DOMAIN ANALYSIS TO ASSESS THE SUSTAINABILITY OF CONSTRUCTION PROJECTS USING SEMANTIC TECHNOLOGIES

Cristiano Geraldo Teixeira Silva (1), Marcello Peixoto Bax (2)

(1) FUMEC University, Brazil, cgts.br@gmail.com

(2) Universidade Federal de Minas Gerais, Brazil, bax.ufmg@gmail.com

Abstract

Evaluating construction projects aims to measure how sustainable the built environments are and this is important to save the environment from increasing deterioration. However, most assessment tasks are carried out in different tools, at all stages of construction and even manually, consuming time and increasing the possibility of the errors. The research problem is to verify how the semantic technologies used would represent an advantageous alternative, in relation to the state of the art, to analyze, systematize and organize information from the domain of construction projects, accelerating the evaluation process. The concepts behind an experiment that employs semantic technologies and a domain ontology to integrate data from BIM (Building Information Modeling) projects and other tabular data external to the project to automate the LEED (Leadership in Energy and Environmental Design) certification tasks are presented. To produce the prototype of this experiment, data from a BIM project and open tabulated data from Belo Horizonte's city are semantically annotated and a knowledge graph is generated in RDF format. Once integrated into the graph, the data is used to assess the certification criteria through consultations. The results suggest that the domain analysis associated with semantic technologies promotes the conceptual extensibility of the constructive elements, enabling their integration with external knowledge bases. It also conceptually organizes data aiming to improve relevant information retrieval tasks.

Keywords: Semantic technologies; Ontology; Sustainable constructive projects; LEED.

1 Introduction

The concern with the use of environmental resources in the construction industry's projects has led to the concept of green building, which is increasingly significant for society. Green construction is meant to ensure quality and safety throughout the building's life cycle, corresponding to the sustainable development idea (Xu et al. 2019). To organize knowledge in this domain means integrating data, information, and documents with information systems' support. Hjørland (2017) states that ontologies promote the development of Knowledge Organization Systems (KOS), leveraging domain analysis methods that lead to knowledge representation techniques in Computer Science, and, ultimately, reinforce the disciplinary integration of Information Science.

We hypothesize that organizing knowledge through ontologies allows the conceptually mapping between construction elements and other complementary data and helps project sustainability analysis. In other words, we claim that the built environment's formal representation by ontologies allows project collaborators to obtain information combining data from multiple and potential heterogeneous sources, enriching the analysis. This representation potentially assists certification tasks through data integration, inference, and queries, making it a plausible innovative differentiator in the industry (Niknam and Karshenas 2017; Sacks et al. 2018; Jiang et al. 2018).

LEED certification (Leadership in Energy and Environmental Design), conceived and granted by the USGBC (US Green Building Council) (USGBC 2020), classifies projects into four levels: certified, silver, gold and platinum. According to the USGBC website, this rating system for sustainability is widely used in design, construction, operation and buildings' maintenance. To obtain the green credential advocated by the USGBC, it is essential to evaluate the construction project using LEED criteria. These criteria are organized into seven categories: sustainable spaces, water efficiency, energy and atmosphere, materials and resources, regional priority credit, indoor environmental quality and innovation.

BIM (Building Information Modeling) technology has received increasing attention in the construction industry (Borrmann et al. 2018). The tools which support this technology interoperate using data in the IFC (Industry Foundation Classes) format, as well as exporting tabular data for information analysis. Expressed in BIM an accurate virtual model of a built environment contains geometry and data necessary to support activities throughout the built environment lifecycle. This technology provides the foundation for the design and construction capabilities, promoting changes in the nature of roles and relationships across teams. BIM is presented as an evolution in the design, allowing new possibilities to represent, process, visualize and retrieve knowledge and information from data (Sacks et al. 2018).

Wong and Zhou (2015 p.157) define green BIM as "a model-based process for generating and managing construction data that is coordinated and consistent throughout the project life cycle, which improves the building's energy efficiency performance and it makes it easier to set sustainability goals". We claim that integrating The LEED building certification based on BIM projects potentially rises opportunities for the creation of new tools to certify green buildings, which can improve compatibility issues and ease of use by certification experts. Thus, LEED certification processes, supported by BIM, can potentially save substantial time and effort (Wong and Zhou 2015).

Evaluating a building to certify its sustainability requires smarter methods. Traditional assessment, performed manually, is labor-intensive and error-prone (Jiang et al. 2018). Interoperability is a crucial aspect of facilitating construction industry processes. Stakeholder information comes from a variety of sources and is obtained, generated and transformed using a variety of data sources and software tools (Borrmann et al. 2018). Considering the IFC format semantic limitations, the search for applying semantic technologies aims to meet the best practices for integrating, retrieving information and generating new knowledge (Pauwels et al. 2017).

Domain knowledge is essential to develop a knowledge organization system (Smiraglia 2015). Thus, domain analysis performed in this study involves understanding the epistemological consensus between the BIM classification system and the rules for meeting the criteria of a LEED certification. Therefore, this work is an applied research using domain ontology and semantic technologies to integrate data from a BIM project and other tabular data in a quest for automating, as much as possible, the LEED Certification tasks. We propose an experiment that gathers data from a BIM project and instantiates them into a simplified ontology. Tabular data from the Belo Horizonte's city in Brazil about the surroundings were aggregated and annotated together using a

Semantic Data Dictionary approach (Rashid et al. 2020) to generate RDF[(1)] graph. Inferences performed on the graph provide new knowledge used to assess the LEED criteria in order to certify a constructive project.

The remaining of the paper is organized as follows: Section 2, presents a brief discussion on Knowledge Organization and Domain Analysis. Section 3 introduces the information needed for LEED certification. Section 4 presents the related works. Section 5 points out the methods used to conduct the experiment. In Section 6 we prototype a framework for testing our ideas. In Section 7 we use the prototype to put forth our solutions to the difficulties in the non-automated methods. Finally, Section 8 summarizes the final considerations.

2 Domain Analysis

The Knowledge Organization (KO) is linked to the modeling of knowledge domains, seeking a common core of concepts. In view of the centrality of KO in relation to Information Science, Hjørland (2008) points out that, in part, this core view is due to the fact that KO is interdisciplinary in nature. For this approach, Hjørland (2008) considers two meanings for the concept of information organization: a broad one, which is concerned with answering how knowledge is constructed, from the point of view of the social division of intellectual work and the social organization of knowledge; and a strict one, dedicated to creating and maintaining KOS in order to intermediate registered knowledge (linked to Information Science), concerned with the description of concepts.

