Journal of Cleaner Production

Development of An Ontology for Construction Carbon Emission Tracking and Evaluation

--Manuscript Draft--

Manuscript Number:	JCLEPRO-D-23-27170
Article Type:	Original article
Keywords:	Carbon Emission; Building Construction Sites; Ontology; Data Integration; Project Management
Abstract:	The concentration and magnitude of carbon emissions from building construction sites highlight the imperative for effective carbon emissions management, which requires the integration of diverse data from multiple sources. Ontologies are widely employed for data integration but there is an absence of domain-specific ontologies for carbon emissions management during construction. We developed a Carbon Emission Management Ontology (CEMO) utilizing a hybrid development approach. CEMO comprises eight modules, namely Project, Quantity of Works, Environment Parameters, Agents, Equipment, Observation, KPI, and Calculation. To assess its efficacy, CEMO was evaluated by experts based on criteria such as clarity, coverage, consistency, and extensibility. The feasibility and effectiveness of CEMO were demonstrated through a case study, showcasing its utility in carbon intensity calculation and evaluation of earned carbon value. CEMO facilitates seamless integration of data from disparate sources, enabling real-time and automated management of onsite carbon emissions as well as disclosure of carbon-related information.

Development of An Ontology for Construction Carbon Emission Tracking and Evaluation

Yujie Lu[a,b], Guanghan Song[a], Peixian Li[c,d,*], Na Wang[a]

[a] College of Civil Engineering, Tongji University, Shanghai, 200092, China

[b] Key Laboratory of Performance Evolution and Control for Engineering Structures of Ministry of Education, Tongji University, Shanghai, 200092, China

[c] College of Architecture and Urban Planning, Tongji University, Shanghai, 200092, China

[d] Key Laboratory of Ecology and Energy-saving Study of Dense Habitat, Ministry of Education, Shanghai, 200092, China

[*] Corresponding Author: Peixian Li, College of Architecture and Urban Planning, Tongji University, No. 1239 Si Ping Road, Shanghai, 200092, PR China. Email: lipx@tongji.edu.cn

Abstract: The concentration and magnitude of carbon emissions from building construction sites highlight the imperative for effective carbon emissions management, which requires the integration of diverse data from multiple sources. Ontologies are widely employed for data integration but there is an absence of domain-specific ontologies for carbon emissions management during construction. We developed a Carbon Emission Management Ontology (CEMO) utilizing a hybrid development approach. CEMO comprises eight modules, namely Project, Quantity of Works, Environment Parameters, Agents, Equipment, Observation, KPI, and Calculation. To assess its efficacy, CEMO was evaluated by experts based on criteria such as clarity, coverage, consistency, and extensibility. The feasibility and effectiveness of CEMO were demonstrated through a case study, showcasing its utility in carbon intensity calculation and evaluation of earned carbon value. CEMO facilitates seamless integration of data from disparate sources, enabling real-time and automated management of onsite carbon emissions as well as disclosure of carbon-related information.

Keywords: Carbon Emission; Building Construction Sites; Ontology; Data Integration; Project Management

1. Introduction

The building industry plays a big part in global carbon emissions, accounting for a third of the total emissions [1,2]. This emphasizes how important it is to reduce carbon footprints. To achieve genuine energy efficiency and carbon emission reduction goals, it is necessary to comprehensively assess carbon emissions throughout a building's entire life cycle, including design, construction, use, and demolition phases. Among these stages, construction phase is a notable carbon emitter, accounting for 12.6% of the building's life-cycle emissions [3,4].

In recent years, scholars have increasingly recognized the pivotal role of building construction in curbing carbon emissions [5]. Many studies have stressed the need to promptly assess carbon emissions' extent and intensity during construction [6-10]. Research demonstrates that construction processes demand substantial resource consumption and the deployment of various equipment, leading to significant short-term carbon dioxide emissions [11]. The rise of smart construction accelerates the adoption of advanced machinery and electrical equipment at construction sites, further amplifying these characteristics [12]. Given the substantial carbon emissions associated with construction sites, meticulous management of construction-related emissions is paramount [13]. Besides carbon emissions, construction also impacts the local environment and society through air and water pollution, noise, and traffic congestion. Thus, carbon emission management in construction is pivotal for both environmental and social sustainability, as well as residents' quality of life.

The energy consumed by construction machinery constitutes the main source of carbon emissions in construction, which is significantly impacted by various factors, including construction quantity, machinery efficiency, shift schedules, and energy types. To assess the potential for reducing carbon emissions onsite and optimize construction decisions, construction managers require diverse data, including geological, mechanical, and engineering parameters. For example, selecting the most suitable machinery and optimizing shift schedules demand careful evaluation. Nonetheless, carbon-related data captured onsite is controlled by multiple stakeholders who may use disparate data management systems. For instance, machinery renters have information about machinery performance and maintenance, while contractors possess the construction environment data. The structure, format, and representation of this data are heterogeneous and fragmented. In most projects, managers must calculate carbon emissions manually to track deviations between the preset and actual emissions and make appropriate management decisions.

Real-time and fine-grained management of carbon emissions during construction pose significant challenges due to (1) the laborious task of collecting diverse and fragmented data, (2) difficulties in directly analyzing data of various formats, accuracy, and frequency, and (3) inefficient manual calculations of carbon emission deviations. An efficient method is required to integrate heterogeneous carbon emission-related data from construction sites, to enable real-time tracking and refined management of onsite carbon emissions. Traditional data fusion methods, such as support vector machines [14] and convolutional neural networks [15], offer partial solutions for the problem of system and structural heterogeneity between data. However, they are not effective in handling the more profound issue of semantic heterogeneity.

Addressing the challenging issue of semantic heterogeneity across multiple data sources is a current area of focus [16], and ontologies provide a key technology for resolving the challenge of data fusion in the presence of semantic heterogeneity. Integrating ontologies into data fusion processes enables a better understanding of the semantic meaning underlying data knowledge and the extraction of implicit relationships within the conceptual information of the data. Due to their ability to improve data interoperability, cross-domain linkage, and logical inference and proofs, various ontologies have been developed to address the challenges faced by the Architecture, Engineering, and Construction (AEC) industry. For example, there are ontologies for bridge rehabilitation project management [17], construction workflows [18], and value-for-money assessments [19].

Several studies have used ontologies to tackle carbon emissions in the construction industry, such as Hou et al. [20], who used Semantic Web technology to represent embodied carbon emissions and energy consumption related to building materials and structural design. Abanda [21] developed a data interoperability system using ontology and building information modeling (BIM) that can automatically calculate the embodied energy consumption, carbon emissions, and material cost of buildings. Despite the existing ontology studies focusing on building materials, they have not addressed the issue of direct carbon emissions from the energy consumption of construction machinery. In summary, the existing ontologies in the AEC industry are developed for specific domains and have limited generality and migration ability. In other words, there is currently no domain ontology specifically developed to address the inherent challenges of integrating heterogeneous data sources related to carbon emissions at construction sites.

This study developed an ontology to address the aforementioned research gap by integrating the diverse data sources related to construction carbon emissions, thereby facilitating the fine-grained management of carbon emissions during construction. This paper presents an ontology CEMO that provides a standardized terminology set for integrating heterogeneous data related to construction carbon emissions, enabling rule-based implicit knowledge inference. Additionally, the ontology developed in this paper enables linking KPI-related formulas with integrated data to support swift and automated carbon emission KPI calculations. This study has the potential to enhance the efficiency and automation of onsite carbon emission management, which is valuable to industry practitioners and academic researchers involved in smart construction and building digital twins.

The paper is organized as follows. Section 2 reviews related works on construction site carbon emissions and ontology. Section 3 describes the development process of the ontology, including a detailed description of each ontology module. Section 4 provides verification and evaluation of the developed ontology. Section 5 provides examples of the ontology's applications, including the automatic computation of carbon emission indicators and semantic reasoning of earned carbon value. Section 6 discusses how this ontology promotes data fusion and onsite carbon management. Lastly, Section 7 presents the main contributions, limitations, and future research, and provides concluding remarks.

2. Literature review

2.1 Construction carbon emissions tracking and management

Carbon emissions on construction sites can be categorized as either direct or indirect emissions [4]. Direct emissions result from the activities of fuel-powered machinery, such as excavators. Indirect emissions arise from the consumption of secondary energy sources, such as electric energy. Typically, indirect emissions on construction sites exceed direct emissions[11]. According to a study, seven types of electricity-driven machinery, including tower cranes, were responsible for over 70% of the total carbon emissions onsite [22]. Consequently, effective management of electricity-driven machinery is vital for reducing carbon emissions during construction.

Construction carbon emissions can be influenced by several factors, including construction quantity which is considered the most important [12], construction climate, geological environment, machinery performance, operational practices, and machinery work patterns. The quantity is related to various factors, such as the number of above-ground floors [23], building height[24], depth below ground level [25], building volume [26], and surface area of maintenance structures [25]. The performance of machinery, energy consumption, shift management, and operational practices by human workers all have an impact on carbon emission reduction in the construction industry [27]. However, most studies have only focused on single factors at a time. One of the main challenges is the lack of integration of information from various stakeholders that is often stored in different formats.

Various methods are typically employed to quantify construction carbon emissions, including the input-output method, the actual measurement method, and the carbon emission factor method. The input-output method is commonly used for macro-scale analyses, but its implementation can be arduous and resource-intensive [28-30]. Its efficacy is limited by its reliance on mean emission data from diverse entities, rendering it unsuitable for process analysis due to a high degree of uncertainty [31]. In contrast, the actual measurement method is more accurate, yet it is also costly, requiring rigorous adherence to representative and precise test entities [32]. The carbon emission factor method is a straightforward approach that allows for the quantification of carbon emissions during the life cycle of projects [33]. This method requires easily obtainable data and uses uncomplicated calculation methods, making it an ideal candidate for measuring carbon emissions during construction projects. Therefore, this study adopts the carbon emission factor method to quantify carbon emissions in construction projects.

The selection of appropriate carbon emission factors and the accurate collection of energy and resource consumption data are critical steps in utilizing the carbon emission factor method. Carbon emission factors from internationally recognized organizations, such as the Intergovernmental Panel on Climate Change (IPCC), Carbon Emission Accounts and Datasets (CEADs), and Ecoinvent, are typically considered authoritative sources. However, using external life cycle inventory (LCI) databases improperly can result in misleading calculations [34]. To address this issue, certain principles have been established for screening carbon emission factors, and numerous methods for localizing LCI databases have been developed. For example, Lu et al. [35] have created a BIM-based approach to transform non-native LCI data into native data. Data on energy and material consumption are usually obtained through the bill of quantities [36] or the BIM model [37]. While accurate, the bill of the quantities collection process is complicated and time-consuming. Meanwhile, the BIM model method is typically more suitable for the design phase and not ideal for tracking carbon emissions during construction. Therefore, emerging technologies have been explored for tracking material consumption on site, such as Li et al.’s [38] deep learning method based on the YOLOv3 detector for automatically detecting and counting the number of steel bars consumed on construction sites. Additionally, researchers have employed Cyber-Physical Systems (CPS) to collect machinery operation data during the construction of prefabricated buildings, while integrating carbon emission factors to visualize real-time carbon emissions on construction sites [39,40]. In conclusion, developing a real-time carbon emission monitoring system based on CPS is meaningful for the optimized management of carbon emissions on construction sites.

The key to achieving optimal control over carbon emissions on construction sites is to track real-time relationships among predefined and actual carbon emissions and construction schedules to detect and address any deviations. CarbonTrust has identified that deviations between predefined and actual carbon emissions may arise from inadequate design and testing, poor construction practices, overly complex construction environments, and poor measurement data [41]. The earned value method (EVM), a powerful tool for characterizing deviations, is widely used in the management of construction schedules and costs [42, 43]. Several studies have attempted to extend EVM to control carbon emissions on construction sites. For example, Abdi et al. [44] developed a model for measuring CO_2 performance metrics in projects, which can be used at any point in the project life cycle. Kim et al. [45] developed a system to integrate CO_2, cost, and schedule management, enabling project managers to forecast and monitor carbon emissions and costs based on construction schedules. However, due to the different data formats used for construction quantity, schedules, and CPS-based real-time carbon emissions, most calculations for carbon emission deviations are still made manually.

2.2 Ontology and its application in AEC industry

The semantic web, as a classic information model, is often considered a standardized framework that is comprehensible to both humans and computers and can be used to extract, represent, and organize relevant domain knowledge [46, 47]. The core technology that supports the implementation of the Semantic Web is ontologies [48], which consist of five fundamental elements: concepts, relations, functions, axioms, and instances. These are typically defined as "an explicit specification for representing shared concepts in a particular domain." Semantic Web technology can leverage the Web Ontology Language (OWL) to model basic ontology entities such as classes, attributes, and relationships, and then construct individual instances of ontologies with a uniform format based on the Resource Description Framework (RDF) [49].

Cross-domain data linkage using Semantic Web technology has gained increasing attention in recent years. Ontology-based approaches have been explored to connect cross-domain information in the Architecture, Engineering, and Construction (AEC) industry, including BIM [50], GIS [51], sensor data [52], design [53], and construction

[54], to the web of linked building data [55]. Furthermore, ontologies often incorporate logical statements to facilitate implicit knowledge inference. For instance, to offer information support for dynamic evaluation or management decisions for bridge rehabilitation projects, Wu et al. embedded evaluation rules for Semantic Web Rule Language (SWRL) representations into ontologies.

In the field of AEC, three fundamental applications of ontologies have been recognized: computing key performance indicators (KPIs), improving building performance, and detecting and diagnosing faults [56]. Zhong et al. [57] developed an ontology framework that integrates Building Information Modeling (BIM) data, sensor-based environmental data, knowledge related to building regulations, and design requirements, enabling its application in building environmental monitoring and automated compliance inspections. Hong et al. [58,59] developed an ontology to standardize the systematic expression of energy-related occupant behavior in buildings, promoting interoperability between occupant behavior models and building energy modeling programs. Li et al. [60] have developed a semantic ontology-based model for fault detection in building energy systems. This model exhibits strong effectiveness in tailoring fault detection solutions for various operational scenarios, along with high levels of interpretability, reliability, and automation. Specifically, in the context of KPI calculation, ontologies streamline the extraction of vital information from various buildings by associating contextual building data with KPI computation formulas [56]. Nevertheless, there currently lacks an ontology model for quantifying carbon emissions in the construction phase of buildings, which constitutes one of the primary motivations behind our ongoing research.

