Building energy performance assessment using linked data and cross-domain semantic reasoning

Citation for published version (APA):

Document license:
TAVERNE

DOI:
10.1016/j.autcon.2021.103580

Document status and date:
Published: 01/04/2021

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Building energy performance assessment using linked data and cross-domain semantic reasoning

Shushan Hu, Jiale Wang, Cathal Hoare, Yehong Li, Pieter Pauwels, James O’Donnell

Abstract

Cross-domain information is essential for building energy performance assessment. The heterogeneous nature of this information is a major source for inefficient assessments. The semantic web provides a flexible pathway for addressing recognised interoperability issues. However, further implicit knowledge in cross-domain information could provide meaningful solutions for such assessments. This paper aims to develop a conceptual framework that links cross-domain information, infers implicit knowledge, and empowers building managers with insightful assessments. The framework integrates Web Ontology Language (OWL) ontologies, Resource Description Framework (RDF) instances, and a set of predefined rules to infer implicit knowledge, which can satisfy data requirements of performance metrics and enable meaningful performance assessments. Then building managers can identify inefficient building operations and improve energy efficiency while maintaining desired building functions. This approach reduces burdensome intervention from the managers when compared with traditional solutions. A demonstration highlights the engineering value by evaluating energy performance of a university building.

1. Introduction

Buildings account for a globally significant portion of primary energy consumption (e.g. 40% in U.S.), a figure that is predicted to grow in years to come [1]. Building energy performance assessment plays a key role in narrowing the recognised gap between predicted and measured building energy performance. This assessment analyses the level of the indoor thermal comfort and energy consumption of heating, cooling, ventilation, lighting, hot water, and accounts for the presence of people and domestic appliances. Assessment results provide rich data for Measurement and Verification (M&V) and Monitoring and Targeting (M&T) related to building energy management. M&V quantifies energy savings with minimum uncertainty and explores if the desired performance is achieved. M&T aims to identify inefficient operations and improve energy efficiency by rectifying these exceptions.

The widespread application of M&V and M&T has resulted in significant benefits for building energy management, for example, Gallagher et al. [2] proposed a mixed machine learning model with artificial neural network, support vector machine, and K-Nearest neighbour clustering to calculate the baseline energy consumption and enable accurate quantification of energy savings. Sun et al. [3] summarised that data-driven approaches for building energy prediction overcome limitations of physical models, due to their ability to discover statistical patterns from a large amount of available data without expert knowledge. The Construction Operation Building Information Exchange (COBie) specification incrementally gathers and systematically manages relevant information among building stakeholders as they engage in a project to enable better facility management and building performance throughout the building life cycle [4].

Building energy performance assessment is a complex activity because of the diversity and individualism of buildings, as well as the diverse goals of stakeholders. The heterogeneous and isolated nature of available information is one of the key factors attributed to inefficient assessments of building energy performance [5]; this is the domain scope for this paper. Effective cross-domain information sharing offers exciting opportunities for analysing the causes of the well-recognised...
Recent years have seen an increase in studies that highlight the potential for linking and integrating cross-domain building information using semantic web technologies [18]. For example, ifcOWL describes sensors as elements of a system deployed in a building with specific measuring capability [13]. Based on the work by Rasmussen et al. [14], the Linked Building Data Community Group (LBD) defined the Building Topology Ontology (BOT) that allows expression of the building topology for any building. Furthermore, Curry et al. [15] developed a linked dataspace to create a well-connected graph of relevant information for the purpose of building management.

Such individual and integrated contextual data silos contain rich information about the energy performance associated with delivering correctly functioning spaces [16]. There is also considerable implicit knowledge contained within these datasets that has the potential to play a key role in the energy performance assessment of buildings. For example, a deduced relationship hasGeoLocation enables linking IfcSpace instances in ifcOWL and instances of Sensor in SSN. Alternatively, more specific links can be detailed between related datasets, such as spaces and sensors, using dedicated properties like hasSensor. Schneider et al. published a study including a smart home dataset illustrating how smaller datasets representing one building can be combined into one data graph that contains sensor data according to SSN [16]. The resulting linked dataset is a solid basis for cross-domain semantic reasoning processes.

Building on available datasets, semantic reasoning offers enormous potential to infer implicit knowledge from cross-domain building information [17]. This technique has the ability to derive logical inferences that are not explicitly expressed in Resource Description Framework (RDF) instances or in OWL ontologies. Many researchers applied semantic reasoning to improve the data interoperability issue in the Architecture, Engineering, Construction (AEC) domain, for example, Pauwels et al. implemented a semantic rule-checking process for design and construction data by defining a novel set of rules [18] and Zhong et al. proposed a reasoning framework to support building environmental monitoring and compliance checking with data from Building Information Modelling (BIM) tools [19]. These semantic reasoning based approaches currently focus on BIM-sourced information (including IFC). However, architecture and equipment data in a BIM only provide part of the information needed for a comprehensive analysis of building energy performance. Cross-domain information such as meteorological, occupancy, and sensing data has not yet contributed to semantic reasoning.

This paper aims to develop a conceptual framework that links cross-domain building data, infers hidden links between instances, and empowers building managers with insightful energy performance assessment information. The framework first proposes to leverage the RDF standard to construct a linked dataspace that presents cross-domain building data in a semantic and homogeneous manner. The approach then develops an automated and efficient pathway to infer implicit knowledge and hidden links from the dataspace and construct an enriched knowledge base that can satisfy datum stream requirements of key performance metrics. Finally, these hidden links are used to evaluate performance metrics and provide building managers with comprehensive energy performance assessment. Compared to traditional solutions the automated reasoning approach could dramatically reduce intervention and interpretation from building managers.

