Article

E-ISSN 1808-5245

Application of inferences in ontological modeling

Marcello Peixoto Bax[1]

[1]Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil;
bax.ufmg@gmail.com; ORCID https://orcid.org/0000-0003-0503-3031

Cristiano Geraldo Teixeira Silva[2]

[2]Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, MG, Brasil;
cgts.br@gmail.com; ORCID https://orcid.org/0000-0001-5010-3081

Abstract: Data modeling supported by ontologies favors the organization of knowledge and the use of inference machines to obtain new knowledge about data and support expert decision-making. However, it is necessary to discuss the techniques to promote inferences and explore the semantics between the data. The research problem is to verify how the use of semantic technologies would represent an advantageous alternative, in relation to the state of the art, to analyze, systematize and organize the information of a domain, accelerating a decision-making process. It presents the concepts behind the development of an artifact that employs semantic technologies and a domain ontology to integrate data from Building Information Modeling projects and other tabular data external to the project to automate Leadership in Energy certification tasks and Environmental Design. To demonstrate the artifact, data is semantically annotated and a knowledge graph is generated in Resource Description Framework format. Once integrated into the graph, inferences are applied to evaluate the certification criteria through queries. The results suggest that the application of these technologies promotes semantic extension, facilitates integration with other knowledge bases, and conceptually organizes data to better retrieve information and generate new knowledge.

Keywords: inference; domain analysis; semantic interoperability

1 Introduction

Semantic Web technologies provide a modeling environment that can handle heterogeneous data, supporting interoperability across multiple knowledge domains, integrating distributed data, and employing reasoning engines to automatically infer new knowledge (Terkaj; Sojic, 2015).

For data modeling to be efficient, it is necessary to organize the knowledge derived from documents and information with the support of information systems. According to Hjørland (2017), supporting ontologies favors the development of knowledge organization systems and reinforces the integration of Computer Science with Information Science. Works in this aspect apply knowledge from methods and results of domain analytical studies on professional cognition, knowledge representation in computer science and artificial intelligence.

A domain ontology is a conceptual model for representing a certain portion of reality according to a shared conceptualization. To develop a quality ontology, it is necessary to satisfy the general requirements of a conceptual modeling language, that is, it must have a foundation ontology as an underlying theory (Guizzardi *et al*., 2011).

Modeling for the Semantic Web can be performed using descriptive logic, on which ontologies are based. The reasoning systems treat the semantics of the ontologies, creating other relations by inferences. Another paradigm is based on turning a subset of the semantics into rules that are used by a rule engine to infer new knowledge. However, a solution that has stood out in the industry is the creation of rules and restrictions with the Protocol and RDF Query Language (SPARQL) query language (Bassiliades, 2018).

Therefore, this work presents research on the use of domain ontology and semantic technologies to integrate data and automate tasks to infer new knowledge. As an experiment, data from a civil construction project were serialized into a grounded ontology, tabular data from Belo Horizonte city about the surroundings were aggregated and annotated together using a semantic data dictionary to generate Resource Description Framework (RDF) graphs. SPARQL inferences performed on the graph provided new knowledge about meeting the criteria to certify the sustainability of a constructive project according to the Leadership in Energy and Environmental Design (LEED) certification system (USGBC, 2020).

Section 2 of the article presents a brief discussion on Knowledge Organization and Domain Analysis. In Section 3, the theoretical bases for the organization of information through conceptual modeling are described. Section 4 presents the techniques for carrying out inferences in ontologies, Section 5 addresses the base domain for the development of the artifact, Section 6 points out the methodologies addressed for the development of this work, Section 7 reports the applied research process in the context of the sustainability of constructive projects and Section 8 demonstrates the experimentation of the artifact and describes solutions adopted in the face of difficulties in the traditional method. Finally, Section 9 summarizes the final considerations.