Several domain ontologies beyond the AEC domain deserve attention, and they can be divided into two categories: (1) conceptual ontologies related to construction environment or carbon emission accounting, and (2) ontologies related to informational monitoring. The first category includes noteworthy examples such as Friend-of-a-Friend Ontology (FOAF) [61], Weather Ontology (WO) [62], Energy and Resource Ontology (ERO) [63], Mathematical Modelling Ontology (MAMO) [64], and EM-KPI ontology (EKO) [65]. FOAF offers a machine-understandable depiction of individuals and their relationships. WO defines terms and relationships related to weather phases, while ERO comprises various energy-related concepts, such as energy parameters and carbon emission factors. MAMO ontology provides data relevant to modeling. EKO, which calculates energy consumption metrics for building operation and maintenance, was created by Li et al. [65]through the reuse of MAMO and the mathematics modeling ontology OntoMODEL [66]. The second category of ontologies encompasses SOSA/SSN [67,68], Time [69], and Om/QUDT [70,71]. SOSA/SSN offers a complete collection of representations for sensor observation processes using terms such as observation value, attribute, and object. It is important to note that the SSN ontology lacks time and unit terms. As a result, detection times and units typically utilize the Time ontology and Om/QUDT vocabulary in their descriptions.

Previous studies on carbon emissions in the construction industry have primarily focused on total emissions or emissions from building materials, with inadequate attention paid to onsite energy-related emissions. The heterogeneity of data owned by different stakeholders has made it challenging to achieve automated and refined management of carbon emissions. While many tools have been developed for construction data management, they lack proper domain ontology generalization. Therefore, no ontology is currently available for the integration of relevant information to track and manage carbon emissions on construction sites.

3. CEMO development

The development of the ontology in this study was guided by the frameworks proposed by Uschold and Grunninger [72], Zhou [73], Zheng [18], and SKEM [74], as detailed in **Fig. 1.** Stages 1-5 of the research framework are comprehensively described in the following subsections, with the evaluation stage presented in Section 4.

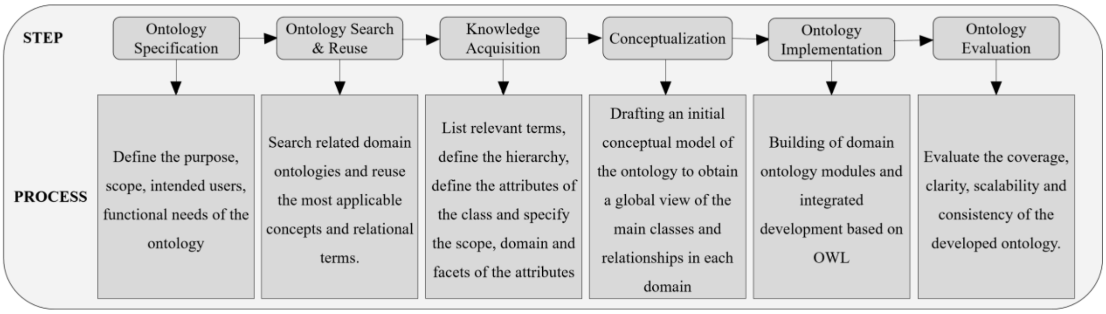

Fig. 1 Ontology development approach.

3.1 Ontology specification

The ontology specification involves defining the ontology's purpose, scope, intended users, and functional requirements. The competency question (CQs) technique is a preliminary sketch that serves as the primary tool to ensure ontology quality in early development stages. CQs can identify any knowledge relevant to the targeted ontology. Generally, an ontology's scope, purpose, and intended users are identified through the use of three standard natural language questions.

- **What is the purpose of the ontology?** The purpose of the CEMO ontology is to integrate key information that influences carbon emissions on building construction sites and provide a data source for monitoring, evaluating, and predicting carbon-related information for observation targets, such as sub-projects, machinery, individuals, teams, and others.
- **What is the scope of the ontology?** The ontology incorporates entities related to energy consumption and carbon emissions on construction sites, as well as their corresponding relations and attributes, along with the equations used to calculate carbon performance.
- **Who are the intended users of the ontology?** The ontology is intended for stakeholders with an interest in, influence over, or impact on carbon emissions at construction sites, including project managers, project engineers, and supervisory teams. These primary users can analyze and reduce carbon emissions to achieve their carbon management goals. Other stakeholders involved in machinery rental, maintenance, construction outsourcing, and energy supply should also be informed

of carbon emission status by accessing relevant data to participate in decision-making.

By combining the aforementioned purpose, scope, and intended application of the ontology, a set of core CQs has been defined to determine if the developed ontology covers the relevant domain of knowledge. These CQs have also been utilized to define ontology's concepts, properties, relations, and axioms. Several examples of the CQs are presented below:

- What background and environmental information is available for construction projects?
- What are the quantity and budget carbon emissions of each sub-project?
- What kind of machinery is used for specific engineering sub-projects?
- How can the energy consumption and carbon emissions of construction machinery be monitored?
- Who are the stakeholders involved in carbon emissions on construction sites?
- How are carbon emissions calculated for construction projects?
- How can the deviation between construction site carbon emissions and construction progress be evaluated?
- What indicators can be used to assess carbon emissions from construction sites?
- How can indicators be calculated?
- What factors contribute to construction carbon emissions?

We divided the domains into categories based on the knowledge category to which the CQs belong. For example, the CQ "What type of machinery is used for specific engineering sub-projects?" falls under the construction machinery domain, as the answer is related to the machinery involved. Each CQ is allocated to the appropriate domain. We identified ten primary domains, including observation, data property, project background, project content, engineering quantity, energy consumption, carbon emissions, machinery, personnel, and KPIs. Additionally, we included time and unit terms as two auxiliary domains to provide precise data descriptions.

3.2 Ontology reuse and knowledge acquisition

Using terms from existing ontologies can reduce the time and cost of developing an ontology from scratch, whilst also ensuring clear and unambiguous alignment with other ontologies [17]. We started by screening gold standard ontologies, such as SSN [68], which are well-established in the field of semantics for describing classes and properties associated with sensed data. Subsequently, we used an online engine to filter and identify existing ontologies that were most suitable for the relevant domain of CEMO, providing a source of crucial classes and attributes for the subsequent step. Table 1 presents examples of ontology terms that were reused.

Table 1 Modules of CEMO

Domain	Scope	Reused terms		New terms
		Existing Ontology	Examples of terms	Examples of terms
Core 1 Project (pro) & Engineering (eng)	Project background, engineering information, carbon emissions status, and management measures	N.A.	N.A.	ProjectName, ProjectType, Engineering
Core2 Agent (age)	Organizations and individuals involved in construction projects, and dynamic and static information about individuals	FOAF [61]	Gender, Member, Age, FamilyName, FirstName	Subcontractors, Occupation, OccupationalGrade
Core 3 Environment (env)	Geological, hydrological, and meteorological parameters related to construction sites	WO [62]	Weather, Humidity, WeatherCondition	EarthRock, Geology, Hydrology, MeteorologicalParameters
		EKO [65]	WindSpeed, WindDirection	
Core 4 EngineeringQuantity (qua)	Quantity for work scheduled and work performed	N.A.	N.A.	QuantityForWorkScheduled, QuantityForWorkPerformed
Core 5 Equipment (equ)	Basic information, technical parameters, and maintenance information of construction machinery and equipment	N.A.	N.A.	EquipmentName TechnicalParameter, MaintenanceRecord

Domain	Scope	Reused terms		New terms
		Existing Ontology	Examples of terms	Examples of terms
Core 6 Observation (obv)	Observation object, observation value, property, and observation time	SOSA [67]/ SSN [68]	Observation, FeatureOfInterest, Property,	RealtimeMeteorological, Weight, Distance, CarbonEmissionFactor
		ERO [65]	EnergyTypy, EnergyParameter,	
Core 7 Calculation (cal)	Evaluated objects, calculation formulas, input parameters, and output results	EKO [65]	KPIValue, KPICalculation, KPIEvaluatedObject,	KPIParameter, Operator, BinaryOperator
Core 8 Key Performance Indicator (kpi)	Indicators related to energy consumption and carbon emissions on construction sites	EKO [65]	KPI	EarnedValueLogicalIndicator, CarbonEmissionIndicator, BudgededEmissionForWorkScheduled
Supplementary 1 Time (time)	Time involved in index evaluation, observation, and information statistics	Time [69]	TemporalEntity, Instant, Interval,TemporalDuration, DurationDescription	startedAtTime, endedAtTime, generatedAtTime
Supplementary 2 Unit (unit)	The unit of each value	Om [70] /QUDT [71]	UnitOfMeasure, CompoundUnit, SingularUnit.	CarbonEmissionIntensity Unit, EnergyEfficiencyRatioUnit,

The primary objective of knowledge acquisition is to identify crucial terms associated with carbon emissions from construction sites, including project entities such as machinery, energy consumption, and carbon emissions. Critical attributes, including construction machinery and relationships between construction project entities, are also identified. Several construction manuals, construction standards, and engineering case reports were consulted as resources for identifying domain terms. We reviewed various sources such as China's Building Construction Handbook II, Smart Construction Standard, Green Construction in Building Construction, and 12 construction case reports, which cover almost all terms related to building construction. Furthermore, we examined relevant studies that could provide insights into construction carbon emission calculation and sensor data.

We sought the input of three experts from the top 100 Chinese construction companies, each possessing extensive construction management experience and specialized knowledge, and a fourth expert from academia specializing in green construction, carbon emissions, and smart construction. The expert panel meticulously classified, filtered, and supplemented knowledge based on their extensive expertise and construction experience. The inclusion of these experts was crucial in eliminating deficiencies resulting from possible developer bias or misconceptions, and in procuring missing knowledge from documents. As an illustration, the expert panel suggested placing the engineering quantity system at the same level as the indicator system rather than making it a component of it. All experts boast more than 5 years of experience in construction management.

Table 1 presents the definitions, scope, and selected terms for the identified domains of CEMO. Domain knowledge classes have a hierarchical structure based on engineering logic. Some classes, such as foundation engineering, structural engineering, and decoration engineering, can be considered subclasses, while "Engineering" can be viewed as a superclass in the engineering domain. **Fig.2** depicts a partial view of the class hierarchy in CEMO. The relationships between classes or between a class and an attribute form a Subject-Predicate-Object (SPO) triple. Ontologies have two types of properties, object properties, and data properties, as described in the reference [75]. An object property refers to the semantic relationship between two classes, such as the "hasConsumed" relationship between "Machinery" and "Energy". A data property enables quantitative or qualitative descriptions of classes. For example, "EngineeringQuantity" is quantitatively described using "hasValue". The acquired terminology, except for reused terms, is defined based on common terms or expert advice and can be modified or expanded to suit specific project conditions.

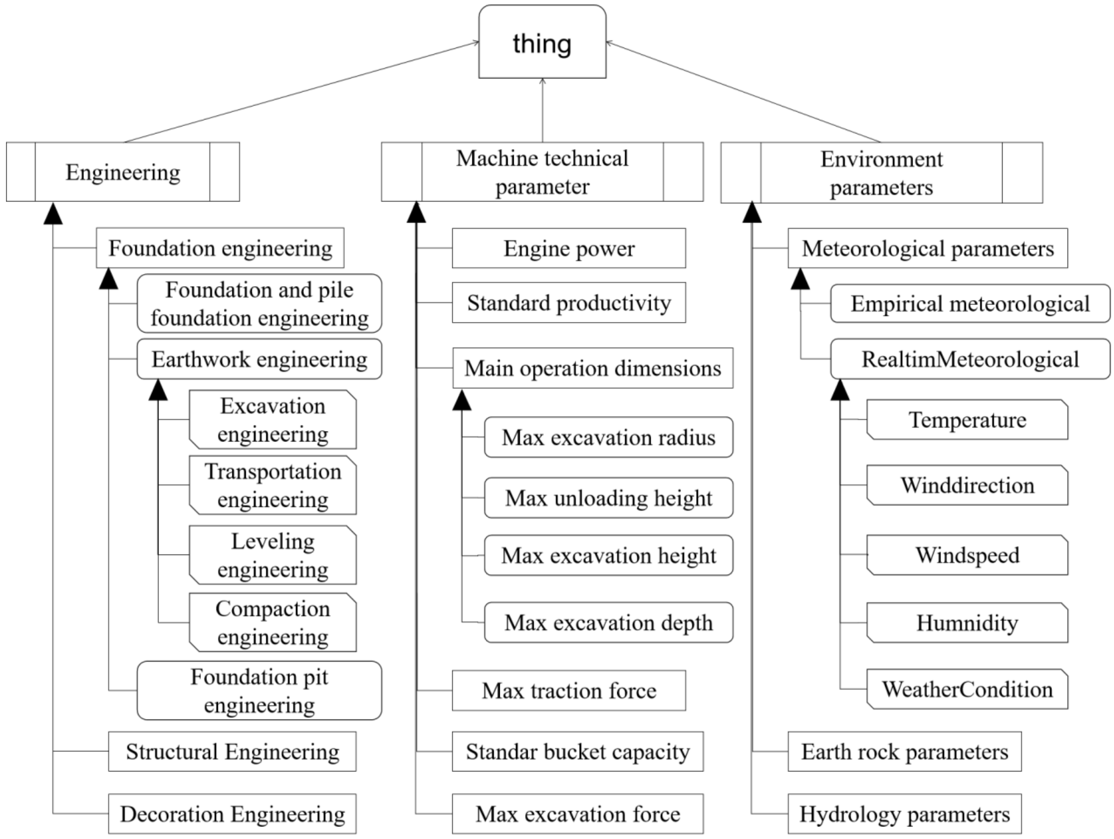

Fig. 2 Overview of the class hierarchy (partial)

3.3 Ontology conceptualization

Based on the ontology specification and extracted and reused terms, we have created an initial conceptual model for CEMO. **Fig.3** presents an overview of the primary classes and relationships in the relevant domain. The proposed ontology comprises 11 ontological modules: project, engineering, engineering quantity, environment, machinery, personnel, observation, data property, indicator, status, and KPI. As the observation, engineering quantity, and indicator modules provide the necessary data and related parameters for calculations, they are linked to the KPI module. The property module describes the properties of the observation data. Moreover, the KPI, observation, engineering quantity, and indicator modules have corresponding entities for evaluation, which may include engineering projects, engineering subprojects, construction teams, machinery, or individuals. Additionally, to identify the impact of the external environment on carbon emissions, the engineering background and natural environment modules are essential. The natural environment module provides meteorological parameters for observation. Furthermore, the KPI module provides parameters for assessing the project's carbon status. Subsequently, each detailed model of every module was developed using the conceptual model above as a guide.