This research design an assessment solution for building energy performance through cross-domain semantic reasoning upon information in a linked dataspace and leverages a significant body of previous research to do so [18, 20–22]. The solution uses three sources of information (i.e. RDF instances, OWL ontologies, and inference rules) to establish a knowledge base that supports data requirements of metrics for assessment of building energy performance. The definition of a set of inference rules (e.g. IF-THEN statements) infers sensor, weather and occupancy information for evaluating performance metrics. A demonstration case validates the effectiveness of the approach with energy performance assessment for a school building in Hubei University, China.

The remainder of this paper is structured as follows: Section 2 presents the case for addressing the data interoperability issue in the AEC domain with linked data and semantic reasoning. The framework that uses cross-domain semantic reasoning to assess energy performance of buildings is illustrated in Section 3. Section 4 lists a set of inference rules for performance-related building information. Section 5 provides a conceptual demonstration of this work and Section 6 discuss engineering findings arising from the demonstration. Section 7 reflects on the broader application potential of the approach.

2. Using semantic web technologies to improve data interoperability within the AEC domain

By virtue of advanced Information and Communication Technology (ICT), a wider range of building information, than was previously possible, can be accessed across the entire building life cycle [23]. These data open ample opportunities for building energy improvements through field-oriented computation such as energy benchmarking and energy performance evaluation [24]. Yet, the limited interoperability among diverse data silos in modern buildings significantly contributes to the current difficulties associated with energy management activities [25].

A quite successful solution to the data interoperability issue is to link cross-domain data sources through semantic web, which uses a subject-object-predicate structure to describe the relationships between specific objects. In order to construct a linked dataspace with contextually represented cross-domain building information, previous research has defined some OWL ontologies (e.g. ifcOWL, SSN) for specific domains of information. Indirect links between RDF instances in different domains are essential for building energy performance assessment but require significant efforts from engineering experts. Semantic reasoning offers a feasible pathway to obtain these indirect links. However, current reasoning based approaches mainly focus on architecture information originating from BIM-based formats and do not involve other domains of information (e.g. sensor and weather information) that is the precondition for evaluating key performance metrics.
2.1. Linking cross-domain building information with RDF

Assessment of energy performance in buildings requires information from a wide range of data silos, most of which are unstructured and heterogeneous in nature [26]. The RDF standard in semantic web provides a practicable pathway to link cross-domain data sources. Such RDF-based approaches use OWL ontologies to accomplish homogeneous descriptions of building information. For example, the BuildingSMART linked data working group defined the ifcOWL ontology7 that can serve as a domain ontology for BIM information [11,12]. ifcOWL provides an alternative representation of the EXPRESS schema of IFC models. A free IFC-to-RDF conversion service10 converts IFC files into RDF data following the ifcOWL ontology.

The W3C LBD group proposed a simple BOT ontology that only covers the core geometrical concepts of a building (i.e. Site - Building - Storey - Zone - Space - Element). This group is in the process of developing three other ontologies (i.e. a PRODUCT ontology, a GEM ontology, and a PROPS ontology), which aim to represent a classification of building elements, geometric representations, and building-related properties respectively. When combined, these four ontologies align with W3C recommendations for adjacent domains, e.g. the geo-spatial ontologies [27], Smart Appliances REference11 (SAREF) [38], Domotics intelligent device12 (DogOnt) [29], and SSN [13]. Also, a Building Automation and Control Systems (BACS) ontology now supports monitoring and controlling functions to improve building operations in terms of energy efficiency and indoor comfort [30].

The deployment of various sensors within a building facilitates continuous measurement of building energy performance [31]. The characteristics of these sensors and their observations form part of an integrated cross-domain data set. In order to enable semantic representations of sensor descriptions and observations, the W3C SSN-XG group developed the SSN ontology with an associated RDF data model [13]. This ontology enables advanced access and formal analysis of sensor information.

Occupant behaviour is a pivotal consideration during building design, operation, management and retrofit, and is a crucial factor when assessing building energy performance [32]. Hong et al. defined an ontology to represent energy-related occupant behaviour in buildings [33]. This ontology parses occupant behaviour into four key components: drivers of behaviour, needs of occupants, actions of occupants, and systems acted on by occupants. This approach facilitates occupant information modelling through improved interoperability between occupant behaviour models and building energy modelling.

Mario et al. defined the ThinkHome ontology as part of an energy efficient smart home system. This ontology consists of concepts related to building information, appliances, indoor thermal comfort, and external weather [34]. Abanda et al. proposed the PV-TONS ontology for photovoltaic (PV) devices in buildings where the intended use of the technology and the ontology is to size and select PV system components for different buildings [35].

In order to enable an automated and easy-to-use process for evaluating building energy performance, Corry et al. defined a performance assessment ontology to analyse building energy performance [36]. This ontology connects existing RDF data silos including ifcOWL, and SSN instances. Hu et al. presented an ontology describing rule-based performance metrics in RDF that in turn provides a flexible and automated mechanism for building energy performance assessment [21]. The design of a hybrid storage structure links time-series related to building energy performance and contextual data silos in a linked dataspace [37].

Together, existing OWL ontologies form an integrated knowledge graph (linked dataspace) for building energy performance assessment (Fig. 1). The knowledge graph provides the underpinning data requirements for performance metrics using relationships between concepts from diverse data silos and domain-oriented relationships in each individual data silo. However, many relationships between RDF instances in the graph are presented in an implicit and indirect way, for example, the link between a ‘Metric’ instance and a ‘Sensor’ instance (Fig. 1). These indirect links are important for the evaluation of building energy performance but are manually inferred, which is a burdensome and error-prone process.

2.2. Semantic reasoning in ABC

Semantic reasoning expands directly defined characteristics of RDF data and derives implicit assertions that are more difficult to obtain through the Simple Protocol and RDF Query Language (SPARQL) [17]. Predefined semantic inference rules typically use first order predicate Logic or description logic to make conclusions from a sequence of statements. Additionally, it is feasible to define new IF-THEN rules using the Semantic Web Rule Language (SWRL), as an example, to enable semantic reasoning within the first order logic [38].