Therefore, KO is a key sub-domain of Information Science, which is dedicated to the conceptual order of knowledge (Smiraglia 2015). A pragmatic approach to KO, seen as a predominant logic by Hjørland (2002) focuses on information retrieval. Thus, considering the domain analysis derived from the pragmatic view of the KO, it is necessary to elucidate the main concepts of this area of knowledge and its main approaches, which are important for the development of the prototype presented in this article.

Domain analysis is a proposal developed by Hjørland and Albrechtsen (1995), being conceptually deepened by other studies in the course of subsequent years (Hjørland 2017).

Smiraglia (2015) states that the KO paradigm encompasses domain analysis as a way of visualizing the emergence and coherence of a domain and as a way of mastering the parameters of the universe in which this domain operates.

The pragmatic method of domain analysis, suggested by Hjørland (2002) is based on the determination of objectives and uses of the generated knowledge. The use of domain analysis is linked to the socio-cognitive aspects of the relations between society and the knowledge produced by it. Domain analysis explores ontological relationships, that is, generic relationships in thesauri and classification systems; studies social groups such as scientists, professionals or students; and studies epistemologies, paradigms, traditions, theories, which is important because people tend to organize themselves according to their views (Hjørland 2004).

Smiraglia (2015) brings a similar view of domain analysis, is characterized by the study of the theoretical aspects of a given environment, constituting a means for the generation of knowledge. According to him, domain analysis makes it possible to observe the evolution of knowledge, the sharing of information from different domains and the migration of paradigms from the same domain.

For Hjørland and Albrechtsen (1995), the object of domain analysis is the development of collective information and knowledge structures. For the authors, each domain has its particularities, its ideological discourses and for this reason, they cannot be treated as similar. That is, each domain deserves to be treated in a specific way, as it has different forms of interpretation, depending on the object of study. KO, information structure, research and relevance criteria are interrelated with the work of specific communities. The central question, according to Hjørland (2002), is how to assess the domains of knowledge of specialists in the area.

For the methodological approach of domain analysis, the use of the eleven approaches in Hjorland (2002) is predominant, as it encompasses everything from the epistemological formation of a domain to metric studies for its analysis. For the author, domain analysis approaches should not be used separately and should be combined, at least in two, to characterize and define a domain. Thus, the development of the prototype presented in this article involves the approaches of "building special classifications and thesauri" and "implementing specialized indexing and retrieval of information, through information and communication technologies".

The domain knowledge of the civil construction industry was guided by the understanding of BIM technology and the organization of information for processes such as obtaining a sustainability certification. The following section presents the results of communities in this domain that work on the standardization of terms and their relationships for the development of KOS.

3 Information Organization for LEED Certification

The process for obtaining LEED certification requires analysis, measurement and knowledge management. Knowledge management methods are important for collecting data and documenting work progress. Learning in this process must be documented and made available to increase productivity and reduce future failures (Ofori-Boadu et al. 2012). Considering the BIM use, there is a need to model a structure that represents the knowledge contained in the LEED classification system to be integrated into the BIM classification system (Nguyen et al. 2016). Thus, in addition to understanding the LEED manual indicators (USGBC 2020), this study analyzed the construction industry domain.

Groups of researchers in Linked Data and Semantic Web have worked to develop classification systems for the architecture, engineering and construction on the Web, relating data in place of documents (Borrmann et al. 2018; Curry et al. 2013). The Linked Data Working Group (LDWG) in buildingSMART International and the Linked Building Data Community Group (LBD CG) in the W3C stand out. In these standardization groups, ontologies are proposed to capture construction data using Semantic Web technologies such as OWL and RDF (W3C OWL Working Group 2012). Thus, domain ontologies are built aligned according to the terminology of other ontologies in such a way that standardization is maintained.

The buildingSMART maintains the Industry Foundation Classes (IFC), which is the format for exchanging data from a BIM model between software. IFC is based on the EXPRESS language and concepts, which consists of providing general and broad definitions of products and data from which more detailed and specific models of tasks that support specific exchanges could be defined. The EXPRESS language uses terms such as types, entities and properties, and rules that must be used to create a specific schema (Pauwels and Terkaj 2016). In this sense, IFC was designed to handle all information about a built environment, throughout its life cycle, from feasibility and planning, analysis and simulation in the design and construction phase, to occupation and operation (Sacks et al. 2018).

The LDWG group, under the domain of buildingSMART, focuses primarily on the ifcOWL ontology production. This ontology is designed to be a direct translation of the IFC schema into an OWL representation, according to the summary in Table 1, allowing to continue using this established standard to represent construction data. In addition, it allows one to explore Semantic Web technologies in terms of data distribution, data model extensibility and information retrieval, and reuse general-purpose software implementations for data warehousing, consistency checking, and knowledge inference (Pauwels and Terkaj 2016).

Table 1 - Summary of main conversions

IFC	**OWL**	**Observations**
Entity *Defined Data*	owl:Class	These are the model entities themselves, with their own set of attributes and constraints.
SubTypeOf *SuperTypeOf*	rdfs:subClassOf	Differentiation between general and specific concepts.
Select Types	owl:ObjectProperty	Allow a data to be a subClassOf or UnionOf.
Types e *Simple Data*	owl:DatatypeProperty	Simple data types.
Attributes	owl:ObjectProperty	Non-functional object property. Explicit domains and ranges
Enumerations Data type	owl:oneOf owl:NamedIndividual	Possible values list.
Select Data type	owl:unionOf	Related to classes.
Inverse	owl:inverseOf	Concept inversion.
Attribute of entity data type *Cardinality*	owl:AllValuesFrom or owl:qualifiedCardinality or owl:maxQualifiedCardinalit	Attributes converted to constraints.
Rule	SWRL rules	Rules definition.

Source: Adapted from Paywels and Terkaj (2016).

The ifcOWL ontology is extensive and different from many existing ontologies in many other domains, which generally have a narrower scope and rely on extensions enabled by linked data principles. For this reason, there are attempts to simplify this ontology, such as IFCWoD

(Farias et al. 2015), SimpleBIM (Pauwels and Roxin 2016), BIMSO and BIMDO (Niknam and Karshenas 2017), BimSPARQL (Zhang et al. 2018) and BOT (Rasmussen et al. 2018).