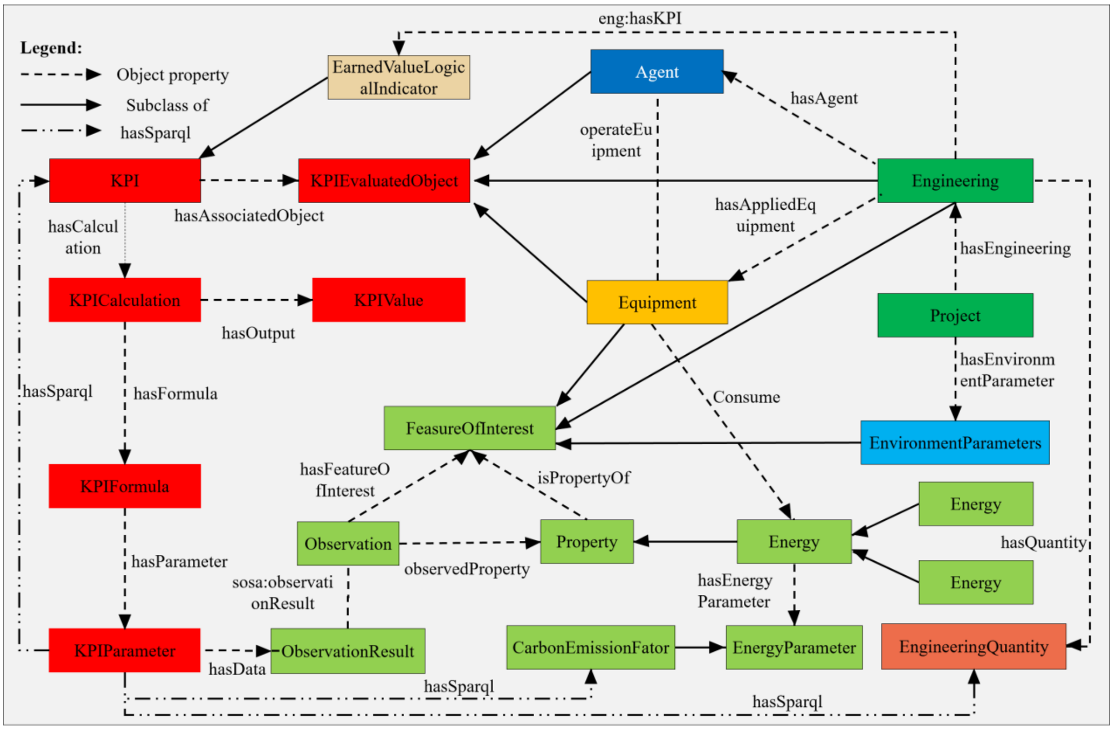

Fig. 3 The initial conceptual model of CEMO

3.4 Ontology implementation

Ontology development through modularization serves to not only formalize its structure but also guarantee the uniformity of terms, enabling flexible application and easy management. With reference to the conceptual model depicted in **Fig.3**, we have established exhaustive ontology modules according to distinct themes. In the ensuing subsections, we provide detailed information on each ontology module's model. For ease of comprehension, all components are denoted with their relevant module names as a prefix.

(1) Project module and Engineering module

Fig. 4 illustrates the intricate model of the construction project. Notably, each construction endeavor encompasses an array of *EnvironmentParameters*, *Engineering*, and *Projectbackground* parameters, such as *ProjectName*, *ConstructionSite*, *ConstructionArea*, and *ConstructionDuration*, as these elements are intimately linked to the carbon emissions emanating from construction activities. Additionally, each engineering feature comprises a comprehensive suite of details, including *ConstructionPlan*, *EngineeringQuantity*, *Machine*, *CarbomEmissionState*, and *ManagementMeasure*. Significantly, the engineering feature is a subclass of both *FeatureOfInterest* and *KPIEvaluatedObject*, highlighting the intricate interconnections between the diverse facets of the project. We have categorized *engineering* subclasses based on the building construction process. **Fig.5** illustrates the *Engineering* class subdivided into three subclasses: *StructuralEngineering*, *FoundationEngineering*, and *DecorativeEngineering*, representing the building's construction content. These subclasses can be further divided based on the specific engineering content.

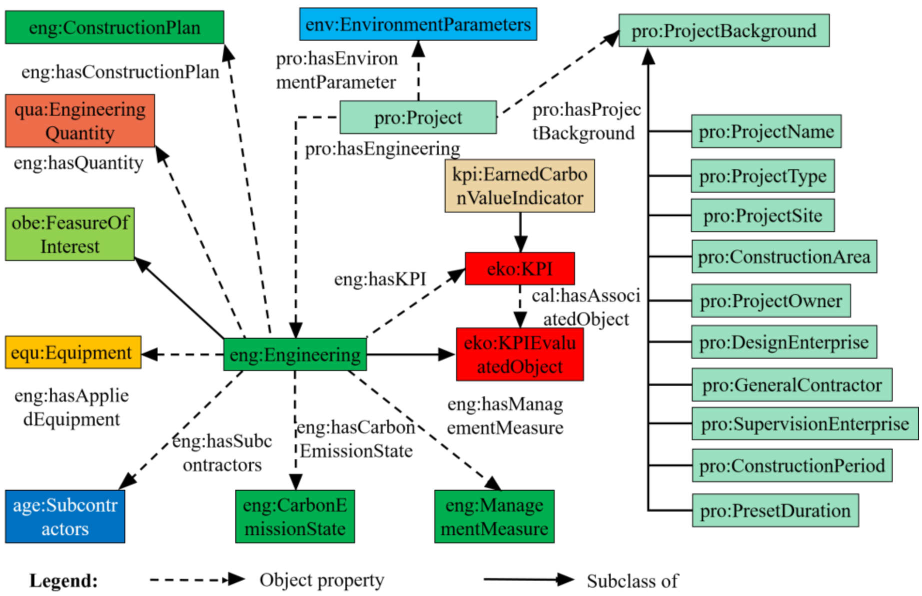

Fig.4 Detailed model of the project module

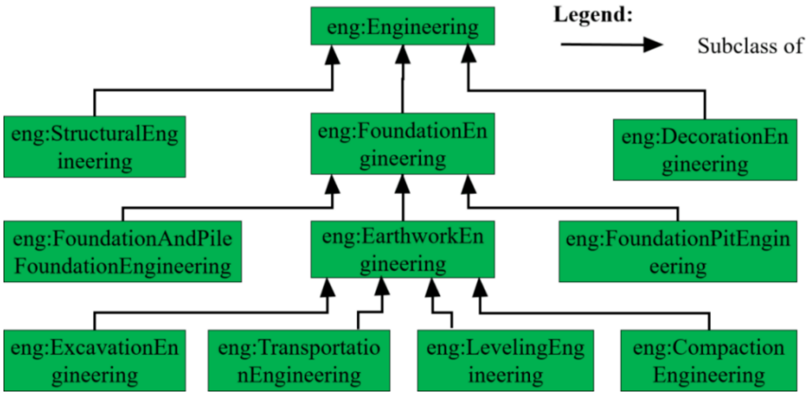

Fig. 5 Detailed model of the engineering module

(2) Quantity of works module

Fig. 6 illustrates a detailed model of the quantity of works module. The Quantity of Works class describes the amount of construction work and is divided into two subclasses: *QuantityWorkScheduled* and *QuantityWorkPerformed*. The former refers to the planned volume of work, including *TotalBudgetedQuantity (TBQ)*, *BudgetedDailyQuantityForWorkScheduled (BDQWS)*, and *BudgetedQuantityForWorkScheduled (BQWP)*. The latter pertains to real-time data based on construction progress statistics and can be further subdivided into *ActualQuantityForWorkPerformed (AQWP)* and *ActualDailyQuantityForWorkPerformed (ADQWP)*. Additionally, each specific engineering quantity corresponds to a *TemporalEntity* and *UnitOfMeasure*. Notably, the budgeted and actual daily engineering quantity can be updated to weekly, bi-weekly, or monthly actual or budgeted engineering quantity, as required by the project's specific needs.

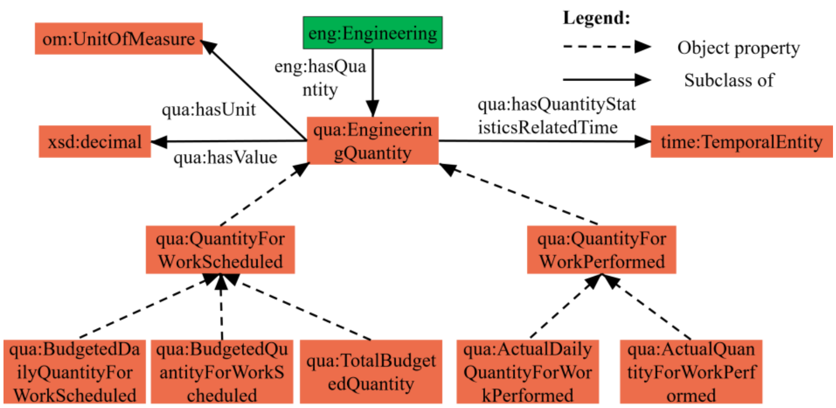

Fig.6 Detailed model of the Quantity of works module

(3) Environmental parameters module

The detailed modules of the environmental parameters of the construction project are presented in **Fig. 7**. The environment parameters class is defined to describe natural environmental information regarding the construction site. It is subdivided into three subclasses: *Earthrock*, *MeteorologicalEnvironment*, and *Hydrology*. *Earthrock* describes various construction land types and characteristics, such as *Density* and *EarthrockType*.The *Meteorological* class has subclasses, including *EmpiricalMeteorological* and *RealtimeMeteorological*, which describe the meteorological environment of the construction site.RealtimeMeteorological primarily reuses Weather ontology's terms [62] to describe real-time weather phenomena, including humidity, temperature, wind direction, wind speed, dust, and weather conditions. Moreover, *RealtimeMeteorological* is a property subclass in the Observation module, commonly obtained through sensors.

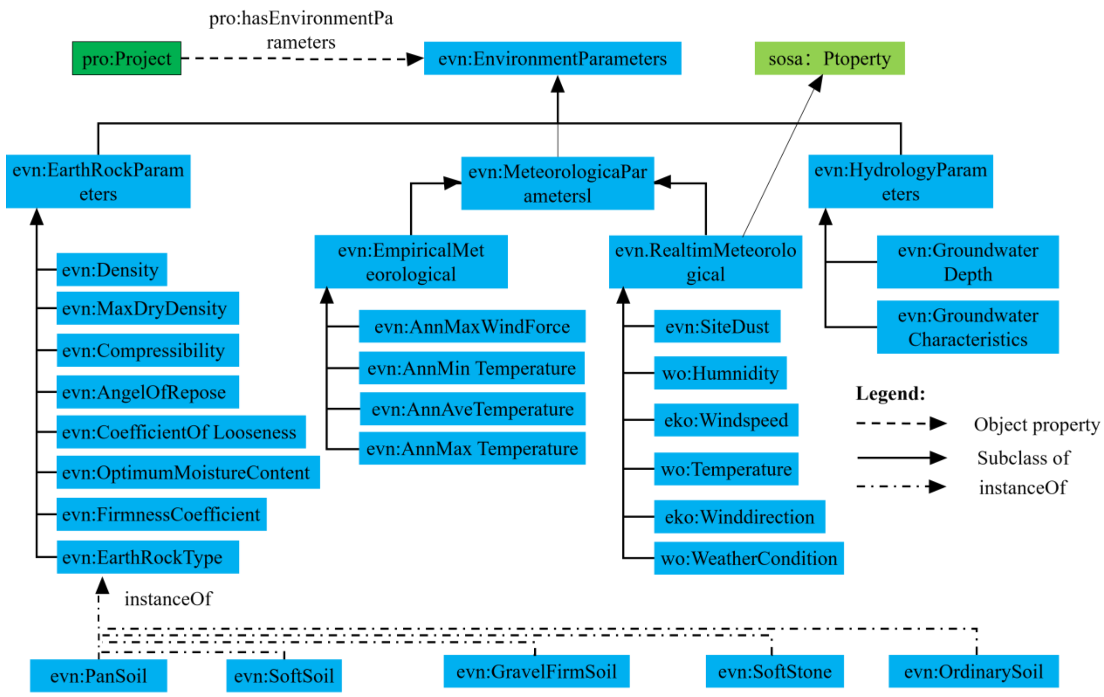

Fig.7 Detailed model of the Environmental parameters module

（4）Construction Equipment module

Fig. 8 presents a comprehensive model of the construction equipment module, where the *ConstructionEquipment* used in the project is the primary source of energy consumption and carbon emission during the construction stage. Each construction equipment unit includes a description of its engineering application, energy consumption, operators, basic information, technical parameters, and maintenance data. All of these elements are essential in understanding and monitoring the equipment's energy usage and carbon emission. Basic Information records crucial data about the construction machinery, such as its *EquipmentName*, *EquipmentType*, and *EquipmentManufacturer.* The TechnicalParameter outlines the performance parameters of the construction machinery, with each type of equipment having several such parameters. For instance, this study employs the example of an excavator to illustrate the construction machinery's technical parameter information. Proper maintenance of construction machinery has a direct impact on its energy efficiency and carbon intensity. The *MaintenanceRecord* captures seven essential maintenance details, including the maintenance address and maintenance time. Similar to the engineering class, the mechanical class can function as either a monitored object (*FeatureOfInterest*) or an evaluated object (*KPIEvaluatedObject*).