Some approaches use semantic reasoning to address particular issues (e.g. compliance checking, cost estimating, error detection) in the AEC field [22]. Pauwels et al. introduced a process of acoustic regulation compliance checking for BIM models by defining N3Logic rules in combination with a domain ontology and an instance model in order to check whether a building model is compliant with European acoustic regulations [39]. Pauwels et al. also presented a benchmark of three inference rule checking approaches implemented in SPIN, EYE, and Stardog [18]. This paper obtains a performance benchmark for these approaches and gives an indication of main features and decisions.

Abanda et al. explored the development of an ontology based on new rules of measurement for building cost estimation during the tendering stage [40]. This methodology adopts descriptive logic-based reasoners to syntactically validate the proposed ontology. Lee et al. stated an ontological inference process for building cost estimation [41]. This solution defines two OWL ontologies, a work item ontology and a work condition ontology, to perform the estimations.

Wicaksono et al. proposed an ontology based information model for building energy management that offers an expressive representation and reasoning capability [42]. The model integrates a set of SWRL rules to detect occurrences of anomalous activities in building operations. Zhang et al. further developed another approach for health and safety measures in buildings [43]. The approach combines a number of ontologies and SWRL rules to analyse construction projects including jobs, tasks, and resources required to enable safe execution of projects.

When viewed collectively, previous approaches enable a paradigm shift in the way that information in AEC can be automatically inferred for specific requirements. However, semantic reasoning for building energy performance assessment is still in development due to its requirements for cross-domain information from sources. The integration of this information demands cross-domain semantic reasoning to process indirect and implicit relationships between RDF instances in multiple data silos.

3. Performance assessment through linked data and semantic reasoning

This research proposes a cross-domain semantic reasoning solution that aims to empower building managers when assessing building energy performance (Fig. 2). The approach infers implicit knowledge from linked dataspaces to enable comprehensive evaluations of key performance metrics. For example, energy use intensity (EUI) infers energy consumption of a known space by integrating sensor descriptions and observations in SSN and space information in ifcOWL. In addition to available OWL ontologies and RDF instances, we define a set of cross-
domain inference rules to deduce implicit knowledge related to building energy performance. The reasoning process can output an enriched knowledge base that satisfies query requirements related to performance metrics.

Semantic reasoning provides an effective pathway to deduce explicit knowledge by validating implicit and potential links between RDF instances [44]. Reasoning engines generally leverage the description logic principles to obtain extra knowledge from RDF data and OWL ontologies. Additional inference rules can enable reasoning within the first order logic which is in turn used to realise a powerful inferencing process by chaining and integrating rules with the description logic principles.

Aligning with an approach taken by a relevant research project [21], the approach creates a framework with three distinct layers (see Fig. 2) that clearly distinguish among: 1) support services for building energy performance assessment, 2) linked data that enable semantic reasoning, and 3) cross-domain building data sources. Each layer is described in detail in Sections 3.1–3.3.

3.1. Support services layer

The support services layer provides functional interfaces to end-users. In the context of this paper, the comprehensive assessment of building energy performance evaluates real time and historical periods of building operation. Two query types (i.e. SPARQL and SQL) prepare datum streams for processing into performance metrics [37]: 1) SPARQL queries on contextual building information in a knowledge base, and 2) SQL queries on time-series data stored in a relational database.

3.2. Linked data layer

The linked data layer generates a contextual knowledge base using three key sources: (1) A set of OWL ontologies defining class, data, and object properties that are the main input for the Terminological Box (TBox); (2) Cross-domain RDF instances assisting the Assertion Box (ABox) to assert facts based on the concepts defined in OWL ontologies; (3) A list of inference rules (i.e. IF-THEN statements) being another source for the TBox to infer implicit links between RDF instances. A semantic reasoning engine parses the three information sources to generate an enriched knowledge base.

As demonstrated in the context of a SSN graph, relationships among classes, properties, and instances contain implicit knowledge hidden in TBox and ABox (Fig. 3). The domain of ssn:observes should be an instance of ssn:Sensor and the range is an instance of ssn:Property. A transitive relationship (i.e. rdfs:subPropertyOf) hints that all resources related by one property are also related to another. The owl:inverseOf relationship defines an inverse link between properties such as the relationship between ssn:observes and ssn:observedBy.

The rdfs:subClassOf property states that one class (i.e. child) is a subclass of another class (i.e. parent) and all instances of the child class are also instances of the parent class. This property has a transitive character (Fig. 3). For example, SN-ABC01 is an instance of both ssn:
CapacitiveThinFilmPolymer and ssn:Sensor. The same link also exists in the two pairs: 1. cur_humidity and ssn:Property, 2. room_air and ssn:FeatureOfInterest. In conclusion, a positive assertion is that the room_air entity has a relative humidity of 52% and the coil entity has a temperature of 10.5°C.

3.3. Sources layer

The sources layer contains cross-domain building data silos related to building energy performance (in Fig. 2). The approach categorises this information into four types:

1. OWL ontologies (e.g., ifcOWL, SSN) that define schemata for information from diverse domains;
2. contextual building data in RDF that is accessible through a set of available adapters/services [12,45];
3. a set of predefined inference rules that infers indirect and hidden links among RDF instances; and
4. a relational database to manage time-series data from monitoring of building operations

4. Cross-domain inference rules

Cross-domain inference rules can extract more implicit knowledge from semantic data silos than the standard reasoning potential from OWL ontologies [44]. For purpose of constructing an enriched knowledge base to evaluate building energy performance, this study defines a set of meaningful inference rules based on available OWL ontologies and specific reasoning purposes (Fig. 4).