2 Domain analysis

Knowledge Organization (KO) is linked to the modeling of knowledge domains, seeking a common core of concepts. Bearing in mind the centrality of KO in relation to Information Science, Hjørland (2008) points out that, in part, this central view is because KO is interdisciplinary in nature. For this approach, Hjørland (2008) considers two meanings for the concept of knowledge 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 the creation and maintenance of Knowledge Organization Systems to intermediate registered knowledge (linked to Information Science), concerned with the description of concepts.

Therefore, KO is a key subdomain of Information Science, which is dedicated to the conceptual order of knowledge (Smiraglia, 2015). A pragmatic approach to KO, seen as a predominant rationale 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, important for the development of the artifact presented in this article.

Domain analysis is a proposal developed by Hjørland and Albrechtsen (1995) conceptually deepened by other studies over the 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.

For the methodological approach of domain analysis, the use of eleven approaches by Hjorland (2002) prevails, as it ranges 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 artifact presented in this article involves the approaches of construction of special classifications and thesauri and implementation of indexing and specialized retrieval of information, through information and communication technologies.

3 Organization of information

Rationale ontologies describe the categories that are used for conceptual modeling. This work highlights the Unified Foundational Ontology (UFO) that has been developed over the last few years by bringing together axiomatic theories that deal with the main categories of concepts used in conceptual modeling. In conceptual modeling, UFOs are organized into Universals, which refer to types (classes, entities) and Individuals, which refer to type instances (Guizzardi, 2005).

In UFO, these types or classes are organized according to the ontological/metaphysical nature of their properties and other principles and are classified into: kinds, subkinds, phases, roles, categories, mixins and rolemixins. Some types are persistent (endurants) or occurrences (occurents); others have essential or accidental properties; and still others can be counted or ordered (sortals). These types can be rigid or anti-rigid, in the sense that their instances can change their original type. That is, if a class is rigid, it means that its instances maintain their unique "identity principles" during their existence. Thus, a thing that is something never becomes something else. This is equivalent to stating that a rigid class cannot change its instances of a rigid class to another rigid class (Suchánek, 2022). In summary, classes are understood as:

a) kinds and subkinds represent essential properties of objects. They are rigid or static types. A kind captures the essential properties of objects in a domain and can be subdivided into subkinds. As an example, kind "Person" can be specialized in subkinds "Man" and "Woman";

b) phases represent types of classes that have properties intrinsic to objects, while Roles represent types of classes that have properties that objects have in a relational context, that is, contingent relational properties (extrinsic). As an example, a person may be an adult and play the role of a student;

c) categories represent essential properties that are shared by entities of different types. As an example, the geospatial location of a building, a park or a bus stop;

d) mixins represent shared properties for things with different identity principles. RoleMixins is equivalent to a Role for types that aggregate instances with different identity principles. As an example, we can aggregate different roles of components of a computer in a RoleMixin.

In addition to the relationships between classes, the Relator class stands out, which represents the objectification of relational properties between things. As an example, a contract class might contain the relational properties between customer and supplier.

The UFO is divided into incremental layers called: ontology of endurants (UFO-A), ontology of perdurants (UFO-B) and ontology of social aspects (UFO-C) (Guizzardi *et al.*, 2011). Recently, gUFO was developed which provides a lightweight implementation of UFO suitable for Web Ontology Language (OWL) Semantic Web applications. Using a subset of UFO-A and UFO-B, a lightweight ontology can be designed to provide an artifact to structure a knowledge base (Almeida *et al*., 2019).

4 Inferences

Ontology editors support the ontology creation task. Protégé is an example of this type of tool and provides resources for making inferences. It is an extensible, cross-platform environment developed at Stanford University. In addition to being an ontology editor, this tool contains a library of modules that can be installed (plug-ins) to add more functions to the environment. The main functions of Protégé are loading and saving OWL and RDF ontologies; edit and view

Semantic Web Rule Language (SWRL) classes, properties, and rules; define logical class characteristics as OWL expressions; run reasoners for inferences, such as description logic classifiers; and edit OWL individuals for Semantic Web markup.