The IFCWoD (IFC Web of Data) implements an adaptation of the IFC model to the OWL, considering the modeling constraints required by the object-oriented structure of the IFC schema. The developed model has a structure based on RDF and OWL resources for data inference, thus taking advantage of all the modeling constraints required by the EXPRESS language object-oriented structure. The authors present comparisons with other solutions demonstrating that the proposed model provides the query execution time optimization (Farias et al. 2015).

The SimpleBIM implements code rewriting rules through direct programming using Semantic Web technology APIs (Application Programming Interface), the use of SPARQL queries and the use of logic rules in an OWL-DL based inference engine (Pauwels and Roxin 2016).

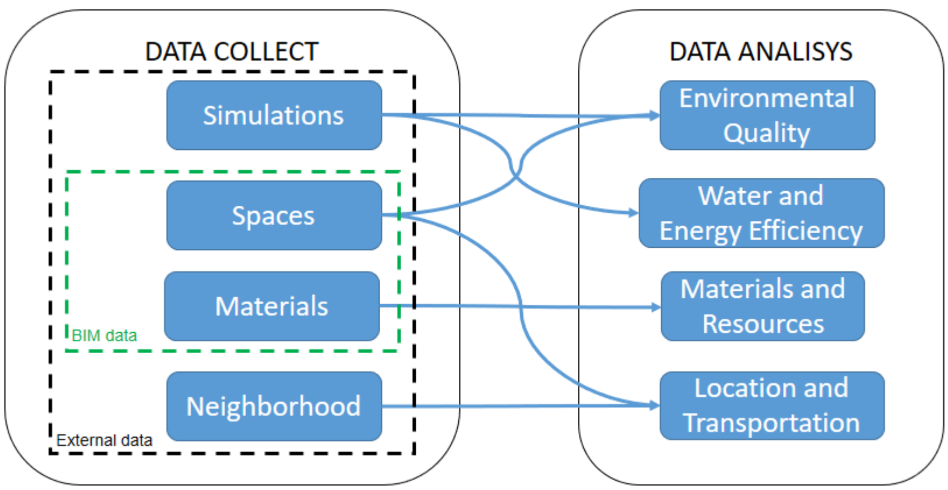
BIM Shared Ontology (BIMSO) and BIM Design Ontology (BIMDO) are another approach to modularity with ontologies projected from scratch and therefore have no connection with IFC. The creation of these ontologies demonstrates that the semantic representation of construction information makes it easier to find and integrate construction information distributed across multiple knowledge bases. BIMSO has a minimal core and is based on the UNIFORMAT II classification system, and BIMDO provides specific terminology for the project (Niknam and Karshenas 2017).

The BimSPARQL also simplifies ifcOWL data, declarative rules with procedural programming to implement extended functions for queries. Through the query language for RDF, spatial and logical reasoning is performed considering the application with data from various sources (Zhang et al. 2018).

A simplification approach that seeks to define an explicit OWL ontology is the Building Topology Ontology (BOT) that involves publishing building products and associated properties on the Web using the Linked Building Data principles. The authors highlight that the existing ontologies in the scope of buildings redefined the same basic elements (for example, spaces, floors and elements) and their relationships. So, a minimal and extensible ontology for this single objective was proposed with the development of a community effort by the W3C LBD Community Group (Rasmussen et al.2018).

This way, for the elaboration of experimentation with the BIM data, the adoption or creation of a simplified ontology is a solution for linking with other data. The process to organize the data for LEED certification includes the use of different tools, documents and comparisons that prove the sustainability of the built environment. Comparing these data requires understanding the criteria available in the manual for certification. These criteria involve analyzing the materials used, the availability around the building, the spaces used, the consumption of water and electricity and the internal environment quality (USGBC 2020). Figure 1 presents the collection steps, involving different data sets, and data analysis, applying inferences for scoring, addressed in the experiment of this research.

Figure 1 – Certification actions



Source: the authors

4 Related Works

It is possible to find works that investigated multidisciplinary interdependencies in green building projects, with a focus on computational optimization and collaboration in the elaboration of the projects (Azhar et al. 2011; Geyer 2012; Hong et al. 2019). It also identifies recent studies related to ontologies uses to classify materials and perform automated analysis of construction characteristics for green certification.

Zhang et al. (2019) use ontology, with SWRL rules (Semantic Web Rule Language), to infer real-time scores of green projects in a computerized social communication environment.

In this same line of research, Jiang, Wang and Wu (2018) present an ontology using SRWL rules. The experimental results demonstrated that the BIM knowledge base can serve for the sustainability of the construction, as well as the sharing, maintenance and knowledge acquisition among the different project participants. It can also be observed in Xu et al. (2019) that using logical inferences to evaluate a criterion allows one to research aspects that need improvements in the building and help project managers to use BIM data.

These related works do not address, however, the integration with open data for environment analysis, nor the application of the inference for quantitative analysis, which are key requirements for our research project. Our proposal seeks to model more complex relationships than those found in the above works, intending to perform a more comprehensive LEED certification analysis.

5 Methodology

This research is of an applied nature since generates knowledge to automate the assessment of sustainable buildings. For the domain analysis, knowledge from previous studies and the understanding of the criteria for LEED certification was used. Regarding the approach, this research is configured as descriptive, prescriptive and qualitative.

By providing more information about the representation and organization of information through ontologies, this research presents a descriptive approach. The prescription occurs using the Design Science methodology and the Wieringa (2009) regulatory cycle, which consists of a logical structure with useful guidelines for solving problems in the elaboration of an experiment (Bax 2013). The qualitative character is justified by the study of methodologies for the construction and integration of ontologies, as well as the inferences application to generate new knowledge (Creswell 2014).

Our experimental procedure to automate the LEED certification process involves developing a simplified ontology, based on the ifcOWL, to integrate BIM data and certification criteria together. Data sets were used as a model for preparing the ontology. Because it involves the elaboration of a simplified ontology, the basis for guide 101 stands out (Noy and McGuinness 2001). This is a simple method, based on development experience with the Protégé software, and useful in explaining more specific steps in the ontology construction, such as defining properties of classes and relationships, creating data, cardinality and instance constraints, among others.

The Systematic Approach for Ontology Construction (SABiO) is also considered in the ontology development of this work. The SABiO approach focuses on the domain ontologies development and also proposes supporting processes. SABiO distinguishes between reference and operational ontologies, providing activities that apply to the domain ontologies development (Falbo 2014).