4.1. Reasoning purposes and OWL ontologies

The reasoning purposes of inference rules are to enable evaluation of performance metrics by satisfying their data requirements. Performance metrics present standardised quantitative descriptions for building energy performance and reflect ambitions from a multitude of stakeholders throughout the main phases of the building life cycle (i.e. design, construction, commissioning, and operation). Stakeholders are entities (i.e. individuals or organisations) with a legitimate interest in a system, e.g. government, occupants, and managers. These stakeholders might have different interests in building energy performance [46], examples of
which include:

- building occupants actually inhabit the buildings and represent a strong interest in a healthy and comfortable indoor environment;
- building managers are responsible for building energy performance and must track energy use of the building to ensure that all energy systems work correctly;
- Clients and developers initiate the design and construction process and are interested in a value proposition that balances investment in building performance with the requirements of the future occupants.

The rules presented in this section focus on reasoning purposes and identify indirect links that provide a meaningful basis for an analysis of building energy performance. These rules require a number of particular OWL ontologies, namely, the ifcOWL, SSN, and occupancy ontologies. The ifcOWL ontology provides an expressive representation of a BIM that contains architecture and mechanical design information from the design to construction phases. The SSN ontology describes sensor deployment and observations during the commissioning and operation phases. The occupancy ontology represents occupancy schedules and occupant behaviour during operation.

Individual data silos first need to be linked together to bridge domains and subsequently enable cross-domain rule inference. Various techniques exist for this purpose. The chosen approach aligns with that of Schneider et al. and links data together in a modular fashion, namely through predefined object properties which have rdfs:domain in one ontology and rdfs:range in the other ontology [16].

### 4.2. Three sets of inference rules

Architectural and equipment information from BIM provides contexts for building performance analysis. Three additional types of information (i.e. sensor description, weather report, occupant behaviour) are essential for building energy performance. This research defines three sets of cross-domain inference rules for the three types of information in order to satisfy reasoning purposes from performance metrics:

1. the first set infers sensor properties;
2. the second set deduces weather reports; and
3. the third set observes occupancy schedules.

These rules work as key components of the TBox and constitute an integrated approach that can deliver an enriched and explicit knowledge graph for assessing building energy performance.

As logical conjunctions of declarative IF-THEN structures, an inference rule consists of an antecedent (i.e. IF-part) and a consequent (i.e. THEN-part), distinguishable by a ⇒ symbol. The IF-part states a complex partial knowledge graph that follows related OWL ontologies and the THEN-part presents inference results as a single property. Both the IF and THEN parts consist of a set of atoms ((atom₁ ∧ ... ∧ atomₙ)) that can take a number of forms including C(x), P(x,y), sameAs(x,y), and builtin(r,x,y), where C indicates an OWL description or data range, P hints at an OWL property, r presents a built-in relation, where x and y are variables.

The authors previously defined rules as a checking environment for building design and construction [18]. The approach taken in this paper reuses these rules to analyse the hierarchical structure of buildings. Listing 1 presents an example that retrieves IfcSpace entities from a building storey. The first 5 lines (i.e. IF-part) show how ifcowl:IfcRel-Aggregates works as a bridge to connect floors and attached spaces in an IfC model. The last line (i.e. THEN-part) shows a simple and direct object property (i.e. bep:hasifcspace) between the ifcBuildingStorey and IfcSpace individuals.

Listing 1: retrieves all IfcSpace entities within a floor.

1 ifcowl:IfcBuildingStorey(?flr)
2 ∧ ifcowl:IfcRel-Aggregates(?rel)
3 ∧ ifcowl:relatingObject_IfcRel- Decomposes(?rel,?flr)
4 ∧ ifcowl:relatedObjects_IfcRel- Decomposes(?rel,?spa)
5 ∧ ifcowl:IfcSpace(?spa)
6 ⇒ bep:hasifcspace(?flr,?spa)

### 4.2.1. Rules inferring sensor properties

The first set of inference rules reason about various sensors for the assessment of building energy performance. The resulting rules query sensors that measure building and system status to satisfy data requirements of metrics such as thermal comfort and energy consumption. The syntactic structure of these rules relies on class and property definitions in the ifcOWL and SSN ontologies.

Listing 2 illustrates an example rule for retrieving temperature sensor instances within an IfcSpace instance. In this case, a bep:hasGeoLocation property links instances of ifc:IfcSpace (i.e. ifcOWL ontology) and ssn:Sensor (i.e. SSN ontology). This link across domains is used within the IF-part of the rule (first 3 lines). The rule furthermore (i.e. lines 4–7) checks two data properties (i.e. ssn:observedProperty and ssn:hasFeatureOfInterest) of sensors to confirm the air temperature sensing property. Finally, this rule outputs the id of confirmed sensors. This id can be used for querying time-series data from relational databases (lines 9), similarly to how it was proposed in Petrova et al. [47]. The THEN-part shows how this rule generates a direct data property (i.e. bep:hasTemperatureSensor) that links the id value of available temperature sensors directly to the space.

Listing 2: A rule retrieving information of temperature sensors deployed in a IfcSpace entity.

1 ifc:IfcSpace(?spa)
2 ∧ ssn:Sensor(?sen)
3 ∧ bep:hasGeoLocation(?sen,?spa)
4 ∧ ssn:observedProperty(?sen,?obpr)
5 ∧ swrlb:equal(?obpr, "electricity"^^xsd:string)
6 ∧ ssn:hasFeatureOfInterest(?sen,?foi)
7 ∧ swrlb:equal(?foi,"air"^^xsd:string)
8 ∧ rdfs:label(?sen,?senid)
9 ⇒ bep:hasTemperatureSensor(?spa, ?senid)

Another rule in Listing 3 aims to discover sensors that measure the electricity consumption of ductless Heating, Ventilation and Air Conditioning (HVAC) devices and link their ids directly to the space node. This rule (i.e. lines 1–5) applies an indirect mechanism to identify links between IfcSpace instance and sensor instances. The name of an IfcSpace instance is used to check whether the ssn:hasFeatureOfInterest property of a sensor contains the IfcSpace name. In order to confirm the sensing property of HVAC electricity consumption, the rule (i.e. lines 6–9) furthermore checks whether the ssn:hasFeatureOfInterest property contains ‘HVAC’ and the data property of the ssn:observedProperty is equal to ‘electricity’.