In view of the objective of applying inferences to support decision-making, this tool is considered as an option for this purpose. The reasoners added to this tool allow making inferences based on the specifications in the created ontology.

The use of SWRL rules, associated with the use of reasoners in Protégé, allows the implementation of inferences. Depending on the inference, decidability can be performed in terms of the T-Box (concepts or terminology box) or the A-Box (individuals or facts) of a given ontology. In terms of the T-Box, the security of the description logic imposes a limitation that only previously defined classes or terms in the ontology can be used in SWRL rules. This means that no new or unknown individuals can be added to the terminology box. Regarding the A-Box, the security of the description logic determines that the variables used in the consequent rule must also exist in its antecedent, that is, no anonymous or isolated individual can be introduced in the A-Box of the ontology (Lawan; Rakib, 2019).

Decidability can be recovered by restricting the form of admissible rules, imposing an adequate safety condition. SWRL is an option for developing rules and ontology-based applications when supported by the Protégé ontology editor, as well as popular rule engines and ontology reasoners such as Jess, Drools, and Pellet (Bassiliades, 2018).

Applied to the context of civil construction, inferences about spatial data, materials used and other information about the project can provide knowledge about sustainable aspects of a built environment. Rule in Figure 1 exemplifies an inference about verifying the minimum percentage of green area in a building to obtain a score for LEED certification. The calculation to obtain the minimum percentage is performed from the green area (?a), divided by the total (?t) of the building (?x). If the antecedent is true, that is, the percentage is greater than 30%, a triple for the score is created.

Figure 1 - SWRL rule

```
sebim:ifcBuilding(?x) ^ sebim:ifcSpace(?y) ^ sebim:ifcAreaStyle(?z) ^ sebim:ifchasSpace(?x, ?y) ^ sebim:ifchasAreaStyle(?y, ?z) ^ sebim:ifcArea(?y, ?a) ^ sebim:ifcArea(?x, ?t) ^ swrlb:divide(?c, ?a, ?t) ^ swrlb:greaterThanOrEqual(?c, 0.3) -> sebim:score(?x, 1)
```

Source: Created by the authors.

The presented expression uses features of the main internal SWRL libraries for common operations involving constraints, lists, string, comparison, boolean operations, URIs and date/time. Furthermore, SWRL built-ins are especially useful as they allow for special definition of specific domains, arbitrary methods called “user-defined built-ins” – an important feature of the SWRL formalism that allows users to define new internal libraries to special tasks. Key built-in functions such as math operators, and string and date operations were defined in the original SWRL specification (Horrocks *et al*., 2005). Protégé allows implementation through the SWRLTab plug-in, where the rules are grouped with the ontology in a single OWL file, which can be imported into any domain ontology.

In view of the use of knowledge graphs, the possibility of consultations on implicit and inferred knowledge stands out. Within this context, Coppens *et al*. (2013) propose an extension to the SPARQL query language to support remote reasoning, in which the data consumer can define the inference rules. Binding regimes supported by a SPARQL endpoint provider are applied with an additional reasoning step using the inference rules defined by the data consumer. At the same time, this solution offers possibilities to resolve interoperability issues when querying remote SPARQL endpoints, which can support federated query frameworks. These frameworks can be extended to provide remote and distributed reasoning (Coppens *et al*., 2013).

SPARQL is a generic query language that allows federated queries between heterogeneous and linked data sources (Gearon; Passant; Polleres, 2013). Compared to specific query languages, SPARQL is especially applicable to scenarios where data from multiple sources is needed (Krijnen; Beetz, 2018).

As a demonstration, rule present in Figure 2 corresponds to converting rule (1) from SWRL to SPARQL. The INSERT clause will create a new triple in the database referring to the score for the green area criterion.