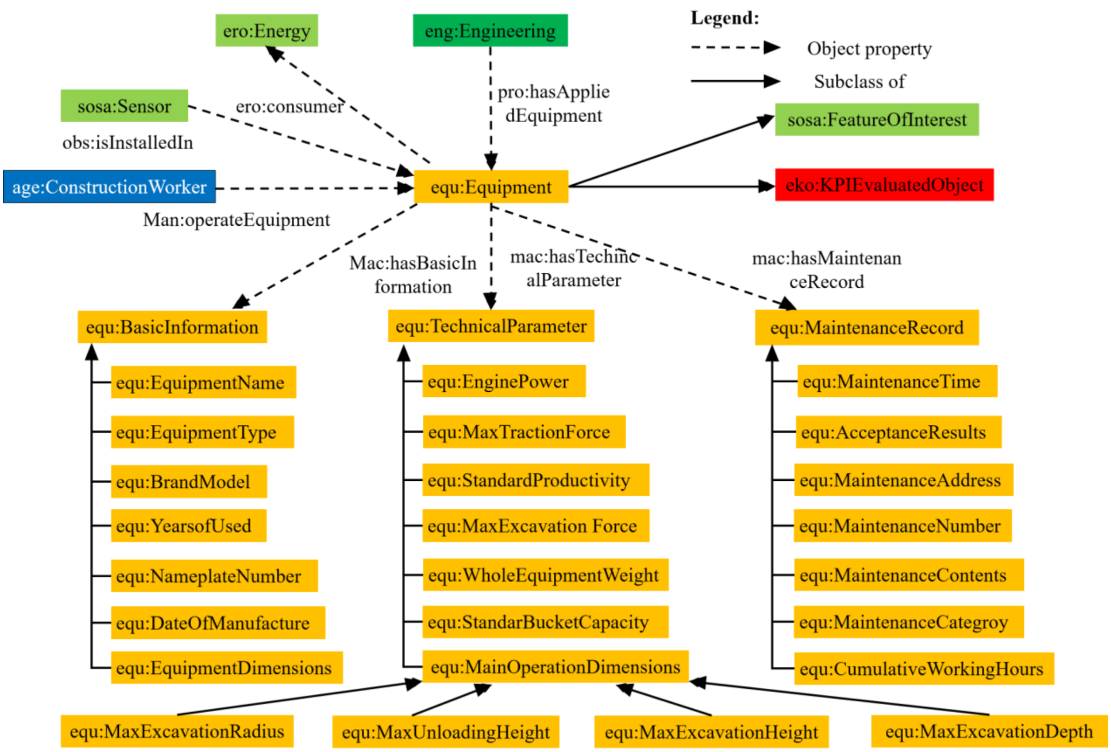

Fig. 8 Detailed model of the construction equipment module

（5）Agent module

Fig. 9 depicts the detailed model of Agents. The Agent class defines an entity responsible for construction activities and is categorized into various subclasses, including *CivilEngineeringConstructionTeam*. Typically, *Agents* employ skilled operators to operate construction equipment. For instance, *CivilEngineeringConstructionTeam* hires experienced excavator operators. Operators' experience and proficiency in operating equipment directly impact project progress, energy consumption, and carbon emissions. The module defines eight Static parameters to describe operators' basic information, such as *WorkNumber, Name, Gender, and Age.* Additionally, the *DynamicInformationOfOperator* class is subdivided into six subclasses, such as *AccumulatedWorkingTime* and *WorkEfficiency*, to represent real-time operator information.

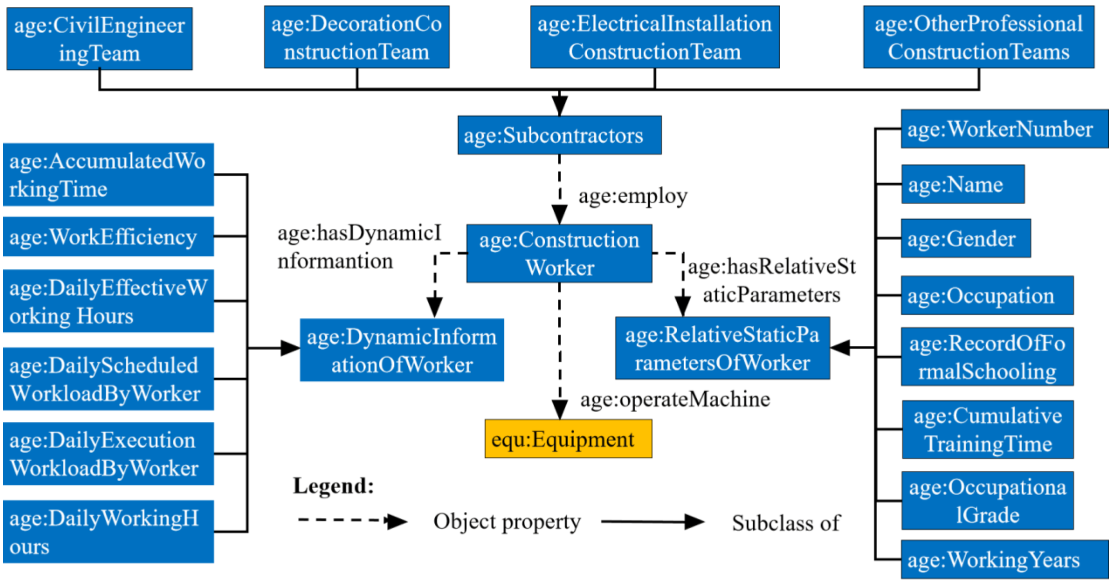

Fig. 9 Detailed model of the agent module

（6）Observation module

Fig.10 depicts a comprehensive model of the Observation module that largely adopts terms from the SOSA [67]/SSA ontology [68], Time ontology [69], and Om [70]/QUDT [71]. Fe*atureOfInterest* and *ObservationValue* represent core classes of the observation module, with the feature of interest encompassing operational engineering, construction equipment, construction teams, or their respective subclasses. Each observation corresponds to an *ObservationValue* that is accompanied by a *UnitOfMeasure*. Furthermore, each observation is associated with a TemporalEntity, which can be further subdivided into Interval and Instant. Finally, a specific sensor must be linked to the observation to capture the data source.

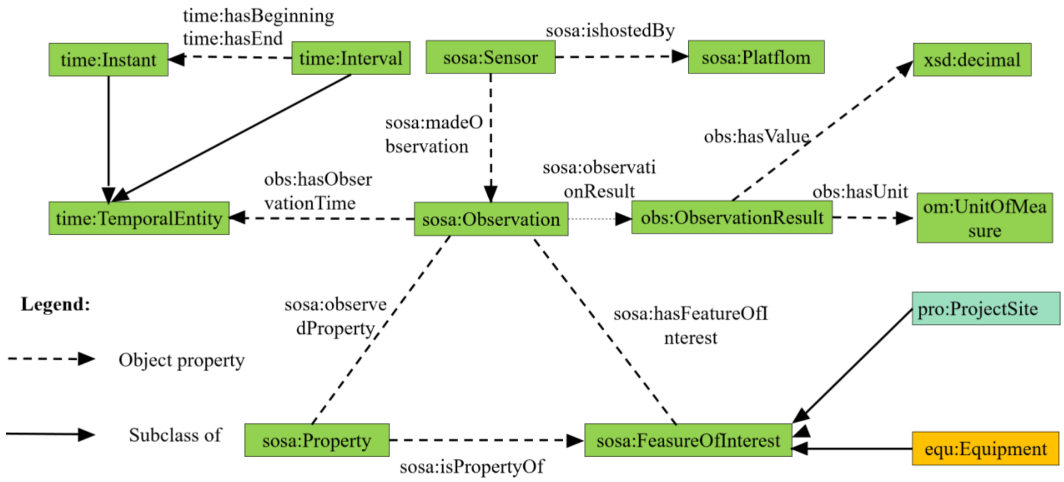

Fig. 10 Detailed model of the Observation module

Fig.11 illustrates another critical class of the Observation module, the PropertyClass, which comprises four domain information types, including weight and energy. These properties are essential for carbon emission calculation and performance analysis. The Energy class has several subclasses, such as *Electricity, NaturalGas, Gasoline*, and *DieselOil*, representing the different types of energy consumed on the construction site. *Energy* is associated with *EnergyParameter* via *hasEnergyParameter*. *EnergyParameter* describes the energy consumption characteristics at the construction site and is classified into *CarbonEmissionFactor, EnergyProvider, EnergySource*, and *EnergyType.*

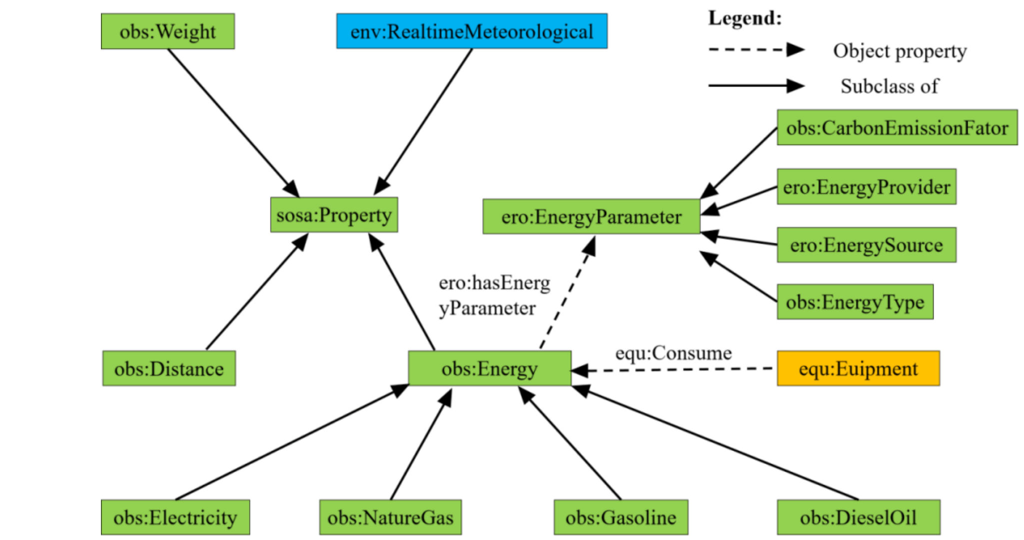

Fig. 11 Detailed model of the data property module

（7）KPI module

Fig.12 depicts a detailed model of the KPI module that delineates various key performance metrics for evaluating carbon emissions and energy consumption at construction sites. The *EarnedCarbonValueIndicator* (*ECVIndicator*) is a pivotal class of the indicator module, developed using earned value methodology and a novel contribution to this paper. *ECVIndicator* primarily characterizes deviations in carbon emissions during the construction process. This class is further divided into four subclasses: *CarbonEmissionIndicator*, *EnergyConsumptionIndicator*, *VarianceIndicator*, and *IndexIndicator*. To describe distinct aspects of carbon emissions, *CarbonEmissionIndicator* is subdivided into *CarbonEmissionValue* and *CarbonEmissionIntensity* subclasses. *CarbonEmissionValue* is further categorized into *EmissionPerformed* and *EmissionScheduled*. CarbonEmissionIntensity describes the EmissionPerformed and EmissionScheduled per unit engineering quantity. Additionally, each carbon emission indicator corresponds to an energy consumption indicator. Earned value logic sets *EnergyConsumptionVariance*, *CarbonEmissionVariance*, and *ScheduleVariance* to describe deviations between scheduled and performed values. Corresponding to the VarianceIndicator, the PerformanceIndexIndicator has three subclasses: *EnergyConsumptionPerformanceIndex*, *CarbonEmissionPerformanceIndex*, and *SchedulePerformanceIndex*.

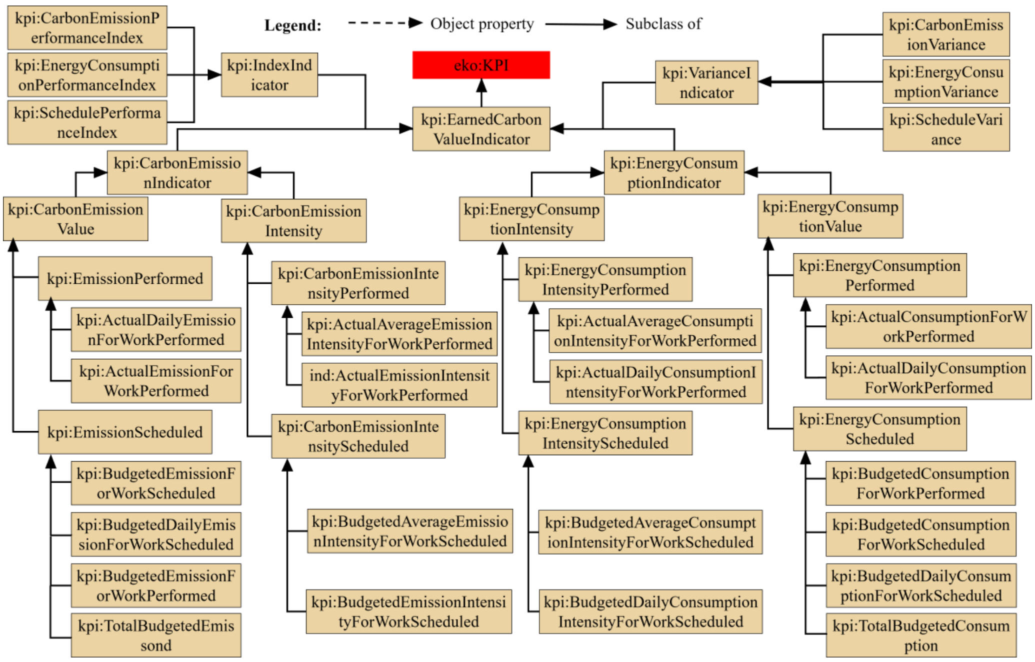

Fig. 12 Detailed model of the Indicator module

(8) Calculation module

The detailed model of the calculation module (See **Fig. 13**) is constructed using several ontologies, including EM-KPI Ontology [65], KPI Ontology [76], Om [70]/QUDT [71], and time ontology. The model provides a comprehensive representation of the KPI Class and its components, such as *Indicator, KPIEvaluatedObject, KPIcalculation,* and *TemporalEntity.* The calculation module can evaluate operational engineering, construction equipment, construction teams, and their subclasses. *KPICalculations* involve input *KPIFormulas* and output *KPIValues*, with the former described using the Operator class and *KPIParameter* class. The operator class comprises three main subclasses: Unary operators, Binary operators, and Aggregation. KPIParameter describes the parameters used in indicator calculations and can be categorized as *SimpleParameters*, *KPISingleParameter*, and *KPIListParameter. KPISingleParameter* and *KPIListparameter* can query related single instances or series of instances from the semantic web using the SPARQL statement. Depending on the input *KPIParameter*, the output KPIValue can be a *SimpleValue* or *ListValue*, and the *KPIValue* unit is based on Om [70]/QUDT [71] concepts and patterns. Each indicator undergoes evaluation during a specific period, with time description terms provided by Time ontology [69].