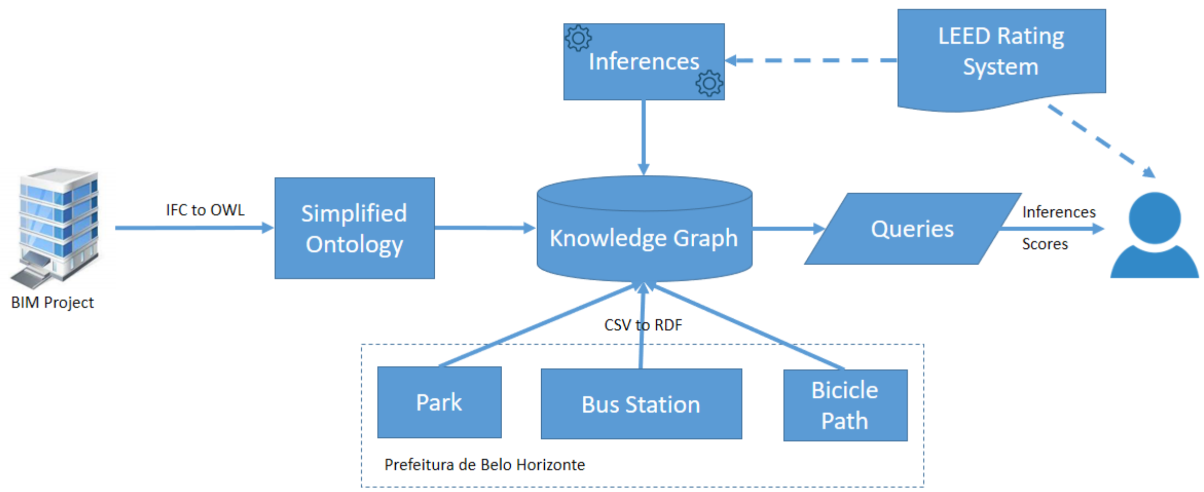
The procedures performed to simplify the ifcOWL ontology were based on the proposal by Pauwels and Roxin (2016), which is directly related to the presence of elements and EXPRESS format resources.

6 An Experiment to Automate LEED Certification

Figure 2 presents the orchestration of methods that make up the workflow, still manual, which will later be integrated into the architecture of a prototype. This process requires the knowledge of information technologies and, moreover, can be considered complex for the use of a certification specialist. The implementation seeks to validate the prototype as a potential solution to the problem of this research.

At the beginning of the workflow, a IFC file from a BIM project is converted, as described in Section 6.1, into a simplified ontology (Section 5.1) for evaluating the LEED criteria. To integrate external data, open data, in CSV (Comma-separated values) format, are obtained, mapped to ontology concepts and converted into RDF graphs (Section 5.2). The datasets are inserted into the Parliament triplestore to perform queries and inferences in SPARQL (Section 5.3). A final query is performed to present the inferences result about the scores obtained for the certification.

Figure 2 – Workflow for the prototype architecture

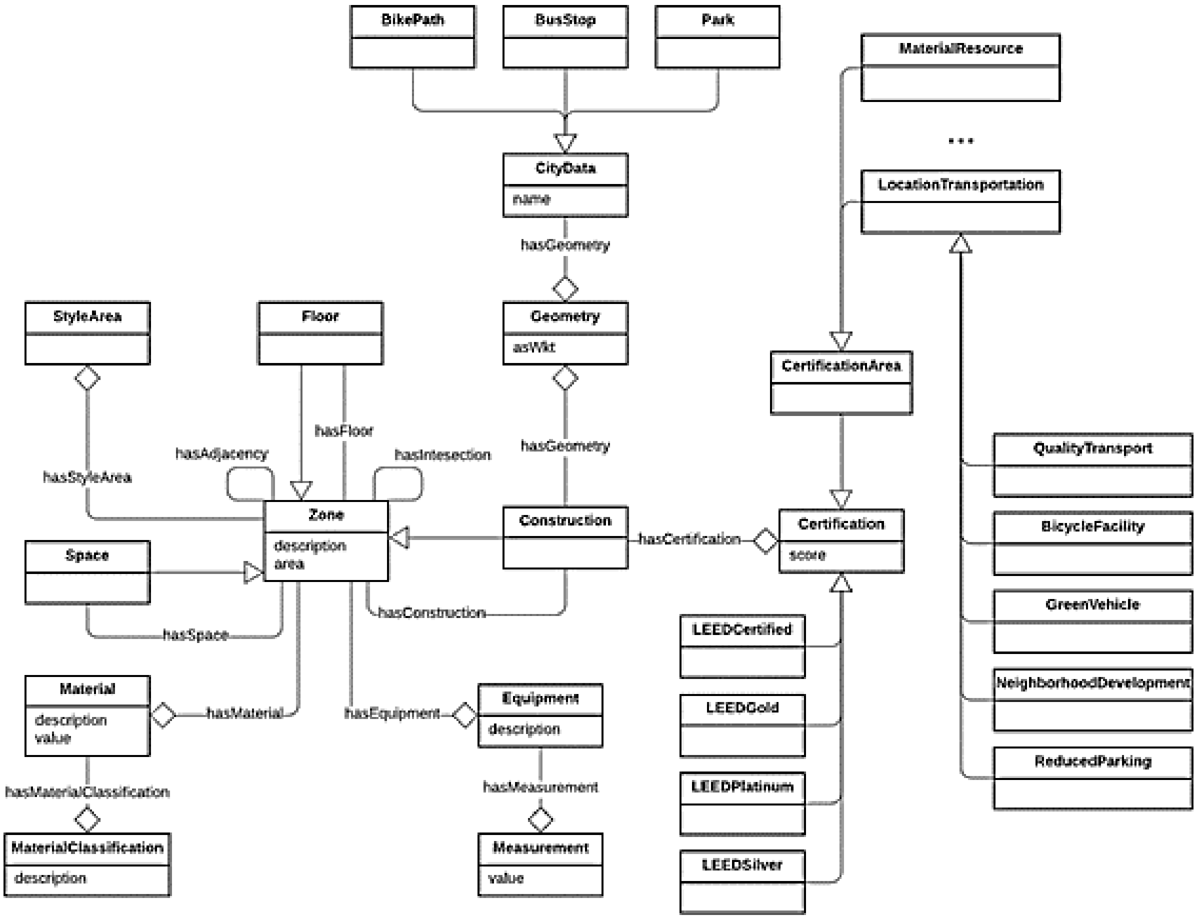


Source: the authors

6.1 The SEBIM Ontology

The SEBIM (Semantic BIM) ontology (Figure 3) was created in Protégé 5.5, with classes representing the data needed to organize the information and analyze the LEED certification criteria. The ontology reuses concepts from the ifcOWL (Pauwels and Terkaj 2016), SimpleBIM (Pauwels and Roxin 2016), BOT (Rasmussen et al. 2017) and BIMSO (Niknam e Karshenas 2017), to semantically organize the BIM data. LEED criteria covering the use of materials and spaces, among others, are considered. These criteria, in the ontology context, support the inferences that assess the building state.