Listing 3: A rule observing sensors related electricity consumption of HVAC devices.

1 ifc:IfcSpace(?spa)
2 ∧ rdfs:label(?spa,?span)
3 ∧ ssn:Sensor(?sen)
4 ∧ ssn:hasFeatureOfInterest(?sen,?foi)
5 ∧ swrlb:contains(?foi,?span)
6 ∧ swrlb:contains(?foi, "HVAC"^^xsd:string)
7 ∧ ssn:observedProperty(?sen,?obpr)
8 ∧ swrlb:equal(?obpr, "electricity"^^xsd:string)
9 ∧ rdfs:label(?sen,?senid)
10 ⇒ bep:hasHVACEnergySensor(?spa, ?senid)

### 4.2.2. Rules deducing weather reports

The second set of rules deduces weather reports generated by

1 ifc:IfcSpace(?spa)
2 ∧ rdf:label(?spa,?span)
3 ∧ ssn:Sensor(?sen)
4 ∧ ssn:observedProperty(?sen,?obpr)
5 ∧ swrlb:contains(?obpr, "weather"^^xsd:string)
6 ∧ swrlb:contains(?obpr, "temperature"^^xsd:string)
7 ∧ swrlb:contains(?obpr, "air"^^xsd:string)
8 ∧ rdfs:label(?sen,?senid)
9 ⇒ bep:hasWeatherSensor(?spa, ?senid)
meteorological equipment in weather stations. Weather reports consist of outdoor environmental parameters such as air temperature, humidity, and wind speed. These parameters are essential for evaluating performance metrics (e.g., indoor thermal comfort, energy consumption of lights and HVAC system). In this case, these rules rely on the ifcOWL and weather ontology (ThinkHome).\(^{13}\)

Listing 4 shows a rule that retrieves pertinent weather information for a building over a specific time period. The rule relies on an indirect relationship between building instances in an IFC model and weather reports in a geospatial data silo (i.e., lines 1–9). Such a link needs to be available in the data (data linking phase) to make this rule function. Furthermore, this rule obtains a city name?town through two ifcOWL object properties (i.e., ifcowl:buildingAddress?ifcBuilding and ifcowl:town?IfcPostalAddress) (lines 2–5), and checks whether that name is equal to the name of a considered geospatial point (?weaaddr - line 9).

Finally, the weather state for this location is retrieved, if available (line 10–11) in the same rule. The output of the rule shows how a simple property (i.e., bep:hasWeatherState) can be inferred to directly link a building and an associated weather state.

Listing 4: A rule inferring outside weather information of a building.

\[
\begin{align*}
1 & \text{ifcowl:IfcBuilding(?blg)} \\
2 & \text{\land ifcowl:buildingAddress?ifcBuilding(?blg, ?addr)} \\
3 & \text{\land ifcowl:town?IfcPostalAddress(?addr, ?town)} \\
4 & \text{\land expchasString(?town, ?tona)} \\
5 & \text{\land vgs84:pos:Point(?weaaddr)} \\
6 & \text{\land rdfs:label(?weaaddr, ?weadna)} \\
7 & \text{\land swrlb:equal(?tona, ?weadna)} \\
8 & \text{\land vgs84:pos:location(?weacur, ?weaaddr)} \\
9 & \text{\land weaOn:WeatherState(?weacur)} \\
10 & \Rightarrow \text{bep:hasWeatherState(?blg, ?weacur)}
\end{align*}
\]

4.2.3. Rules observing occupancy schedule

The third set of rules observes the occupancy schedule to support building energy performance assessment. Occupant behaviour can dramatically affect overall building energy consumption, as much as doubling or tripling predicted energy consumption \cite{48}. The occupancy schedule is therefore one of the key factors for building energy performance simulation and system control strategies. This rule set relies on the ifcOWL and iCalendar ontologies that represent people’s calendars and schedule information through events and to-dos \cite{49}.

Listing 5 illustrates a rule that deduces attendees of an event held in an IfcSpace instance. Within the IF-part of the rule, the direct link (i.e., bep:hasEveLocation) between instances of ifc:IfcSpace (i.e., ifcOWL ontology) and iCal:Vevent (i.e., iCalendar ontology) allows retrieval of relevant spaces and occurring events (lines 1–3). The rule then selects events occurring in the location during the time period in question and represents attendees of these events as outputs of the rule (lines 4–8). The THEN-part of this inference rule shows a straightforward object property (i.e., bep:hasOccupant), through which the rule exports occupancy information of an IfcSpace instance.

Listing 5: A rule retrieving occupancy information in an IfcSpace entity.

\[
\begin{align*}
1 & \text{ifc:IfcSpace(?spa)} \\
2 & \text{\land iCal:Vevent(?eve)} \\
3 & \text{\land bep:hasEveLocation(?eve, ?spa)} \\
4 & \text{\land iCal:dtStart(?eve, ?esta)} \\
5 & \text{\land iCal:dtEnd(?eve, ?eend)} \\
6 & \text{\land time:before(?esta, ?cur)} \\
7 & \text{\land time:after(?eend, ?cur)} \\
8 & \text{\land iCal:attendee(?eve, ?occ)} \\
9 & \Rightarrow \text{bep:hasOccupant(?spa, ?occ)}
\end{align*}
\]

5. Testing performance assessment with cross-domain reasoning

This study now documents results in terms of the effectiveness and engineering value of the approach in the context of a practical application. The case study assesses energy performance of a school building in Hubei University, China. This facility is a six-storey complex spreading over 8000 m\(^2\). The building contains approximately 100 spaces that are categorised into four types: classrooms, washing rooms, corridors and staircases. There are two types of classrooms to accommodate different class sizes. The larger classrooms contain about 100 seats in an area of around 100 m\(^2\) and the smaller classrooms have 74 seats within an area of 70 m\(^2\). Each classroom has an independent air conditioning device.