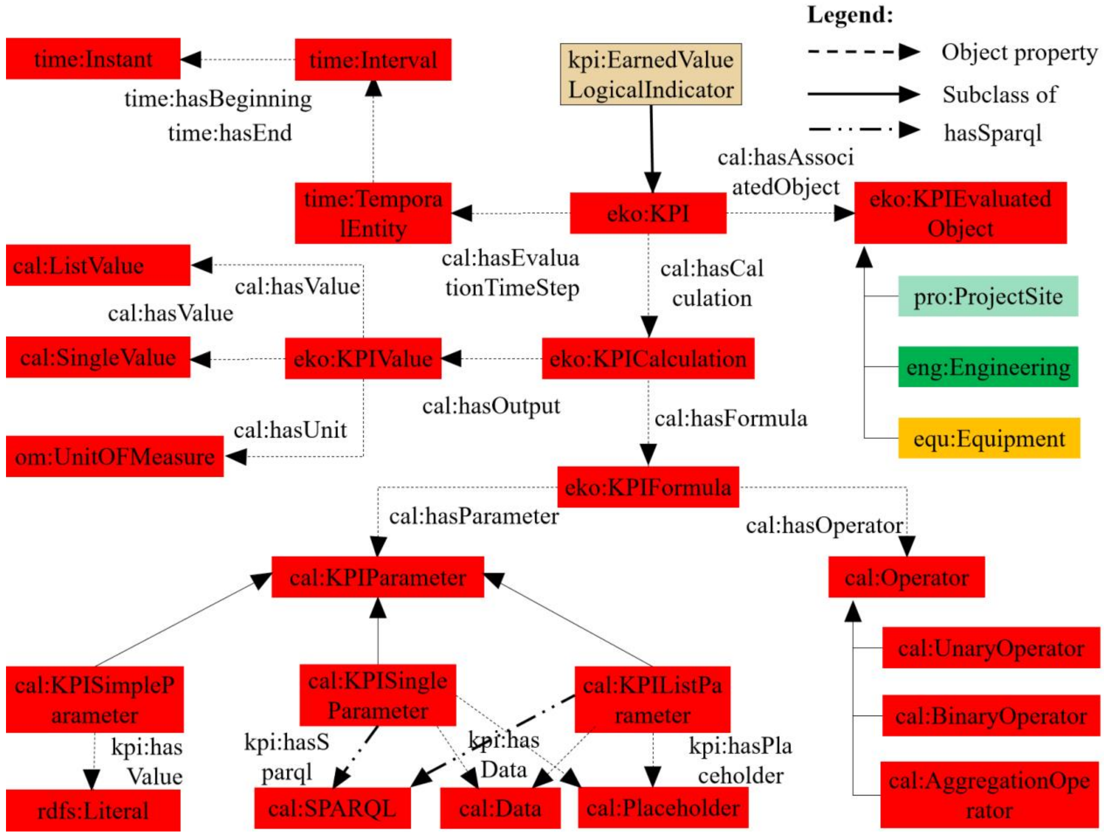

Fig. 13 Detailed model of the Calculation module

CEMO's modules are constructed through the implementation of the standard language, OWL, which is commonly utilized for defining ontologies. The Protégé5.5 development environment, an open-source and freely available platform for editing OWL ontologies, was employed to define CEMO.

4. Verification and evaluation

Ontology evaluation is a critical process in ontology development that entails assessing ontology content using a specific frame of reference [77]. To evaluate CEMO, we selected four criteria from prior research: clarity, coverage, consistency, and extensibility [18,78,79]. Ontology evaluation typically comprises technical developer and user evaluation, aimed at preventing unintentional deviation or self-deception during self-evaluation by technology developers [78]. The following subsections describe each evaluation approach and their respective outcomes.

4.1 Developer evaluation

In this section, we conducted a technical evaluation of CEMO, which included (1) manual evaluation of clarity, coverage, and extensibility, and (2) automatic evaluation of consistency.

Clarity is a vital intrinsic feature of an ontology, referring to its ability to effectively and unambiguously communicate the intended meaning of defined terms. To ensure the ontology's clarity, existing concepts and relational terms were extensively reused in this paper. New terms were defined according to existing professional knowledge and expert input was sought to refine the expressions of terms. The terms in CEMO were formally and unambiguously defined

Extensibility is a critical intrinsic characteristic of ontologies, enabling them to describe specific domains without altering the existing definitions within the ontology. To this end, CEMO was developed as a high-level vocabulary that can be extended with lower-level glossaries for specific contexts. For example, subclasses of the Engineering class can be expanded to encompass infrastructure projects.

Assessment of ontology coverage commonly involves answering competency questions (CQs). We conducted a manual verification of the developed ontology's concepts, relationships, and axioms against the predefined CQs in the ontology specification (see Section 3.1) to assess its comprehensiveness. Our analysis reveals that CEMO fully encompasses the terms and relationships outlined in the ontology specification.

Ensuring consistency is crucial in preventing conflicts or inconsistencies in the description logic of the developed ontology. For ontology consistency checks in this study, we utilized the Pellet reasoner, which can be run as a plugin in Protégé 5.5. Pellet is an open-source OWL DL reasoner based on Java and possesses robust reasoning capabilities [78]. After multiple iterations, the final version of CEMO was determined to be consistent, coherent, and free from logical reasoning conflicts.

4.2 User evaluation

Expert assessment, based on their specialized knowledge of ontology abstraction, classification, and coverage, is a fundamental approach to evaluating ontologies. In developing CEMO, we conducted an expert questionnaire survey to evaluate ontology content from the perspective of potential users. The survey involved four experts, including two distinguished construction management scholars and two seasoned practitioners, with more than five years of engineering experience and a track record of successful participation in at least two building construction projects. These experts provided valuable feedback on CEMO terminology. During the evaluation phase, each expert completed a comprehensive questionnaire comprising four parts: (1) introduction to the research motivation and scope; (2) description of the CEMO taxonomy, relations, and axioms; (3) ontology evaluation; and (4) proposal revision (Table 2). To record expert responses in Section 3, a 5-point Likert scale was employed, with 1 indicating strong disagreement and 5 indicating strong agreement.

The questionnaire results indicate that CEMO satisfies the requirements for clarity, extensibility, coverage, and consistency. Nevertheless, experts provided valuable feedback and suggestions for improvement. Some experts noted that certain concepts were used improperly or lacked clear definitions (GA(No.1=4.8)), which were then revised through discussion. For example, 'Machine' was originally defined as a construction activity facility and energy consumption entity, but experts pointed out that 'Equipment' was a more appropriate and encompassing term. Additionally, the experts recommended extending the CEMO to infrastructure construction projects such as bridges and tunnels, although this may require specific adjustments (GA (No. 3) = 4.6). Lastly, the experts proposed linking CEMO to other ontologies but suggested that new classes and semantic relationships should be added (GA (No. 4) = 4.3).

Table 2 The evaluation criteria and results of the developed ontology

No.	Evaluation criteria	Evaluation Content	Min	Max	Avg
1	Clarity	Concept terms are defined clearly and without ambiguity	4	5	4.8
2	Consistency	The class hierarchies and inter-concept relations are logical and complete	5	5	5
3	Extendibility	The developed ontology can be extended to be used in other areas	4	5	4.6

No.	Evaluation criteria	Evaluation Content	Min	Max	Avg
4		The developed ontology can be extended to interconnect with other ontologies	4	5	4.3
5	Coverage	Classes can cover the purpose of the ontology	5	5	5
6		The developed ontology covered the major concepts and inter-concept relations within the construction carbon emissions domain	5	5	5

5. Ontology Applications

5.1 KPI calculation

A semantic query language can retrieve multi-type static and dynamic information on construction sites, facilitating carbon management decision-making processes. **Fig.14** illustrates the automatic calculation process of indicators based on SPARQL queries. Before performing indicator calculations, relevant background parameters must be specified, including the period and the evaluation object associated with the indicators, which can be directly extracted from the semantic network. Indicators can be represented as formulas, parsed into a tree structure containing operators and parameters. In cases where a parameter is represented by another formula, the entire process requires re-parsing. The resulting calculation process is carried out step-by-step from bottom to top, with parameters divided into explicit and non-explicit values. While explicit values can be established directly, non-explicit values require extraction from the ontology semantic web through the execution of relevant SPARQL queries to replace placeholders. Fig.15 illustrates how to compute the energy consumption intensity of a project by dividing the energy consumption of the project within a specified period by its engineering quantity. Notably, this calculation involves a bottom-to-top approach, where the upper calculation parameters are the results of the lower formula. Parameters can be divided into explicit and non-explicit values, whereby explicit values can be readily established, but non-explicit values require ontology semantic web access. By executing relevant SPARQL queries, non-explicit parameter values can be obtained from the semantic web and inserted in place of placeholders. The operators can then be used to calculate the final result. The proposed semantic query language can retrieve multi-type static and dynamic information from construction sites and support carbon management decision-making processes.

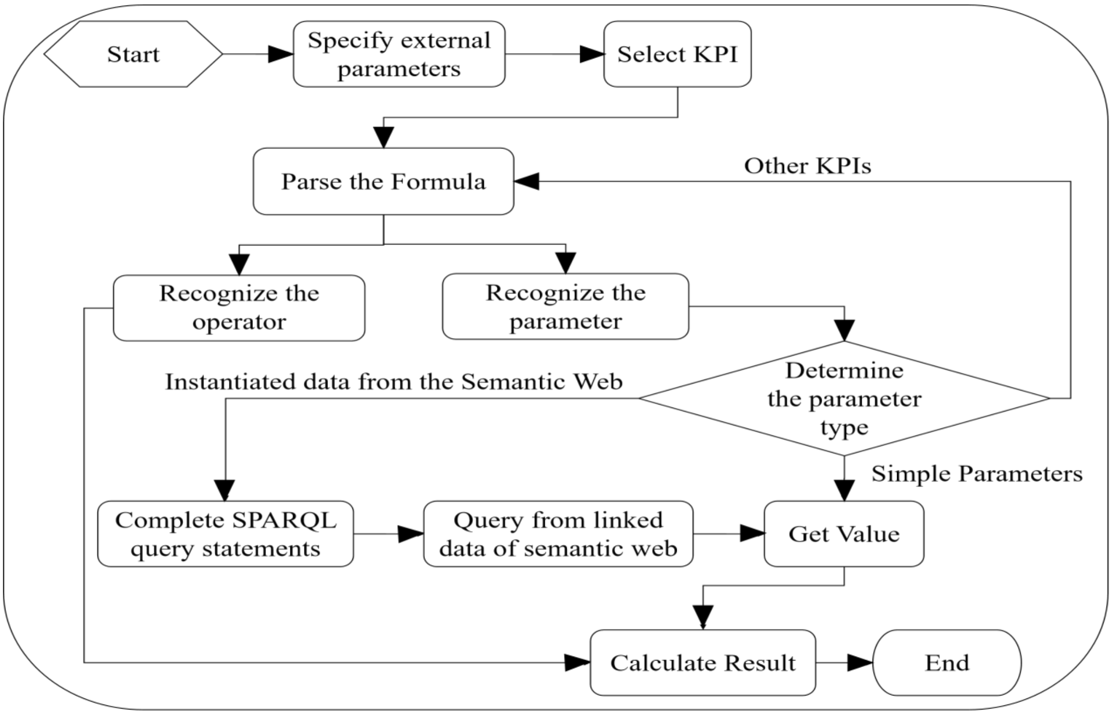

Fig.14 Automatic calculation process for indicators

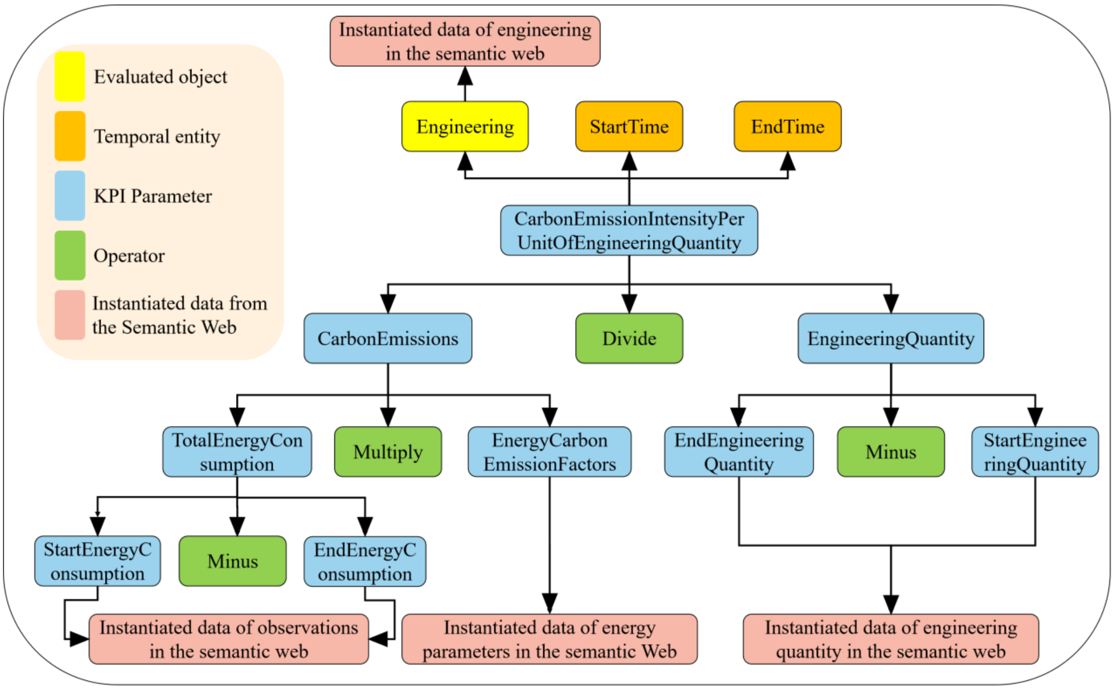

Fig. 15 Instantiated ontology for carbon emission intensity indicator per unit of work volume Engineering quantity

5.2 Semantic reasoning for earned carbon

The semantic inference engine, combined with customized semantic rules and transformed OWL language, facilitates the derivation of additional information to aid in carbon management decision-making. SWRL, a language for semantically presenting rules, consists of antecedent (body) and consequent (head), connected by the symbol “->”. Both antecedent and consequent comprise zero or more atoms, connected by the symbol “^”, with variables in atoms using the standard convention of prefixing them with a “?”. SWRLB is a built-in function of SWRL that enables the comparison of values or strings and basic mathematical calculations. As an example, the following SWRL-1 is expressed: the total carbon emissions (TotalCO2e) of a machine (M) is equal to the quantity of a particular type of energy consumed by the machine multiplied by the carbon emission factor (EF) of the energy.

SWRL-1	*Machine(?M) ^ consume(?M, ?E) ^Energy(?E) ^ Consumed(?E, ?Q) ^ EmissionFator(?E, ?EFE) ^ swrlb:multiply(?TCO2, ?Q, ?EFE) ->TotalCO2e(?M, ?TCO2e)*

Based on the interrelationships between the ECVIndicator metrics, six states of schedule-carbon emissions have been identified for the construction site, each of which requires targeted management measures. The following Rule-1 demonstrates one of the six relationships between the ECV indicators and the schedule-carbon emissions status.