Figure 3 – SEBIM Ontology



Source: the authors

The construction elements were chosen to meet the analysis of the criteria selected for the experiment, but new classes can be included according to other analyses, following the same method. The built areas, namely zones (***Zone*** class), are divided into ***Construction***, ***Space***, and ***Floor*** subclasses. These types of zones are related to each other through the intersection (*hasIntersection*) and adjacency (*hasAdjacency*) properties, making it possible to interconnect zones each other. For example, spaces instances in an apartment (space instance) consist of interconnected zones: bedrooms, bathroom, living room and kitchen. The ***Zone*** relates to the ***StyleArea*** class and can have equipment installed. The ***StyleArea*** is an important classification to organize and identify relevant areas for environmental assessment, such as green areas, parking for green vehicles, among others.

Because many certifications require to assess the materials and equipment used in the construction, two classes abstract these elements, and two properties (*hasMaterial* and *hasEquipement*) represent their inclusion in a ***Zone***. The linking of the zone with equipment involves the internal location of sensors, air conditioning equipment, photovoltaic panels, among others. The zones relationship with materials (*hasMaterial*) seeks to identify compliance with sustainability requirements. The materials are related to their raw composition and to their classification, according to the type of use. The attributes are not displayed on the diagram.

In order to be able to apply the certification criteria, a ***Construction*** is related to the ***Certification*** class (*hasCertification*). In the LEED manual, the criteria are divided into themes to assess sustainability. Therefore, the criteria are subclasses specialized according to ***LocationTransportation***, ***WaterEfficiency***, ***EnergyAtmosphere***, ***ResourcesMaterials***, ***RegionalPriorityCredit***, ***IndoorEnvironmentalQuality*** and ***Innovation*** classes. Between the ***LocationTransportation*** and ***ResourcesMaterials*** classes, there are other groupings, suppressed in the diagram due to space limitations. The ***LocationTransportation*** subclass, as well as the other subclasses, separately record each score received to organize the information. The score received is assigned to a subclass, as done in a Query (2) (in Section 5.3) for the ***TransportQuality*** subclass.

The attributes have numerical values (*Certification.score, Material.value, Zone.area* and *Measurement.value*); alphanumeric descriptions (Zone.*description*, *Material.description* and *Equipment.description*); and geospatial data (*asWkt* [(2)]). The latter being necessary to integrate external open data about the built surroundings.

6.2 Semantic Data Annotation for Some LEED Certification Criteria

Publishing open data on the Web is an action by city halls that promotes the smart city (Isotani and Bittencourt 2015). On the Belo Horizonte's city website (http://bhmap.pbh.gov.br/v2/mapa/idebhgeo), it is possible to find data sets about services and locations, but still without using the Semantic Web potential. The amount of data is excessive for human interpretation and makes data interconnection expensive and error-prone.

Open data on transport facilities and parks are used to assess criteria in the ***LocationTransportation*** class. These data have geometric coordinates of city areas and points, following OGC (Open Geospatial Consortium) standardization, which defines a vocabulary to represent geospatial data. This data format is required to perform queries by applying GeoSPARQL functions, which return triples according to proximity and relationships between coordinates (Perry and Herring 2011).

Other datasets, in tabular formats, can also be integrated with data from a BIM project. To do this, it is necessary to convert these formats by applying semantic technologies. Typically, these datasets are usually accompanied by a data dictionary that describes the data. The Semantic Data Dictionary (SDD) is an alternative to represent machine-readable metadata for a dataset (Rashid et al. 2020). SDD formalizes the assignment of a semantic representation to the data, annotating the columns of these sets and their values, using vocabulary concepts and best-practice ontologies. Approaches to mapping and semantic representation of this data promote discovery, interoperability, reuse and traceability. The formalization of tabular data in RDF format is performed by converting its structure into an OWL ontology.

Faced with options for converting tabular data to RDF (Ding et al. 2011; Jeremy Tandy et al. 2015; Rashid et al. 2020), Semantic Data Dictionaries were chosen, implemented in the *sdd2rdf* tool. Objects and their attributes are represented and identified through relevant ontologies that constitute this information in a formally precise and machine-readable manner (Rashid *et al.* 2020).

The RDF graph generated by execution of the *ssd2rdf* script contains the data formalization and support integration to query city locations by spatial data. Thus, data on cycle paths, parks and bus stops are converted into RDF.

6.3 Assess Certification Level Using Inferences

World Wide Web Consortium (W3C) standards, including RDF and OWL, provide semantic interpretations for RDF graphs that allow one to infer additional RDF instructions from explicitly given statements. Many applications that rely on these semantics require a query language such as SPARQL, which is a well-established language to combine arbitrary RDF and return attributes and transformations from them. This language has become an industry standard for representing rules and constraints in Semantic Web models, being well supported by multiple engines and databases. Thus, triples returned in a SPARQL query can also be considered inferences in a reasoning process (Coppens et al. 2013; Terkaj and Sojic 2015).

Logic rules allow one to automatically calculate construction project scores on the internal data analysis clauses of the BIM project. According to Bassiliades (2018) SWRL has become a popular choice for rule-based applications. However, as SWRL has existed for more than ten years and has not yet reached the industrial world, we chose to use SPARQL over the generated RDF. Since these data are integrated with others, SPARQL is a solution to centralize inferences. Compared to specific query languages, SPARQL is especially applicable to scenarios where data from multiple sources is required (Krijnen and Beetz 2018).

The logical inferences resulted from data processing, were made with SPARQL, applying an INSERT clause to insert the inferred triple in the dataset, attributing the score to the class instance corresponding to the criterium. Query (1) displays the inference made to rate and score the bicycle facility which evaluates a bicycle area installation, calculated on the quantity of the units in the building.