The case study has significant potential for energy performance enhancement. The building is a relatively complex facility containing tens of big zones with own energy consuming devices and serving a diverse range of classes over a long teaching day. For example, appropriate teaching schedules for classrooms can maximise the number of occupants within classrooms during the teaching periods and take advantage of the residual energy from previous lectures in the same zones as a pre-heating/pre-cooling process. Due to the complexity of the building and the interoperability issue among cross-domain data silos, the building managers have difficulty assessing the overall energy performance of the building, its systems and components.

This case study implements an executable wrapper for the theoretical approach presented in this paper in a way that facilitates continuous analysis of building energy performance. The demonstration presents assessment results to the managers through an intuitive dashboard. The developed demonstration comprises five components:

1. Preparation of cross-domain information related to building energy performance, namely, a Building Energy Performance Simulation (BEPS) model and performance metrics organised by a method called scenario modelling (Section 5.1).
2. A software framework which implements the newly defined cross-domain semantic reasoning approach based on Stardog (Section 5.2).
3. Assessment results of building energy performance as evaluated through a BEPS of the building (Section 5.3).
4. System performance metrics that give an indication of the sizes of the data, the query response and overall performance times of the framework (Section 5.4).

5.1. Preparation of cross-domain information

In order to demonstrate the applicability of the approach, this case study prepares six types of RDF information and corresponding OWL ontologies related to building energy performance. All of this data supports scenario-based building energy performance assessment.

The considered cross-domain information includes:

- BIM data according to the ifcOWL ontology, obtained by transforming an IFC file to an RDF syntax using the online IFC to RDF service.
- Sensor information according to the SSN ontology, which provides the meta-data for describing the deployment of the sensors and the output of the BEPS model \cite{50}.
- Weather information collected from an on-line system (i.e., weather underground) and encapsulated with the core of the ThinkHome ontology \cite{51}.
- Occupancy information according to the iCalendar ontology, which represents the lecture timetable \cite{49}.
- Performance metric data and contextual information about the building, according to a performance metric ontology \cite{21}.
- Time-series data from deployed meters and the energy simulation results that are managed by a relational database (i.e., MySQL).

\[^{13}\text{https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/}\]
The case study prepares a fusion of measured and simulated building performance data for evaluation of performance metrics. Some necessary sensors are absent in the building including those for space temperature, relative humidity, meters for energy consumption of space equipment. In order to address this issue, the demonstration constructs a BEPS model (Fig. 5) to generate sufficient time-series data that supplements the measured data. As shown in Fig. 6, four datum streams come from real measurements including outside air temperature, building energy consumption and area, and the number of occupants.

Besides the architectural design (Fig. 5), the BEPS model contains extensive equipment to provide an indoor environment appropriate for teaching and learning. Each classroom has a computer, a projector, and more than ten fluorescent tubes (i.e. T5). Ductless mini-split air conditioners satisfy the heating and cooling requirements of each classroom. Such modular systems have the advantage of a small form factor and offer highly flexible control over a range of spaces. These systems consist of two main components: an outdoor compressor/condenser and an indoor air-handling unit.

This study applies EnergyPlus as the simulation engine for the BEPS model. EnergyPlus can generate precise whole building energy simulation results through a modular system integrated with a heat balance-based zone simulation with time-steps of less than an hour. The system enables accurate calculation of radiant cooling and heating loads, heat exchange between zones, moisture adsorption and desorption in elements of the facility, and interzone airflow. EnergyPlus is a commonly used tool among researchers and industry professionals alike.

The case study executed the BEPS model of the classroom building over the entire year of 2018, covering two teaching semesters. In addition to key inputs such as the weather data, building geometry, building cross sections, material properties and HVAC systems, the model uses the lecture timetable for the building to represent the occupancy profile and operational schedule for the lighting systems, teaching equipment and HVAC devices. The requested output of the BEPS contains:

1. zone environmental parameters (e.g. temperature (°C), and occupancy status (People)) with a time interval of 10 min,
2. energy consumption of devices (e.g. air conditioner, light, indoor equipment) with a time interval of 1 h.

5.1.1. Scenario modelling

This case study applies a method called ‘scenario modelling’ to organise key performance metrics for insightful and multi-granular assessment of building energy performance. O’Donnell et al. [52] designed scenario modelling to enable the explicit and unambiguous coupling of building function with other pivotal aspects of building operation. The method takes education and technical expertise of building managers into consideration and provides a flexible solution for the assessment of building energy performance.

This demonstration defines a scenario model (Fig. 6) to reflect some key concerns in the area of thermal comfort and energy consumption in the building by examining two performance aspects in parallel (i.e. building function and energy consumption). Building function, defined as a controlled environment for a given purpose, is of paramount importance for most organisations and is therefore a priority for building managers. The energy consumption performance aspect focuses on the quantification of energy consumed by equipment deployed in buildings and the identification of unnecessary consumption. The two performance aspects use eight performance metrics to assess energy flows within the building.

A key purpose of the case study is to verify the thermal comfort status of occupants indicating satisfaction about the indoor environment. An appropriate measurement of thermal comfort is essential for energy performance improvement through adaptively controlling the setpoint temperature of HVAC and natural ventilation systems [53]. Fanger et al. defined the PMV model to predict the mean comfort status for many occupants with six parameters. PMV falls short when considering thermal sensation and the preference of individuals with respect to environmental factors such as dry-bulb air temperature and relative humidity, the physiological factors such as skin temperature and heart rate, and behavioral factors such as activity and clothing levels [54]. Personalized Comfort Models (PCMs) predict thermal comfort of individuals by incorporating human thermo-regulation factors [55]. PCMs offer the potential for an improved interpretation of comfort status. The evaluation of a PCM requires some physiological signals including skin temperature and heart rate from wearable sensors which are absent from the case building. Against this backdrop, this study employs adaptive PMV (aPMV) to predict the status of the thermal condition [56]. aPMV enhances the PMV model with an adaptive coefficient to reflect a human’s adaptive functions including behavioral and psychological adaptation.