Rule-1	Engineering(?a)^hasAssociated(?b,?a)^hasEvaluationTime(?b,?et)^Swrlb:date(?er,?year,?month,?day,?timezone)^hasAEWP(?a,?c)^hasBEWS(?a,?d)^hasBEWP(?a,?e)^hasSV(?a,?f)^hasCEV(?a,?g)^hasValue(?c,?h)^hasValue(?d,?i)^hasValue(?e,?j)^hasValue(?f,?k))^hasValue(?g,?l)^Swrlb:greaterThan(?h,?i)^Swrlb:greaterThan(?i,?j)^Swrlb:lessThan(?k,0)^Swrlb:lessThan(?l, 0)⇒CEState_1

In the rule expression, the variables “?aewp”, “?bews”, and “?bewp” represent the actual emission for work performed (AEWP), budgeted emission for work scheduled (BEWS), and budgeted emission for work performed (BEWP), respectively, while “?sv” and “?cev” denote the results of schedule variance and carbon emission variance, respectively. Rule-1 states that if AEWP>BCWS>BCWP & SC<0 & CEV<0, then the schedule-carbon emission state is CEState_1. CEState_1 is linked to detailed interpretations and management measures via the relationships hasInterpretation and hasMeasure, respectively. It is important to note that the Pellet inference engine was utilized to reason about SWRL rules in Protégé 5.5. The next section will provide a specific case to exemplify the SWRL reasoning rule.

5.3 Case study

We selected an engineering project to showcase how the use of CEMO can facilitate the seamless exchange of cross-domain data and carbon-related information. Our case study focuses on a residential building project in Shanghai, with an anticipated excavation volume of 48,000 m^3. The investigation period coincided with the earthwork construction stage of the project, which commenced on June 28, 2022. Given that excavator energy consumption constitutes a substantial proportion of carbon emissions during the earthwork construction stage, we paid particular attention to the carbon emissions resulting from excavators. For this case study, we excluded carbon emissions generated by transport trucks responsible for the off-loading of spoil. Carbon emissions for various sub-projects were allocated based on construction quotas during the planning stage of the project. For earth excavation, the anticipated carbon emissions were 10.73 tCO2e. Specifically, the excavator produced an estimated 223.55 gCO2e per unit volume of excavated soil.

The data involved in this study encompass the project background, natural environment, construction engineering, construction machinery, construction staffing, progress data, energy consumption indicators, carbon emission indicators, and other pertinent information. The statistical data, such as project background, geological and climate environments, construction engineering, construction machinery performance, construction staffing, and other related information, are manually extracted from construction documents and recorded in Excel format. The construction supervision team manually compiles daily construction progress information based on real-time statistics from the construction site, which is subsequently recorded in Excel format. Sensors installed on each construction machine collect data on machine performance and energy consumption every 30 minutes. The energy consumption data collected from the sensors, together with real-time weather data, are exported in CSV format. The aforementioned data were processed using the pure Python package RDFlib5.5, converted into ontology instances, and subsequently stored in RDF format. We created a visual interface based on Python for information querying, formula editing, and indicator calculation.

Fig. 16 illustrates the definition of the indicator formula, using engineering carbon intensity as an example. Before editing the formula, it is necessary to identify the evaluation objects and evaluation indicators. The carbon intensity of a project can be calculated by dividing the total carbon emissions over a selected period by the engineering quantity of the project. The total carbon emissions can be obtained by multiplying the total energy consumption by the carbon emission factor of the energy. The total energy consumption can be determined by subtracting the sensor observation values at the start and end times. Similarly, the total project quantity can be obtained by subtracting the quantity of work for the start and finish time operations. Energy consumption data, engineering quantity data, and energy carbon emission factors can all be queried on the Semantic Web using SPARQL, where engineering information in statements can be represented by placeholders. Once the evaluation object has been identified, the query can be completed by replacing the placeholders with the corresponding engineering information. By using the aforementioned formula, we queried the carbon intensity of earthworks between July 1, 2022, and July 31, 2022. As depicted in **Fig. 17(a)**, the carbon intensity of the earthworks has largely stabilized, with maximum and minimum values of 0.28 $kgCO_2e/m^3$ and 0.21 $kgCO_2e/m^3$, respectively.

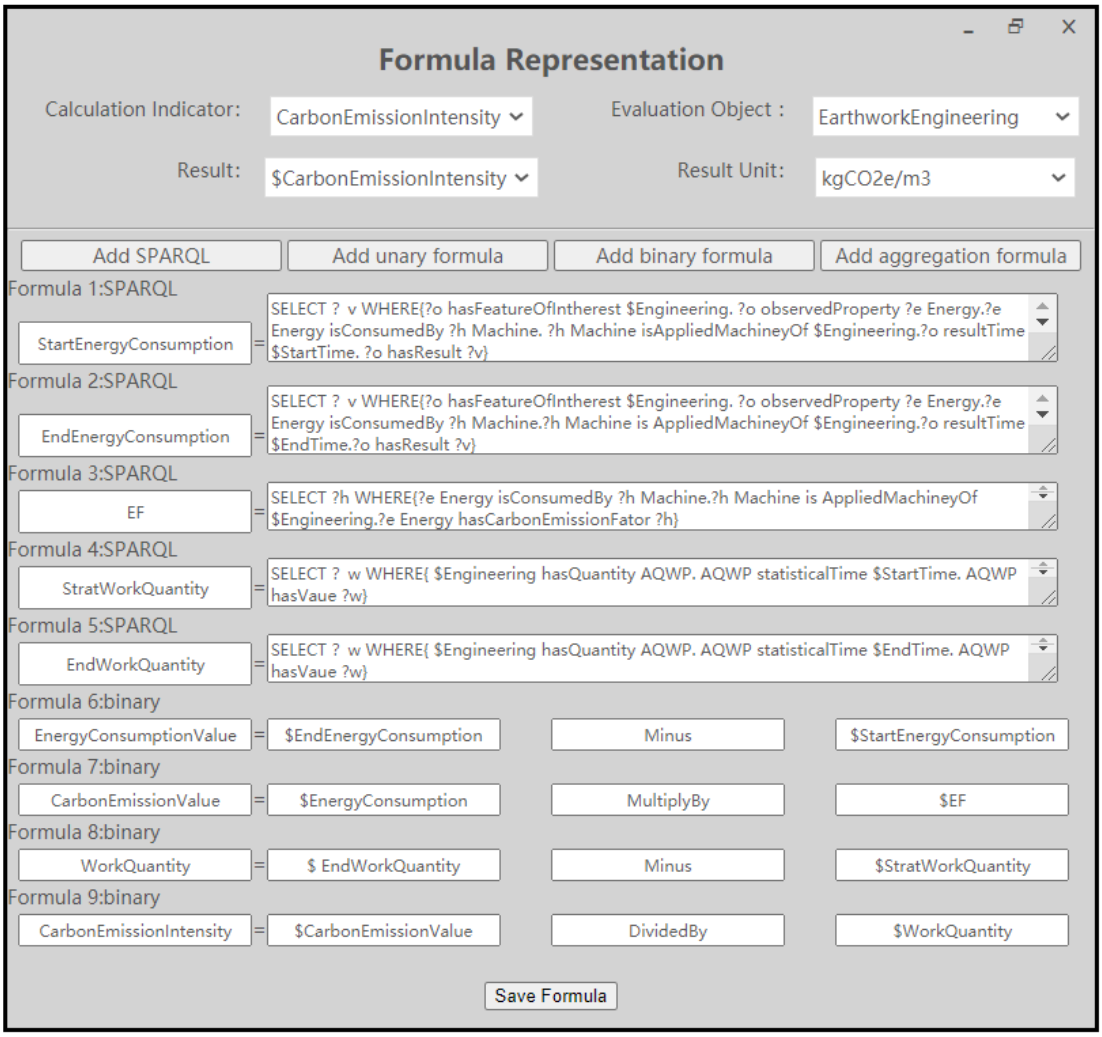

Fig. 16 Definition of carbon emission intensity per unit of engineering quantity

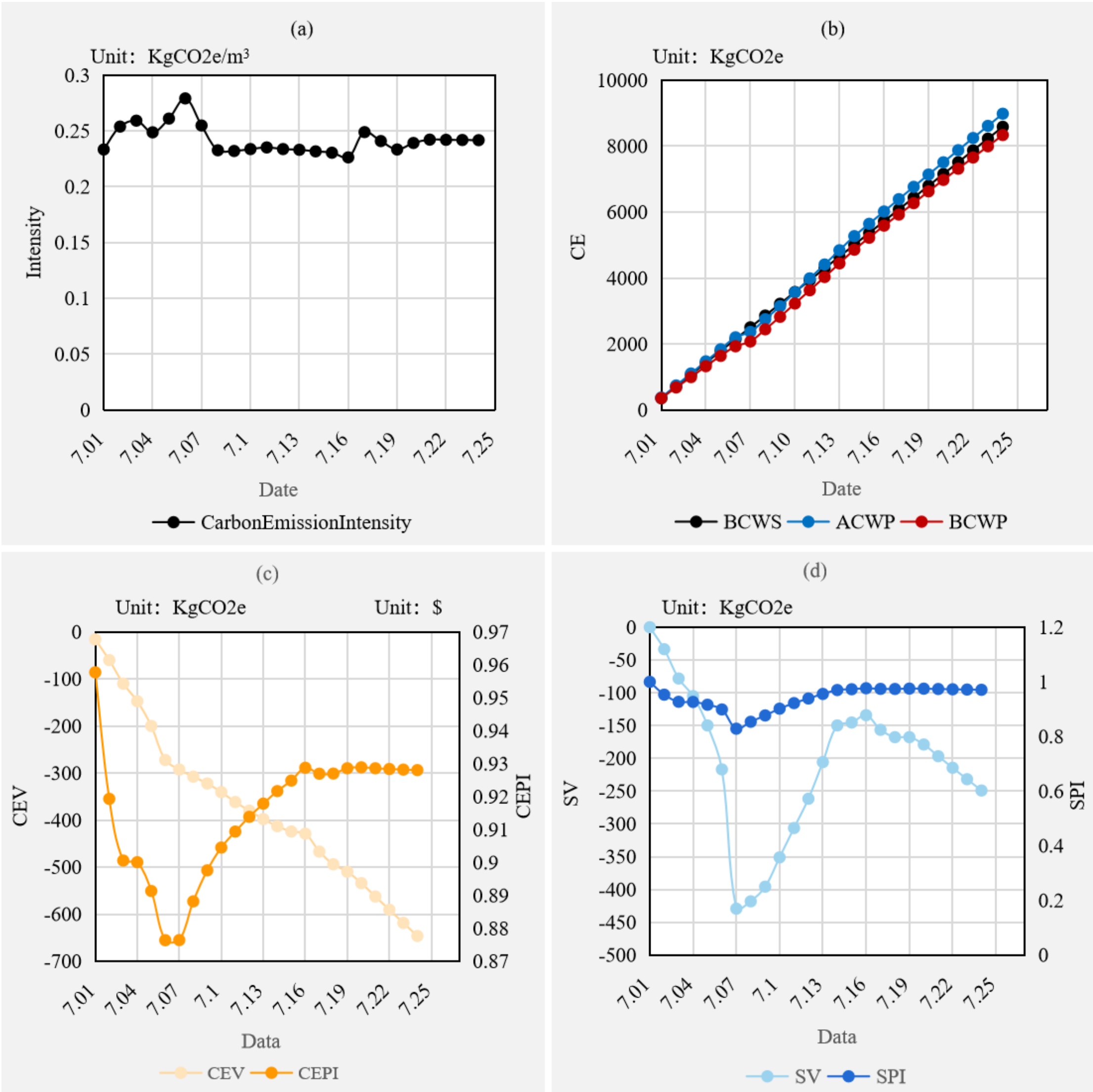

Fig. 17 Calculated results for selected indicators. (a) Carbon emission intensity; (b) Carbon emission ;(c) Carbon emission Variance and carbon emission performance index ;(d) Schedule Variance and Schedule performance index

To acquire a comprehensive understanding of the carbon emissions and construction progress at the construction site, we analyzed the formulas for the construction earned value series of indicators, as illustrated in **Fig. 18**. There are seven series of indicators associated with the logic of earned value: budgeted emission for work performed (BEWP), budgeted emission for work scheduled (BEWS), actual emission for work performed (ACWP), carbon emission variance (CEV), carbon emission performance index (CEPI), schedule variance (SV), and schedule performance index (SPI).

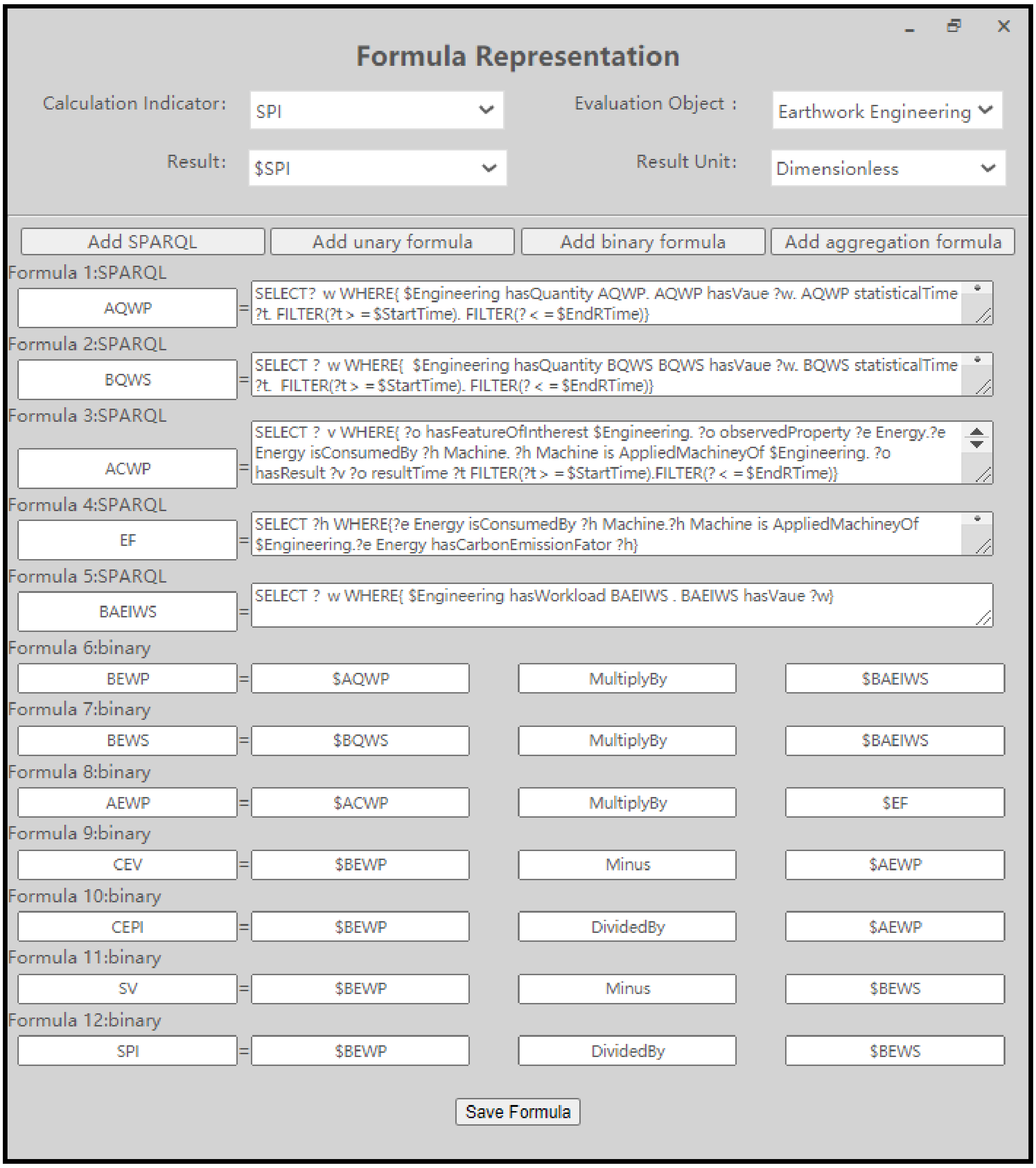

Fig.18 Definition of Earned Value Logical Indicators

Fig. 17 (b-d) presents the query results for the series of earned value logic indicators. The results suggest that BEWS was greater than AEWP and BEWP during the period from July 7 to July 10, while AEWP was greater than BEWS and BEWP during the period from July 12 to July 24. Examples derived from the EVL indicators can provide further information. In this case, we employed an SWRL rule to establish the "hasCarbonEmissionState" relationship between construction content and carbon emission status, as illustrated in **Fig. 19**. The rule language of the Semantic Web can be leveraged to generate automated reasoning, facilitating the dynamic assessment of the relationship between construction progress and carbon emissions at any given point in time, thereby offering valuable insights for decision-making. For instance, this approach can help evaluate the efficiency of construction personnel and equipment and assist in replacing less efficient personnel with more effective ones.