Query 1 – Inference for Bicycle Facilities

```
PREFIX sebim: <http://local/SEBIM#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd:  <http://www.w3.org/2001/XMLSchema#>

INSERT {?cr sebim:score "1"^^xsd:decimal}
WHERE {
   ?cr rdf:type sebim:BicycleFacilities.
   ?build rdf:type sebim:Material.
   ?build sebim:hasMaterialClassification sebim:RackBike.
   ?build sebim:value ?n.
  { SELECT (count(?s) as ?place) WHERE { ?s rdf:type sebim:Space.} }
}
GROUP BY ?cr ?build ?n ?place
HAVING (?place >= ?n)
```

Source: the authors.

Queries for surroundings analysis require using GeoSPARQL functions to capture the geospatial coordinates of the environment. This is accomplished by using Well-Known-Text (WKT) polygon representations, standardized by the OGC. WKT is a text markup language for representing vector geometry objects on a map, spatial reference systems of spatial objects, and transformations between spatial reference systems (Perry and Herring 2011). Query (2) displays the inference that analyzes the bus stops proximity and creates the triple with the score: *<000132 sebim:score 1>*, where *000132* is the ***QualityTransport*** LEED criterium instance for the project in question, which was given the value '1' as its score, by inference.

Query 2 – Inference for Transport Quality

```
PREFIX sebim: <http://local/SEBIM#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd:  <http://www.w3.org/2001/XMLSchema#>
PREFIX uom: <http://www.opengis.net/def/uom/OGC/1.0/>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>

INSERT {?cr  sebim:score "1"^^xsd:decimal }
WHERE {
  ?build rdf:type sebim:Construction.
  ?build geo:asWkt ?pb.
  ?build ?x sebim:Busstop.
  ?bus ?y sebim:Geometry.
  ?bus geo:asWKT ?pl.
  ?cr rdf:type sebim:QualityTransport.
  FILTER (geof:distance(?pb, ?fpl, uom:metre) <= 800)
}
```

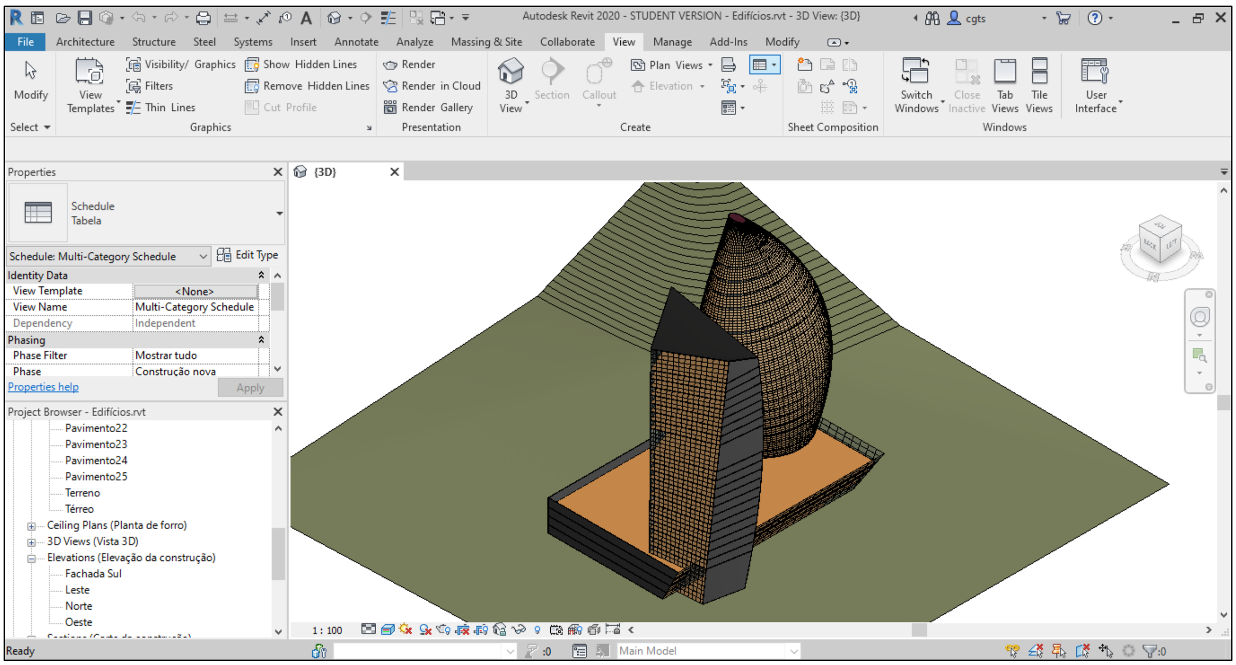
Source: the authors.

We claim that the process of inserting new triples via SPARQL INSERTs is a viable path to accelerating the process of LEED certification criteria evaluation. The initial experiment was based on fictitious instances and attributes to validate the criteria inferences. As the LEED certification provides criteria by typology of the construction (new construction, maintenance, hospitals and sheds), the evaluation of new construction was adopted. Given the complexity of information required to assess all LEED's criteria, those representing different scenarios for data extraction were selected **[(3)]**. For example, an internal scenario where only project data was used (Query 1), and another scenario integrating external data from the Belo Horizonte's city website (Query 2). Thus, considering the 57 criteria for new construction, seven criteria were implemented, as explained in the next section.

7 Prototype

To experiment with the methods used in the prototype, an academic project elaborated by students of the architecture course at FUMEC University was chosen. The modeling, shown in Figure 4, was created in Autodesk Revit [(4)] software and contains glazed buildings with 27 areas (floors).

Figure 4 - BIM Project in Autodesk Revit



Source: the authors.

7.1 Converting Data to the SEBIM Ontology in OWL

In the first attempt to export data from the BIM Project to integrate it with other external data, it was decided to generate the IFC file of the project in Revit. We tried to convert the IFC format into an RDF format. The tools used were the IFCtoRDF (Pauwels and Terkaj 2016) to convert the data using the ifcOWL ontology, and the IFCtoLDB (Bonduel et al. 2018) to convert the data as per the BOT ontology. Both convert the data from the IFC file to their respective ontologies, annotating it. However, due to the limitation of the tool, a semantic superficiality of the relations between the classes of these ontologies was obtained (exclusively subsumption relations), the final organization of the information constitutes a taxonomy rather than an ontology.

Given the insufficient result, a second attempt to export data involved generating tabular files in Revit, instead of exporting in IFC format. This attempt proved to be more effective since it allowed the use of *sdd2rdf* tool to annotate data using the SEBIM ontology. Due to the Revit export filter limitation, two files were generated, one containing "material data", and other "area data".