For the energy consumption performance aspect, the scenario model firstly aims to identify the principal parts of energy consumption in the building and applies three functional metrics to sum the energy consumption of three types of equipment in all zones, namely air conditioners, lighting systems, and teaching equipment (i.e. computer and projector). The case study then tries to verify energy efficiency of the building by using two objectives to calculate energy use intensity including energy use per area unit (kWh/m²) and occupancy unit (i.e. kWh/person). The eight objectives, taken together, constitute a comprehensive model for energy performance assessment of the building. When the model is evaluated, a clear picture emerges of the thermal conditions in the spaces and how energy consumption of equipment relates to design intent.

5.2. Software framework for cross-domain semantic reasoning

The test case requires a prototype software tool to evaluate building energy performance based on cross-domain semantic reasoning (see Fig. 7). This case study applies Stardog 6.0\textsuperscript{14} as the software component to enable predefined semantic inference rules for building energy performance assessment. With three input sources (i.e. RDF instances in ABox, OWL ontologies and SWRL inference rules in TBox), Stardog performs reasoning in a lazy and late-binding fashion which allows reasoning at query time (i.e. reasoning engine). A large-scale triplestore

\textsuperscript{14} https://www.stardog.com/docs/6.0.0/
provides an efficient solution to manage RDF instances for a knowledge base. The framework transforms the information requirements of performance metrics into SPARQL queries and leverages the query runner to execute these queries.

There are two mechanisms to define inference rules: the first is a forward chaining process that starts with the available data and uses inference rules to extract more information iteratively until no more new facts are generated; the second is a backward chaining process that starts with a list of goals and works in a backwards fashion with inference rules related to queries. Based on the comparison and analysis of these two approaches in [18], this study applied the backward chaining process on rules as defined in Section 4 (shown in Listing 6).


```r
1 [ rdf:type swrl:Imp;
2 swrl:head [ rdf:type swrl:AtomList;
3 rdf:rest rdf:nil;
4 rdf:first
5 [ rdf:type swrl:IndividualPropertyAtom;
6 swrl:propertyPredicate bep:hasoccupant;
7 swrl:argument1 <urn:swrl#spa>;
8 swrl:argument2 <urn:swrl#occ>];
9 swrl:body [ rdf:type swrl:AtomList;
10 rdf:rest [ rdf:type swrl:AtomList;
11... ]
12 rdf:first
```

This framework uses a hybrid storage architecture previously developed by the authors in order to improve data query efficiency [37]. Continuous monitoring of building operations can generate a huge volume of time-series records. For example, the building model in this demonstration generates nearly 200 million simulation records when detailed output is requested over a period of one year. The storage architecture leverages a triplestore to manage semantically-described building contextual data and retains time-series performance records in relational databases (corresponding also to [47]). The software uses sensor information from SPARQL queries to automatically generate SQL queries, which include relevant time periods and correspond to the relational database schema. The output of these SQL queries provides detailed datum streams for evaluating key performance metrics.

5.3. Performance assessment results

The approach uses the prototype software tool to evaluate the scenario model instance. The output of the instance illustrates the thermal condition and the energy consumption of related devices in the building (Fig. 8, Fig. 9, and Fig. 10). These results represent an appropriate human-readable format for building energy performance assessment.
A closer look at results of the building function aspect, which consists of three performance metrics, indicates that the thermal condition defined by dry bulb temperature and aPMV is within acceptable limits during three typical teaching days (see Fig. 8). The indoor temperature of classroom 102 is maintained according to its teaching schedule. The aPMV of classroom 102 also locates in a comfort range (around 0) during the occupied time. Regular fluctuations exist in the figure due to scheduled HVAC operations. The indoor temperature and aPMV decrease quite quickly after the HVAC system shuts down, reaching a value close to the outside temperature (between 5 and 18 °C) but this occurs outside of occupied hours.

The evaluation of the first three performance metrics related to energy consumption quantifies the energy use of lights, teaching equipment and HVAC devices (shown in Fig. 9). HVAC devices are the dominant energy consumer (using from 700 to 1000 kWh per day) as opposed to the other two end use categories. With a stable daily lecture schedule in the building, daily energy consumption for lights and teaching equipment remains constant (70 and 180 kWh per day respectively), while a dramatic variation occurs in the daily HVAC energy consumption which inversely correlates with outdoor temperature values.

The last two performance metrics analyse energy use intensity of the building (Fig. 10). The daily energy use per unit area (i.e. kWh/(m² * day)) ranges between reasonable values (between 0.15 and 0.2 kWh/(m² * day)) for a large portion of the building. The demonstration also computes the daily energy usage per occupant, which illustrates the number of people served by the consumed energy. Due to the variation of students attending lectures in classrooms, it is hard to precisely count the number of occupants in the building. The case study therefore sums the occupants for each lecture as the daily occupancy total, which is a high value (around 25,000 People) in Fig. 10. The high occupancy density leads to a low energy use intensity per person in the building (between 0.028 and 0.038 kWh/person), which illustrates that class-rooms can provide more cost-effective delivery of comfort when
5.4. System performance

This section now documents the efficiency of the software tool. A laptop works as the platform for the test environment and has the following specifications: Intel Core CPU i7-6700HQ at 2.6 GHz 4 cores, 16 GB DDR4 RAM memory, 128 GB solid state drive, Ubuntu 18 × 64 operating system, Stardog 6.0, and MySQL 8.0. OWL ontologies used in the tool include nearly 2,100 classes and 2,300 object properties, for which ifcOWL contributes 1,300 classes and 1,600 object properties. The test RDF files contain around 1 million instances with around 700,000 instances within the IFC model; both weather and iCalendar information have around 130,000 instances. Simulation results from EnergyPlus include around 122 million records relating to building energy performance.