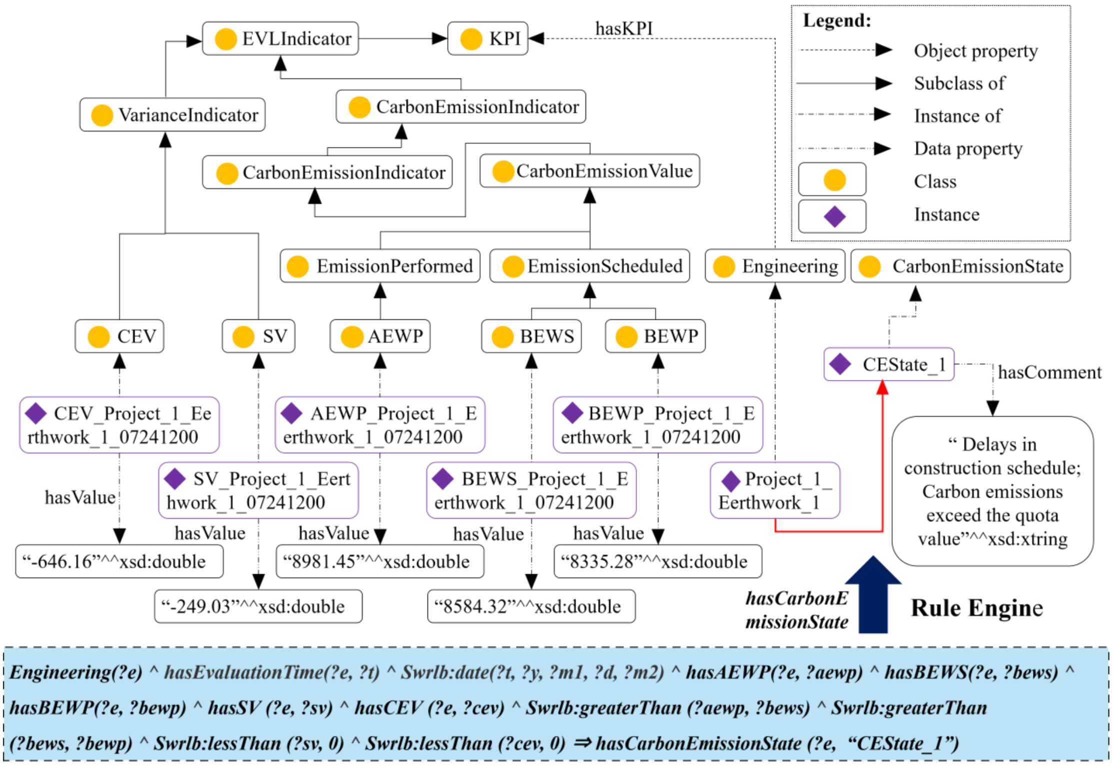

Fig. 19 SWRL Rule for reasoning '*hasCarbonEmissionState*' object property

6. Discussion

6.1 CEMO enables digital twin-based construction site carbon emissions management with its scalability

CEMO, developed through a comprehensive understanding of construction carbon emissions and earned value theory, serves as a paradigm for integrating carbon emissions information and heterogeneous data from multiple sources at construction sites. CEMO links information from various domains, such as engineering, machinery, personnel, predefined carbon emissions, project progress, and the environment, which impact carbon emissions. It also includes terminology for describing engineering carbon performance targets, allowing for assessment at different granularity levels. Furthermore, CEMO leverages existing ontologies, enhancing its scalability and flexibility for integration with other construction site data.

With the wide implementation of modern technologies such as BIM, GIS, IoT, cloud computing, big data, and computer vision in civil engineering construction, digital twin-based intelligent management of construction sites is becoming feasible. The developed ontology in this study supports the integration of diverse dynamic and static data from multiple sources, such as personnel, machinery, law, and environment, in common formats. It also allows for easy merging with other data through methods such as modification, extension, reuse, alignment, etc. Additionally, the RDF data storage of the ontology enables deep mining of construction site carbon emission data using cloud and big data technologies, providing comprehensive information feedback. Thus, CEMO facilitates the implementation of a digital twin system for construction sites, enabling the refined management of carbon emissions.

6.2 CEMO facilitates information retrieval and knowledge reasoning

CEMO demonstrates robust support for information retrieval and implicit knowledge inference through linked data. The case demonstrations highlight CEMO's ability to integrate heterogeneous data from multiple sources and showcase its cross-domain information retrieval and knowledge inference capabilities. Encoded with description logic, CEMO performs implicit knowledge inference based on defined correlation logic, further enhanced by SWRL-based correlation rules. Leveraging SPARQL for information retrieval, CEMO enables various carbon emission information analysis functions beyond carbon performance, such as dynamic relationship analysis between energy consumption and carbon emissions, time-varying law research of construction carbon emissions, optimization of construction machinery deployment considering carbon emissions, and assessment of construction machinery efficiency.

6.3 CEMO facilitates the disclosure of construction carbon emissions

CEMO can serve as a valuable tool in facilitating the disclosure of carbon information in the AEC industry. With an increasing focus on Environmental, Social, and Governance (ESG) disclosure, AEC enterprises are required to enhance their ability to manage carbon footprint data and other relevant information. CEMO contributes to standardizing the types of carbon emission KPIs and the statistical granularity of the data, which is essential for financial institutions and other entities that require accurate carbon emissions information for green financing purposes. Through its dynamic and real-time acquisition of various carbon emissions performance metrics, such as carbon quotas and assessments of carbon reduction actions, CEMO enables individual enterprises to improve their carbon management and carbon asset operation capabilities, while also supporting competent authorities in supervising carbon emissions in the AEC industry. The development of CEMO in this study represents a significant contribution towards building carbon information disclosure platforms for the AEC industry, making it a core technology in this domain.

7. Conclusion

Quantifying carbon emissions on construction sites and assessing real-time carbon emission deviations can greatly aid in the fine-grained management of carbon emissions. However, the fragmented and heterogeneous nature of construction data poses challenges in achieving these objectives. To address this issue and support the optimization of onsite carbon emission performance, this study proposes a carbon emission performance assessment ontology that facilitates cross-domain linking of carbon-related information.

CEMO, developed through collaboration with domain experts, is designed as a modular ontology. It comprises five domain modules, including KPI, Indicator, and quantity of works. The KPI module facilitates interactions between relevant Indicators and objects to be assessed, while other domain ontologies enable cross-domain sharing of engineering-related information. Evaluation results demonstrate that CEMO is a standardized semantic model with extensibility and applicability. A case study showcases its capabilities in supporting cross-domain data searching, carbon emission status assessment, and logical reasoning for decision-making. Furthermore, with the addition of more project-associated data, CEMO can support various functions for carbon emissions management, such as mechanical efficiency assessment, team construction efficiency comparison, and mining of coupling relationships between carbon emission and energy data.

This study makes three fundamental contributions. Firstly, CEMO introduces a comprehensive set of terminology that enables the expression of construction-site carbon emission information, thereby supporting the integration of multi-source and heterogeneous data. Secondly, the linked data based on CEMO facilitates information retrieval and knowledge inference for performance optimization. Lastly, CEMO demonstrates high scalability and the potential for linking with cross-domain information data, such as BIM, enabling digital twin-style management of onsite carbon emissions. Furthermore, an indirect impact of CEMO is its potential to promote the disclosure of carbon information in the AEC industry.

Future research efforts may focus on: enhancing the practicality of CEMO through the incorporation of automatic or semi-automatic data mapping tools, such as deep learning methods for information extraction and data mapping; exploring additional application potentials of CEMO, such as preset prediction and fine-grained control of carbon emissions on construction sites through joint knowledge inference about various factors; and investigating the implementation of CEMO for recommending low-carbon solutions under specific construction conditions, leveraging the accumulation of large-scale engineering cases. These directions could further enhance CEMO's functionality and applicability in the AEC industry.

Acknowledgments

This study was financially supported by the National Natural Science Foundation of China (Grants No. 52078374 and 52108090), the National Key Research and Development Program of China (Grant No. 2022YFC3801700), the Science Foundation for the Science and Technology Commission of Shanghai Municipality (Grants No. 22dz1207100 and 22dz1207800), and the Top Discipline Plan of Shanghai Universities-Class I.

References

[1] Burek, J., & Nutter, D. Life cycle assessment of grocery, perishable, and general

merchandise multi-facility distribution center networks. Energy and Buildings.174(2018), 388-401. https://doi.org/10.1016/j.enbuild.2018.06.021

[2] Eia, U. S. Annual energy outlook 2015 with projections to 2040. United States Energy Information Administration, Washington.(2015).

[3] Peng, C. Calculation of a building's life cycle carbon emissions based on Ecotect and building information modeling. Journal of Cleaner Production, 112(2016), 453-465. https://doi.org/10.1016/j.jclepro.2015.08.078

[4] Hong, J., Shen, G. Q., Feng, Y., Lau, W. S. T., & Mao, C. Greenhouse gas emissions during the construction phase of a building: A case study in China. Journal of cleaner production, 103(2015), 249-259. https://doi.org/10.1016/j.jclepro.2014.11.023

[5] Chou, J. S., & Yeh, K. C. Life cycle carbon dioxide emissions simulation and environmental cost analysis for building construction. Journal of Cleaner Production, 101(2015), 137-147. https://doi.org/10.1016/j.jclepro.2015.04.001

[6] Ahn, C. R., Lewis, P., Golparvar-Fard, M., & Lee, S. Integrated framework for estimating, benchmarking, and monitoring pollutant emissions of construction operations. Journal of Construction Engineering and Management, 139(12)(2013), A4013003. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000755

[7] Rasdorf, W., Lewis, P., Marshall, S. K., Arocho, I., & Frey, H. C. Evaluation of on-site fuel use and emissions over the duration of a commercial building project. Journal of infrastructure systems, 18(2)(2012), 119-129. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000071

[8] Sandanayake, M., Zhang, G., Setunge, S., Li, C. Q., & Fang, J. Models and method for estimation and comparison of direct emissions in building construction in Australia and a case study. Energy and Buildings, 126(2016), 128-138. https://doi.org/10.1016/j.enbuild.2016.05.007

[9] Sandanayake, M., Zhang, G., Setunge, S., Luo, W., & Li, C. Q. Estimation and comparison of environmental emissions and impacts at foundation and structure construction stages of a building–A case study. Journal of cleaner production, 151(2017), 319-329. https://doi.org/10.1016/j.jclepro.2017.03.041

[10] Szamocki, N., Kim, M. K., Ahn, C. R., & Brilakis, I. Reducing greenhouse gas emission of construction equipment at construction sites: Field study approach. Journal of Construction Engineering and Management, 145(9)(2019), 05019012. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001690

[11] Li, L., & Chen, K. Quantitative assessment of carbon dioxide emissions in construction projects: A case study in Shenzhen. Journal of cleaner production, 141 (2017), 394-408. https://doi.org/10.1016/j.jclepro.2016.09.134

[12] Fang, Y., Lu, X., & Li, H. A random forest-based model for the prediction of construction-stage carbon emissions at the early design stage. Journal of Cleaner Production, 328 (2021), 129657. https://doi.org/10.1016/j.jclepro.2021.129657

[13] Xiaodong, L., Fan, Y., Yuanxue, G., & McCarthy, J. Case study of carbon footprint of

residential building construction. Materials Research Innovations, 18(2014), S4-72. https://doi.org/10.1179/1432891714Z.000000000647

[14] E.G. Ortiz-García, S. Salcedo-Sanz, C. Casanova-Mateo, A. Paniagua-Tineo, A. Portilla-Figueras.Accurate local very short-term temperature prediction based on synoptic situation support vector regression banksAtmos. Res., 107 (2012), pp. 1-8

[15] S. Liu, M. Li, Z. Zhang, B. Xiao, X. Cao. Multimodal ground-based cloud classification using joint fusion convolutional neural network.Remote Sens., 10 (6) (2018), p. 822

[16] Ching-Tang,Fan,Yuan-Kai, et al. 2017. Heterogeneous information fusion and visualization for a large-scale intelligent video surveillance system. IEEE Transactions on Systems Man & Cybernetics Systems, 2017,4(47):593-604.