The data from the BIM project to be assessed as sustainable and the data from Belo Horizonte's city were semantically annotated using the semantic dictionary and converted in an integrated way to the RDF standard (with the *sdd2rdf*) and, finally, inserted in the triplestore.

7.2 Executing Inferences

The inferences execution via SPARQL queries generated new triples containing the final score to verify whether or not the LEED certification criteria were met. The criteria analyzed in the project in question are presented in Table 2.

Table 2 – Experiment criteria scoring

Category LEED	**Criteria**	**Max Score**	**Score**
Location and Transportation	Access to quality transport	5	5
Location and Transportation	Bicycle facilities	1	0
Location and Transportation	Green vehicles	1	0
Sustainable Sites	Open space	1	1
Energy and Atmosphere	Renewable Energy Production	3	0
Materials and Resources	Material ingredients	2	2
Environmental Quality	Indoor lighting	2	2

Source: the authors

It can be concluded that only two criteria did not obtain the maximum score. It is also worth mentioning that criteria with a maximum score above 1, may present score scales defined in the LEED manual.

7.3 Discussion of Results

During the bibliographic survey of this research, the difficulties identified (Table 3) in the traditional process of organizing information for LEED certification are used to validate the pertinence of the application of semantic technologies used in the proposed prototype.

The use of BIM is a determining factor for the collaborative work evolution in the construction industry. Recording data, associated with the three-dimensional elements, is essential to organize project information. Even using specialized BIM tools, information retrieval is laborious. Thus, converting the IFC to OWL format could be proposed as a solution to increase data semantics, and facilitate information retrieval for analysis. In addition, integration with other datasets, such as semantically annotated city hall data, is facilitated.

Table 3 – Solutions to the found difficulties.

Difficulty / Research challenge	**Solution**
IFC format extension of BIM technology	Conversion of IFC to SEBIM
Integration with external data	Semantic annotation of tabular data
Information organization	Knowledge graphs
Data analysis (LEED criteria)	SPARQL inference
Data recovery	SPARQL queries

Source: the authors.

In the prototype's first version, the approach based on monotonic ontology was sought, in which the SWRL rule language was used to formalize the rules to infer the scores. However, this approach presented some limitations, especially in the complex rules modeling and ease of maintenance. Thus, the rule-based reasoning model implementation used SPARQL queries, generating new triples for inferences about each criterion score.

Since the evaluation process covers construction project phases and requires reviewing and updating the project to meet the criteria, it is important to consider the information historical organization. In this way, uploading a BIM project new version allows one to compare results and analyze the evolution in meeting the criteria for certification.

The workflow proposed by this research allows implementing a friendly solution for specialists who are unaware of semantic technologies. The experiment can be implemented in frameworks like JenaSemanticWeb and VirtuosoOpenSource, which support GeoSPARQL functions and resources for inferences.

8 Conclusions

This work presents a domain analysis application with the development of a prototype to integrate data and generate inferences with Semantic Web technologies to automate the LEED certification assessment. BIM technology, a reference in the construction industry, is used as the basis of a simplified ontology to integrate data. Open data, in tabular format, is semantically annotated using a semantic data dictionary. With the data integrated into a triplestore, SPARQL queries provided inferences that reveal the BIM project meets or does not the certification criteria. The results suggest that this solution promotes the semantic extension of constructive elements in a BIM project, facilitates integration with other knowledge bases and organizes the data for information retrieval.

In the domain analysis context, it is noteworthy that the Information Science focus is on the knowledge domain to develop databases that are used in Computer Science to computerize access to human knowledge. The process of knowledge organization depends on the use of tools. Therefore, it is necessary to use techniques that preserve the relevance of the human specialist, regardless of the change in access tools.

New experiments will provide enrichment of the knowledge base with the refinement of queries to implement an efficient system to assess the LEED certification degree of building projects. In addition, other information, such as certified buildings in the surrounding area and other external information on materials, can be incorporated into the database to meet the other criteria for evaluation.

The knowledge organization system development provides information retrieval and allows for more efficient and persuasive methods for project management that facilitate decision-making. The built ontology can also be reused and expanded to include other project needs. It is expected that future similar applications will collaborate in the construction projects evaluation such as the authorization of construction in city halls, among other evaluations that require complex analysis.

The prototype experimentation was limited to an academic project with less information than a real professional project. Due to copyright restrictions, this work did not get projects from a real building. Thus, the development of a tool can enable its use in real cases.

This work contributes to the report of the development of a system for knowledge organization that favors information retrieval, allowing more efficient and persuasive methods for project management that facilitate decision making, rather than an experience-based approach. The built ontology can be reused and expanded to include other project needs. It is expected that future similar applications collaborate in the evaluation of construction projects such as building permits in city halls, among other evaluations that require complex analyses.

Notes

(1) Resource Description Framework is a standard model for data interchange on the Web. https://www.w3.org/RDF/.

(2) Geometric coordinates from OGC (Open Geospatial Consortium) standardization.

(3) Mapping performed in this work: https://github.com/cgtsbr/sebim/blob/main/Inferencias.xlsx.

(4) Revit enables three-dimensional modeling in BIM, supports design, simulation of building analysis, and exports data for integration with other tools.

References

Azhar, Salman, et al. "Automation in Construction Building information modeling for sustainable design LEED ® rating analysis". *Automation in Construction*, vol. 20, no. 2, 2011, pp. 217–24.

Bassiliades, Nick. "SWRL2SPIN: converting SWRL to SPIN". *CEUR Workshop Proceedings*, vol. 2204, 2018.

Bax, Marcello Peixoto. "Design science: filosofia da pesquisa em ciência da informação e tecnologia Design science: philosophy of research in information science and technology". *Ci. Inf.*, vol. 42, no. 2, 2013, pp. 298–312.

Bonduel, Mathias, et al. "The IFC to linked building data converter - Current status". *CEUR Workshop Proceedings*, vol. 2159, 2018, pp. 34–43.

Borrmann, André, et al. “Building information modeling: Technology Foundations and Industry Practice”. *Building Information Modeling: Technology Foundations and Industry Practice*, Springer, 2018.

Coppens, Sam, et al. “Reasoning over SPARQL”. *CEUR Workshop Proceedings*, vol. 996, 2013.

Creswell, Jhon W. “Research Design: Qualitative, Quantitative, and Mixed Methods Approaches”. *SAGE*, 2014.