The performance evaluation of the tool includes an assessment of pre-processing time and query execution time. Pre-processing takes 10.9 s and includes loading of OWL ontologies and RDF instances into Stardog. The query contains the execution of SPARQL query on Stardog and SQL query on MySQL. SPARQL queries of the first seven performance metrics return on average around 100 instances within 0.05 s. The queries for the last performance metric (i.e. Energy use per person) export 4100 instances within 0.3 s. Regarding the SQL operations, queries for the first to third performance metrics return around 1,800 records within 0.01 s. Queries for the fourth to sixth performance metrics output around 21,000 records within 0.1 s. Queries for the seventh and eight performance metrics provide around 63,000 records within 0.7 s.

6. Discussion

A number of engineering findings arise from the case study. First, the semantic reasoning method provides a feasible and semi-automated pathway to link cross-domain data and infer implicit knowledge for the evaluation of key energy performance metrics. The building managers currently have difficulty evaluating these performance metrics using time-series building performance data (measured or the output from BEPS models) as stored in CSV files or relational databases. Due to low levels of interoperability among these cross-domain data, the managers need to manually map requirements of metrics to the stored performance data and then perform advanced operations in Excel to filter data in CSV files or design executable SQL queries to process the contents of a relational database. This manual process requires that the managers possess some computer skills but the process remains time consuming, non-reusable, and error prone. The proposed methodology allows further extension of the framework with other OWL ontologies and RDF instances, and inference rules to significantly reduce manual interpretation and intervention from the managers.

A second finding is the proposed solution provides a high degree of flexibility and modularity allowing a wide application for buildings in different climate zones and geographical locations. The RDF standard offers an open and flexible specification to represent cross-domain information in triple instances with a semantic structure (i.e. Subject-Predicate-Object). There are a number of OWL ontologies (e.g. ifcOWL, SSN, and SAREF) available for the description of multiple aspects of building information. Each ontology focuses on one specific type of information and is independent of other ontologies. This design rule for these ontologies results in a high level of modularity. With precise and unambiguous definitions of key performance metrics, existing inference rules can be easily reused for energy performance assessment of many buildings.

Third, the approach can empower building managers to develop insights by assessing historical and real-time building energy performance. Assessment results are shown within an intuitive and continuous dashboard including two types of information: 1) variation of the output of each performance metric in accordance with the time axis and 2) comparison with the outputs of similar performance metrics for the time period in question. Insightful performance assessment data can support the managers during building operation and management and help ensure that buildings operate as expected. For example, these data support M&V to explore whether energy performance of a building achieves agreed values and assists M&T to identify inefficient operations which can then be acted upon for energy performance improvement.

Finally, a number of prerequisites also arise from this approach. For many buildings the evaluation of some important performance metrics is incalculable due to the absence of measured records from specific sensors deployed in buildings. PCM, for example, plays a key role in assessing indoor thermal comfort for improving building energy performance. The evaluation of PCM requires wearable sensors on individuals to obtain measurements such as skin temperature, heart rate, and activity. Installation of these sensors is costly due to the number of occupants in densely populated office spaces. Additional privacy and security concerns may also arise due to tracking of physiological signals. The metabolic rate provides a better understanding of PCM but not directly measurable with existing sensors. An accurate model is required to predict occupants’ metabolic rate with their physiological signals and activity level.

7. Conclusions

Assessment of building energy performance plays a key role in narrowing the performance gap between designed and operational buildings. An interoperability issue among cross-domain information is a major cause of inefficient evaluation of performance metrics from multiple stakeholders. The semantic web enables a paradigm shift by constructing a linked dataset that includes RDF and OWL ontologies. There is considerable implicit and indirect knowledge in the linked dataset that is essential for evaluating building energy performance. Significant technical challenges exist around the inference of this knowledge.

The objective of this research is to develop a conceptual framework that links cross-domain building information, infers implicit knowledge between instances, and empowers building managers with insightful information to support energy performance assessment. In doing so, the framework performs cross-domain semantic reasoning to automatically infer implicit knowledge with three input sources: OWL ontologies, RDF instances, and predefined inference rules. This work defines three sets of inference rules to admit the construction of an enhanced knowledge base by dealing with three types of building information, namely sensor properties, weather reports and occupancy schedules. The use of this implicit knowledge enables comprehensive assessment of building energy performance by satisfying data requirements for performance metrics. A case study highlights the effectiveness and engineering value of the approach using a fusion of measured and simulated performance data for a university building.

With this approach, building managers can conduct better building operation and energy management by accessing comprehensive energy performance assessment in a holistic and detailed manner. The assessment data forms a consolidated basis for M&V from which a calibrated baseline for verifying building energy performance can be established. The managers can perform an M&T process by identifying the causes of inefficient building operations and communicating with other stakeholders to inform better energy-related strategies for energy performance improvement. Another finding is that the approach significantly reduces intervention and reinterpretation by the managers compared with traditional solutions and provides a flexible and modular solution that is applicable for a wide range of buildings. Its inherent flexibility allows further extension of the framework with other OWL ontologies and inference rules related to building energy performance.
Future research is planned to further substantiate the benefits of assessing building energy performance with PCMs, including the deployment of digital wristbands for physiological signals and the development of a deep learning model (e.g., Convolutional Neural Network) for the prediction of the metabolic rate.

Declaration of Competing Interest

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

Acknowledgments

This publication has emanated, in part, from research conducted with the financial support of Science Foundation Ireland under the SFI Strategic Partnership Programme Grant Number SFI/15/SPP/E3125. The opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of Science Foundation Ireland. This work has been partially funded by Natural Science Foundation of Hubei Province, China (Grant No. 2018CFB251).

References