[17] Wu, C., Wu, P., Wang, J., Jiang, R., Chen, M., & Wang, X.Ontological knowledge base for concrete bridge rehabilitation project management. Automation in construction, 121 (2021)., 103428. https://doi.org/10.1016/j.autcon.2020.103428

[18] Zheng, Y., Törmä, S., & Seppänen, O. A shared ontology suite for digital construction workflow. Automation in Construction, 132(2021), 103930. https://doi.org/10.1016/j.autcon.2021.103930

[19] Ren, G., Li, H., Liu, S., Goonetillake, J., Khudhair, A., & Arthur, S. Aligning BIM and ontology for information retrieve and reasoning in value for money assessment. Automation in Construction, 124(2021), 103565. https://doi.org/10.1016/j.autcon.2021.103565

[20] Hou, S., Li, H., & Rezgui, Y. Ontology-based approach for structural design considering low embodied energy and carbon. Energy and Buildings, 102(2015), 75-90. https://doi.org/10.1016/j.enbuild.2015.04.051

[21] Abanda, F. H., Oti, A. H., & Tah, J. H. M. Integrating BIM and new rules of measurement for embodied energy and CO2 assessment. Journal of Building Engineering, 12(2017), 288-305. https://doi.org/10.1016/j.jobe.2017.06.017

[22] Geng Haochen. Research on carbon emission calculation and emission reduction measures for urban housing based on life-cycle assessment.Master's degree thesis(2021), Xi`an University Of Science And Technology. https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202201&filename=1021839727.nh

[23] Xikai, M., Lixiong, W., Jiwei, L., Xiaoli, Q., & Tongyao, W. Comparison of regression models for estimation of carbon emissions during building's lifecycle using designing factors: a case study of residential buildings in Tianjin, China. Energy and Buildings, 204(2019), 109519. https://doi.org/10.1016/j.enbuild.2019.109519

[24] Victoria, M. F., & Perera, S. Parametric embodied carbon prediction model for early stage estimating. Energy and Buildings, 168(2018), 106-119. https://doi.org/10.1016/j.enbuild.2018.02.044

[25] Lotteau, M., Loubet, P., & Sonnemann, G. (2017). An analysis to understand how the shape of a concrete residential building influences its embodied energy and embodied carbon. Energy

and Buildings, 154(2017), 1-11. https://doi.org/10.1016/j.enbuild.2017.08.048

[26] Gardezi, S. S. S., Shafiq, N., Zawawi, N. A. W. A., Khamidi, M. F., & Farhan, S. A. A multivariable regression tool for embodied carbon footprint prediction in housing habitat. Habitat International, 53(2016), 292-300. https://doi.org/10.1016/j.habitatint.2015.11.005

[27] Ye kunhui. Low carbon construction - from construction site to industrial ecology.China Construction Industry Press(2017), Beijing.

[28] Ding, Z., Liu, S., Luo, L., & Liao, L. (2020). A building information modeling-based carbon emission measurement system for prefabricated residential buildings during the materialization phase. Journal of Cleaner Production, 264(2020), 121728. https://doi.org/10.1016/j.jclepro.2020.121728

[29] Hawdon, D., & Pearson, P. Input-output simulations of energy, environment, economy interactions in the UK. Energy economics, 17(1)(1995), 73-86. https://doi.org/10.1016/0140-9883(95)98908-M

[30] Zhang, X., & Wang, F. (2016). Hybrid input-output analysis for life-cycle energy consumption and carbon emissions of China's building sector. Building and Environment, 104(2016), 188-197. https://doi.org/10.1016/j.buildenv.2016.05.018

[31] Sandanayake, M., Zhang, G., & Setunge, S. Environmental emissions at foundation construction stage of buildings–Two case studies. Building and Environment, 95(2016), 189-198. https://doi.org/10.1016/j.buildenv.2015.09.002

[32] Bastviken, D., Wilk, J., Duc, N. T., Gålfalk, M., Karlson, M., Neset, T. S., ... & Sundgren, I. Critical method needs in measuring greenhouse gas fluxes. Environmental Research Letters, 17(10)(2022), 104009. https://doi.org/10.1088/1748-9326/ac8fa9

[33] Li, D., Huang, G., Zhang, G., & Wang, J. (2020). Driving factors of total carbon emissions from the construction industry in Jiangsu Province, China. Journal of Cleaner Production, 276(2020), 123179. https://doi.org/10.1016/j.jclepro.2020.123179

[34] Nebel, B., Alcorn, A., & Wittstock, B. Life cycle assessment: adopting and adapting overseas LCA data and methodologies for building materials in New Zealand. Wellington, New Zealand: Ministry of Agriculture and Forestry (2011).

[35] Lu, Y., & Song, X. Beyond boundaries: A global use of life cycle inventories for construction materials. Journal of cleaner production, 156(2017), 876-887. https://doi.org/10.1016/j.jclepro.2017.04.010

[36] Chen, W., Yang, S., Zhang, X., Jordan, N. D., & Huang, J. Embodied energy and carbon emissions of building materials in China. Building and Environment, 207(2022), 108434. https://doi.org/10.1016/j.buildenv.2021.108434

[37] Wang, J., Wu, H., Duan, H., Zillante, G., Zuo, J., & Yuan, H. Combining life cycle assessment and Building Information Modelling to account for carbon emission of building demolition waste: A case study. Journal of cleaner production, 172(2018), 3154-3166. https://doi.org/10.1016/j.jclepro.2017.11.087

[38] Li, Y., Lu, Y., & Chen, J. A deep learning approach for real-time rebar counting on the construction site based on YOLOv3 detector. Automation in Construction, 124(2021), 103602. https://doi.org/10.1016/j.autcon.2021.103602

[39] Liu, G., Chen, R., Xu, P., Fu, Y., Mao, C., & Hong, J. Real-time carbon emission monitoring in prefabricated construction. Automation in Construction, 110(2020), 102945. https://doi.org/10.1016/j.autcon.2019.102945

[40] Liu, G., Yang, H., Fu, Y., Mao, C., Xu, P., Hong, J., & Li, R. Cyber-physical system-based real-time monitoring and visualization of greenhouse gas emissions of prefabricated construction. Journal of Cleaner Production, 246(2020), 119059. https://doi.org/10.1016/j.jclepro.2019.119059

[41] Carbon Trust Closing the gap: lessons learned on realising the potential of low carbon building design Technical Report CTG047, The Carbon Trust (2012) URL: http://www.carbontrust.com/media/81361/ctg047-closing-the-gap-low-carbon-building-design.pdf

[42] Lipke, W., Zwikael, O., Henderson, K., & Anbari, F. (2009). Prediction of project outcome: The application of statistical methods to earned value management and earned schedule performance indexes. International journal of project management, 27(4), 400-407. https://doi.org/10.1016/j.ijproman.2008.02.009

[43] Anbari, F. T.. Earned value project management method and extensions. Project management journal, 34(4)(2003), 12-23. https://doi.org/10.1177/875697280303400403

[44] Abdi, A., Taghipour, S., & Khamooshi, H. A model to control environmental performance of project execution process based on greenhouse gas emissions using earned value management. International Journal of Project Management, 36(3)(2018), 397-413. https://doi.org/10.1016/j.ijproman.2017.12.003

[45] Kim, J., Koo, C., Kim, C. J., Hong, T., & Park, H. S. Integrated CO2, cost, and schedule management system for building construction projects using the earned value management theory. Journal of Cleaner Production, 103(2015), 275-285. https://doi.org/10.1016/j.jclepro.2014.05.031

[46] Cardoso, J., & Sheth, A. The Semantic Web and its applications. In Semantic Web Services, Processes and Applications (pp. 3-33). (2006),Springer, Boston, MA. https://doi.org/10.1007/978-0-387-34685-4_1

[47] Guarino, N. Semantic matching: Formal ontological distinctions for information organization, extraction, and integration. In: Pazienza, M.T. (eds) Information Extraction A Multidisciplinary Approach to an Emerging Information Technology. SCIE 1997. Lecture Notes in Computer Science, vol 1299. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-63438-X_8

[48] Bali, A., Al-Osta, M., Abdelouahed, G. An Ontology-Based Approach for IoT Data Processing Using Semantic Rules. In: Csöndes, T., Kovács, G., Réthy, G. (eds) SDL 2017: Model-Driven Engineering for Future Internet. SDL 2017. Lecture Notes in Computer Science(), vol 10567. Springer, Cham. https://doi.org/10.1007/978-3-319-68015-6_5

[49] Wang, M. (2021). Ontology-based modelling of lifecycle underground utility information to support operation and maintenance. Automation in Construction, 132(2021), 103933.

https://doi.org/10.1016/j.autcon.2021.103933
[50] Niknam, M., & Karshenas, S. A shared ontology approach to semantic representation of BIM data. Automation in Construction(2017), 80, 22-36. https://doi.org/10.1016/j.autcon.2017.03.013
[51] Mignard, C., & Nicolle, C.Merging BIM and GIS using ontologies application to urban facility management in ACTIVe3D. Computers in Industry, 65(9)(2014), 1276-1290. https://doi.org/10.1016/j.compind.2014.07.008
[52] Tang, S., Shelden, D. R., Eastman, C. M., Pishdad-Bozorgi, P., & Gao, X. A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends. Automation in Construction, 101(2019), 127-139. https://doi.org/10.1016/j.autcon.2019.01.020
[53] Lee, J., & Jeong, Y. User-centric knowledge representations based on ontology for AEC design collaboration. Computer-aided design, 44(8)(2012), 735-748. https://doi.org/10.1016/j.cad.2012.03.011
[54] Zhang, S., Boukamp, F., & Teizer, J. Ontology-based semantic modeling of construction safety knowledge: Towards automated safety planning for job hazard analysis (JHA). Automation in Construction, 52(2015), 29-41. https://doi.org/10.1016/j.autcon.2015.02.00
[55] Pauwels, P., Zhang, S., & Lee, Y. C. Semantic web technologies in AEC industry: A literature overview. Automation in construction, 73(2017), 145-165. https://doi.org/10.1016/j.autcon.2016.10.003
[56] Gilani, S., Quinn, C., & McArthur, J. J. (2020). A review of ontologies within the domain of smart and ongoing commissioning. Building and Environment, 182, 107099. https://doi.org/10.1016/j.buildenv.2020.107099
[57] Zhong, B., Gan, C., Luo, H., & **ng, X. (2018). Ontology-based framework for building environmental monitoring and compliance checking under BIM environment. Building and Environment, 141, 127-142.https://doi.org/10.1016/j.buildenv.2018.05.046
[58] Hong, T., D'Oca, S., Turner, W. J., & Taylor-Lange, S. C. (2015). An ontology to represent energy-related occupant behavior in buildings. Part I: Introduction to the DNAs framework. Building and Environment, 92, 764-777. https://doi.org/10.1016/j.buildenv.2015.02.019
[59] Hong, T., D'Oca, S., Taylor-Lange, S. C., Turner, W. J., Chen, Y., & Corgnati, S. P. (2015). An ontology to represent energy-related occupant behavior in buildings. Part II: Implementation of the DNAS framework using an XML schema. Building and Environment, 94, 196-205.https://doi.org/10.1016/j.buildenv.2015.02.019
[60] Li, T., Zhao, Y., Zhang, C., Zhou, K., & Zhang, X. (2022). A semantic model-based fault detection approach for building energy systems. Building and Environment, 207, 108548. https://doi.org/10.1016/j.buildenv.2021.108548
[61] Dumbill, E. (2002). “Support online communities with foaf: How the friend-of-a-friend vocabulary addresses issues of accountability and privacy. IBM’s XML Watch, http://www-106. ibm. com/developerworks/xml/library/x-foaf2. html.

[62] Weather Ontology. https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/WeatherOntology.owl, Accessed 12th Feb 2017

[63] Energy and Resource Ontology. https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/EnergyResourceOntology.owl, Accessed 14th Mar 2017

[64] Mathematical Modelling Ontology (MAMO). https://bioportal.bioontology.org/ontologies/MAMO/?p=summary, Accessed 12th Feb 2017

[65] Li, Y., García-Castro, R., Mihindukulasooriya, N., O'Donnell, J., & Vega-Sánchez, S. Enhancing energy management at district and building levels via an EM-KPI ontology. Automation in Construction, 99 (2019), 152-167. https://doi.org/10.1016/j.autcon.2018.12.010

[66] Suresh, P., Joglekar, G., Hsu, S., Akkisetty, P., Hailemariam, L., Jain, A., ... & Venkatasubramanian, V. Onto MODEL: Ontological mathematical modeling knowledge management. In Computer Aided Chemical Engineering,25(2008), 985-990. https://doi.org/10.1016/S1570-7946(08)80170-8

[67] Janowicz, K., Haller, A., Cox, S. J., Le Phuoc, D., & Lefrançois, M. SOSA: A lightweight ontology for sensors, observations, samples, and actuators. Journal of Web Semantics, 56(2019), 1-10. https://doi.org/10.1016/j.websem.2018.06.003

[68] Compton, M., Barnaghi, P., Bermudez, L., Garcia-Castro, R., Corcho, O., Cox, S., ... & Taylor, K. The SSN ontology of the W3C semantic sensor network incubator group. Journal of Web Semantics, 17(2012), 25-32. https://doi.org/10.1016/j.websem.2012.05.003

[69] OWL-Time Ontology http://www.w3.org/2006/time#, Accessed 12th Feb 2017

[70] Ontology of units of measure http://www.wurvoc.org/vocabularies/om-1.8, Accessed 12th Feb 2017

[71] qudt.org. QUDT. http://www.qudt.org/, 2017. Last accessed: 2017-07-22.

[72] Uschold, M., & Gruninger, M. Ontologies: Principles, methods and applications. The knowledge engineering review, 11(2)(1996), 93-136. https://doi.org/10.1017/S0269888900007797

[73] Zhou, Z., Goh, Y. M., & Shen, L. Overview and analysis of ontology studies supporting development of the construction industry. Journal of Computing in Civil Engineering, 30(6)(2016), 04016026. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000594

[74] Noy, N. F., & McGuinness, D. L.Ontology development 101: A guide to creating your first ontology, (2001). https://protege.stanford.edu/publications/ontology_development/ontology101.pdf

[75] Zhao, S. Zhang and J. Zhao, "Research of Using Protégé to Build Ontology," 2012 IEEE/ACIS 11th International Conference on Computer and Information Science, (2012),697-700, https://doi.org/10.1109/ICIS.2012.126.

[76] Zhang, Y. Y., Hu, Z. Z., Lin, J. R., & Zhang, J. P. Linking data model and formula to automate KPI calculation for building performance benchmarking. Energy Reports, 7(2021), 1326-1337.

https://doi.org/10.1016/j.egyr.2021.02.044

[77] Gómez-Pérez, A. . Ontology evaluation. In Handbook on ontologies (pp. 251-273).(2004) Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-92673-3_13

[78] El-Diraby, T. A., Lima, C., & Feis, B. Domain taxonomy for construction concepts: toward a formal ontology for construction knowledge. Journal of computing in civil engineering, 19(4)(2005), 394-406. https://doi.org/10.1061/(ASCE)0887-3801(2005)19:4(394)

[79] El-Gohary, N. M., & El-Diraby, T. E. Domain ontology for processes in infrastructure and construction. Journal of Construction Engineering and Management, 136(7)(2010), 730-744. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000178