Curry, Edward, et al. “Linking building data in the cloud: Integrating cross-domain building data using linked data”. *Advanced Engineering Informatics*, vol. 27, no. 2, Elsevier Ltd, 2013, pp. 206–19.

Ding, Li, et al. “TWC LOGD: A portal for linked open government data ecosystems”. *Journal of Web Semantics*, vol. 9, no. 3, 2011, pp. 325–33,

Falbo, Ricardo de Almeida. “SABiO: Systematic approach for building ontologies”. *CEUR Workshop Proceedings*, vol. 1301, 2014.

Farias, Tarcisio Mendes de, et al. “IfcWoD, Semantically Adapting IFC Model Relations into OWL Properties”. *Proc. of the 32° CIB W78 Conference*, 2015, pp. 1–86.

Geyer, Philipp. “Advanced Engineering Informatics Systems modelling for sustainable building design”. *Advanced Engineering Informatics*, vol. 26, no. 4, 2012, pp. 656–68.

Hjørland, Birger. “Domain Analysis: A Socio-Cognitive Orientation for Information Science Research”. *Bulletin of the American Society for Information Science and Technology*, vol. 30, no. 3, 2004, pp. 17–21.

Hjørland, Birger. “Domain analysis in information science: Eleven approaches - Traditional as well as innovative”. *Journal of Documentation*, vol. 58, no. 4, 2002, pp. 422–62.

Hjørland, Birger. “Reviews of concepts in knowledge organization”. *Knowledge Organization*, vol. 44, no. 5, 2017, pp. 349–79.

Hjørland, Birger. “What is Knowledge Organization (KO)?” *Knowledge Organization*, vol. 35, no. 2/3, 2008, pp. 86–101.

Hjørland, Birger, and Albrechtsen, Hanne. “Toward a new horizon in information science: Domain-analysis”. *Journal of the American Society for Information Science*, vol. 46, no. 6, 1995, pp. 400–25.

Hong, Sim-Hee, et al. “Automated management of green building material information using web crawling and ontology”. *Automation in Construction*, vol. 102, 2019, pp. 230–44.

Isotani, Seiji, and Bittencourt, Ig Ibert. “Dados Abertos Conectados”. *NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO Br – NIC.br Centro de Estudos sobre Tecnologia Web – CeWeb.br*, 2015.

Jeremy Tandy, et al. “Generating RDF from Tabular Data on the Web”. *W3C Recommendation*, 2015, pp. 1–29.

Jiang, Shaohua, et al. “Combining BIM and Ontology to Facilitate Intelligent Green Building Evaluation”. *Journal of Computing in Civil Engineering*, vol. 32, no. 5, 2018, pp. 1–15.

Krijnen, Thomas, and Beetz, Jakob. “A SPARQL query engine for binary-formatted IFC building models”. *Automation in Construction*, vol. 95, 2018, pp. 46–63.

Niknam, Mehrdad, and Karshenas, Saeed. “A shared ontology approach to semantic representation of BIM data”. *Automation in Construction*, vol. 80, 2017, pp. 22–36.

Noy, Natalya F. and Deborah L. McGuinness. “Ontology Development 101: A Guide to Creating Your First Ontology”. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, 2001.

Ofori-Boadu, A. et al. "Exploration of management practices for LEED projects: Lessons from successful green building contractors". *Structural Survey*, vol. 30, no. 2, 2012, pp.145-162.

Pauwels, Pieter and Roxin, Ana. “SimpleBIM: From full ifcOWL graphs to simplified building graphs”. *European Conference on Product & Process Modelling*, 2016.

Pauwels, Pieter and Terkaj, Walter. “EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology”. *Automation in Construction*, vol. 63, 2016, pp. 100–33.

Pauwels, P., de Farias, T. M., Zhang, C., Roxin, A., Beetz, J., De Roo, J., & Nicolle, C. A performance benchmark over semantic rule checking approaches in construction industry. *Advanced Engineering Informatics, vol. 33*, 2017, pp. 68–88.

Perry, Matthew, and Herring, John. “GeoSPARQL - A geographic query language for RDF data”. *Open Geospatial Consortium*, 2011.

Rashid, Sabbir M., et al. “The Semantic Data Dictionary – An Approach for Describing and Annotating Data”. *Data Intelligence*, 2020, pp. 443–86.

Rasmussen, Mads Holten, et al. “Demo: Integrating building information modeling and sensor observations using semantic web”. *CEUR Workshop Proceedings*, vol. 2213, 2018, pp. 48–55.

Rasmussen, Mads Holten, et al. “Web-based topology queries on a BIM model”. *5th LDAC workshop*, 2017.

Sacks, Rafael, et al. *BIM Handbook*: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers, and Contractors. John Wiley & Sons, Inc., 2018.

Smiraglia, Richard P. "Domain Analysis for Knowledge Organization: Tools for Ontology Extraction". *Domain Analysis for Knowledge Organization*, Chandos Pu, 2015.

Terkaj, Walter, and Sojic, Aleksandra. "Ontology-based representation of IFC EXPRESS rules: An enhancement of the ifcOWL ontology". *Automation in Construction*, vol. 57, 2015, pp. 188–201.

USGBC. "LEED V4.1: Building Design and Construction". *US Green Building Council*, 2020, pp. 259.

W3C OWL Working Group. "OWL 2 Web Ontology Language Document Overview". *W3C Recommendation*, 2012, pp. 1–7.

Wieringa, Roel. "Design science as nested problem solving". *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, DESRIST '09, 2009.

Wong, Johnny Kwok Wai, and Zhou, Jason. "Enhancing environmental sustainability over building life cycles through green BIM: A review". *Automation in Construction*, vol. 57, 2015, pp. 156–65.

Xu, Zhao, et al. "Study on the Evaluation Method of Green Construction Based on Ontology and BIM". *Hindawi*, 2019, pp. 1–22.

Zhang, Chi, et al. "BimSPARQL: Domain-specific functional SPARQL extensions for querying RDF building data". *Semantic Web*, vol. 9, no. 6, 2018, pp. 829–55.

Zhang, Daxin, et al. "A semantic and social approach for real-time green building rating in BIM-based design". *Sustainability (Switzerland)*, vol. 11, no. 14, 2019, pp. 1–16.

Research Data

Disponível em: https://github.com/cgtsbr/sebim.

Received: 15/10/2021 Accepted: 20/05/2022