illankoon, I., et al. (2018). Green building evaluation system implementation. *Build. Environ.* 133, 32–40. doi:10.1016/j.buildenv.2018.02.012

Doan, D. T., Ghaffarianhoseini, A., Naismith, N., Zhang, T. R., Ghaffarianhoseini, A., and Tookey, J. (2017). A critical comparison of green building rating systems. *Build. Environ.* 123, 243–260. doi:10.1016/j.buildenv.2017.07.007

Feng, D., and Chen, H. N. (2021). A small samples training framework for deep Learning-based automatic information extraction: case study of construction accident news reports analysis. *Adv. Eng. Inf.* 47, 101256. doi:10.1016/j.aei.2021.101256

He, Q. F., Wu, Z. Z., and Chen, X. S. (2024). An integrated framework for automatic green building evaluation: a case study of China. *Front. Eng. Manag.* 11, 269–287. doi:10.1007/s42524-023-0274-0

Hong, S. H., Lee, S. K., and Yu, J. H. (2019). Automated management of green building material information using web crawling and ontology. *Automation Constr.* 102, 230–244. doi:10.1016/j.autcon.2019.01.015

Jiang, S. H., Wang, N., and Wu, J. (2018). Combining BIM and ontology to facilitate intelligent green building evaluation. *J. Comput. Civ. Eng.* 32, 32. doi:10.1061/(asce)cp.1943-5487.0000786

L. Shanghai Research Institute of Building Sciences Co (2019). *Green building evaluation standard of Shanghai*. Shanghai, China: DG/TJ08-2090, Shanghai Housing and Urban and Rural Construction Management Commission.

Lee, W. L. (2013). A comprehensive review of metrics of building environmental assessment schemes. *Energy Build.* 62, 403–413. doi:10.1016/j.enbuild.2013.03.014

Li, X. Q., Shi, T. Y., Li, P., Yang, L. B., and Ma, X. N. (2018). "Acm, BiLSTM-CRF model for named entity recognition in railway accident and fault analysis report," in *Asia-pacific conference on intelligent medical (APCIM)/7th international conference on transportation and traffic engineering (ICTTE) beijing union univ*. Beijing: PEOPLES R CHINA, 1–5.

Ling, J. X., Li, X. J., Li, H. J., An, Y., Rui, Y., Shen, Y., et al. (2024). Hybrid NLP-based extraction method to develop a knowledge graph for rock tunnel support design. *Adv. Eng. Inf.* 62, 102725. doi:10.1016/j.aei.2024.102725

Lu, Y. J., Wu, Z. L., Chang, R. D., and Li, Y. K. (2017). Building Information Modeling (BIM) for green buildings: a critical review and future directions. *Automation Constr.* 83, 134–148. doi:10.1016/j.autcon.2017.08.024

Ma, C. X., Dai, G. W., and Zhou, J. B. A. (2022). Short-term traffic flow prediction for urban road sections based on time series analysis and LSTM_BILSTM method. *IEEE Trans. Intelligent Transp. Syst.* 23, 5615–5624. doi:10.1109/tits.2021.3055258

Meng, F. Q., Yang, S. S., Wang, J. D., Xia, L., and Liu, H. (2022). Creating knowledge graph of electric power equipment faults based on BERT-BiLSTM-CRF model. *J. Electr. Eng. and Technol.* 17, 2507–2516. doi:10.1007/s42835-022-01032-3

Ministry of Housing and Urban-Rural Development of the People's Republic of China (2006). *Evaluation standard for green building, GB/T 50378-2019, China architecture publishing and media Co*. Beijing, China: Ltd.

Mitchell, L. M. (2010). "Green Star and NABERS: learning from the Australian experience with green building rating tools," *Energy efficient cities: assessment tools and benchmarking practices*, 93–124.

Miwa, M., and Bansal, M. (2016). "End-to-End relation extraction using LSTMs on sequences and tree structures," in *54th annual meeting of the association-for-computational-linguistics*. GERMANY: ACL Berlin, 1105–1116.

Motawa, I., and Carter, K. (2013). Sustainable BIM-based evaluation of buildings. *26th World Congr. International-Project-Management-Association (IPMA) Crete, GREECE* 74, 419–428. doi:10.1016/j.sbspro.2013.03.015

Noy, N. F., and McGuinness, D. L. (2001). Ontology development 101: a guide to creating your first ontology, Stanford knowledge systems laboratory technical report KSL-01-05.

Pauwels, P., Zhang, S. J., and Lee, Y. C. (2017). Semantic web technologies in AEC industry: a literature overview. *Automation Constr.* 73, 145–165. doi:10.1016/j.autcon.2016.10.003

Pérez-Lombard, L., Ortiz, J., and Pout, C. (2008). A review on buildings energy consumption information. *Energy Build.* 40, 394–398. doi:10.1016/j.enbuild.2007.03.007

Saka, A. B., Oyedele, L. O., Akanbi, L. A., Ganiyu, S. A., Chan, D. W. M., and Bello, S. A. (2023). Conversational artificial intelligence in the AEC industry: a review of present status, challenges and opportunities. *Adv. Eng. Inf.* 55, 101869. doi:10.1016/j.aei.2022.101869

Shao, R. Q., Lin, P., and Xu, Z. H. (2024). Integrated natural language processing method for text mining and visualization of underground engineering text reports. *Automation Constr.* 166, 105636. doi:10.1016/j.autcon.2024.105636

Shen, L. Y., Yan, H., Fan, H. Q., Wu, Y., and Zhang, Y. (2017). An integrated system of text mining technique and case-based reasoning (TM-CBR) for supporting green building design. *Build. Environ.* 124, 388–401. doi:10.1016/j.buildenv.2017.08.026

Tang, K. H. D., Foo, C. Y. H., and Tan, I. S. (2019). A review of the green building rating systems, 2nd international conference on materials technology and energy (ICMTE) curtin univ Malaysia, sarawak, Malaysia.

The U.S. Green Building Council (2008). *Leadership in energy and environmental design*. Washington DC, USA.

Wu, I. C., and Chang, S. (2013). Visual Req calculation tool for green building evaluation in Taiwan. *Automation Constr.* 35, 608–617. doi:10.1016/j.autcon.2013.01.006

Wu, W. J., Wen, C. F., Yuan, Q., Chen, Q. L., and Cao, Y. Z. (2023). Construction and application of knowledge graph for construction accidents based on deep learning. *Eng. Constr. Archit. Manag.* 32, 1097–1121. doi:10.1108/ecam-03-2023-0255

Zhang, D. X., Zhang, J. Y., Guo, J. N., and Xiong, H. M. (2019). A semantic and social approach for real-time green building rating in BIM-based design. *Sustainability* 11, 3973. doi:10.3390/su11143973