Figure 2 - SPARQL rule

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

INSERT {?w <http://www.semanticweb.org/SEBIM#Score> "1"^^xsd:decimal }
WHERE {
 ?x rdf:type sebim:ifcBuilding.
 ?y rdf:type sebim:ifcSpace.
 ?y sebim:ifchasStyleArea ?z.
 ?y sebim:Area ?a.
 ?x sebim:Area ?t.
FILTER (?a/?t > 0.3).
}
```

Source: Created by the authors.

The triples retrieved in a SPARQL query are similar to the reasoning, since the presence of the entire set of triples is necessary, with the exploration of the definition of conditions and the execution of filters, to make the operation precise and to return a subset, which are the inferred triples. The main mechanism for computing query results in SPARQL is graph matching which is also called simple linking. The RDF triples in the queried data and query pattern are interpreted as nodes and edges of the directed graphs, and the resulting query graph corresponds to the data graph generated from the variables. World Wide Web Consortium (W3C) standards, including RDF and OWL, provide semantic interpretations for RDF graphs that allow inferring additional RDF statements from explicitly given statements. Many applications that rely on these semantics require a query language like SPARQL, but to use SPARQL, basic matching must be defined using binding semantic relationships rather than explicitly created structures. There are different possible ways to define a basic graph pattern matching extension for a linking relationship (Coppens *et al*., 2013). Hawke *et al.* (2011) specify binding regimes to define not only the binding relation to use, but also which queries and graphs are well-formed for the regime, since there are different meaningful ways to use the same binding relation. In this way, it is possible to define a solution in which the end user defines the ontology or the set of rules used to reason about the graph. Reasoning happens after the basic matching of graph patterns, to have a post-reasoning step, which complements the pre-reasoning step offered by the linking regime (Hawke *et al.*, 2011).

5 Domain of the developed artifact

Within the construction industry, Building Information Modeling (BIM) technology has received increasing attention. The computational tools, which meet the concepts of this technology, interoperate using data in the Industry Foundation Classes (IFC) format, as well as exporting tabular data for information analysis. With BIM technology, an accurate virtual model of a built environment contains the geometry and data needed to support activities throughout the built environment's lifecycle. This technology provides the foundation for new design and construction capabilities, as well as driving changes in roles and team relationships. Thus, BIM presents itself as an evolution in the design process, as it allows new possibilities to visualize, represent, process, use and retrieve information (Sacks *et al.*, 2018).

To obtain the green seal, it is essential to evaluate the construction project. LEED certification, conceived and granted by the US Green Building Council (USGBC), is a system that classifies projects into four levels (certified, silver, gold and platinum) and is widely used in design, construction, operation and maintenance of buildings (USGBC, 2020).

BIM-supported LEED certification can save substantial time and resources. With the integration between BIM and sustainable environments, opportunities arise to develop practical tools to certify green buildings, improving compatibility and ease of use by experts (Wong; Zhou, 2015).

Evaluating a building to certify its sustainability requires smarter methods. The traditional assessment, performed manually, is laborious and prone to errors (Jiang; Wang; Wu, 2018). Reconciled with the semantic limitations of the IFC format, the quest to apply semantic technologies aims to meet the best practices for integrating, retrieving information, and generating new knowledge.

The domain analysis carried out in this study involves understanding the epistemological consensus between the BIM classification system and the rules for meeting the criteria of a LEED certification. Knowledge of the subject is essential to present applied research, reporting the development of a knowledge organization system (Smiraglia, 2015).

It is possible to find works that investigated multidisciplinary interdependencies in green building projects, focusing on computational optimization and collaboration in project design (Azhar *et al*., 2011; Geyer, 2012; Hong; Lee; Yu, 2019).

Recent studies related to the use of ontologies to classify materials and perform automated analyzes of building characteristics for green certification are also identified Zhang *et al.* (2019) use ontology, with SWRL rules, to infer the real-time score of green projects in a computerized media 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 acquisition of knowledge among the different project participants. It can also be seen in Xu *et al.* (2019) that using logical inferences to evaluate a criterion allows researching aspects that need improvements in the building and helping project managers to use BIM data.

These surveys listed above do not, however, address the integration with open data for analysis of the environment, nor the application of inferences for quantitative analysis, key points of this project. Our proposal seeks to model more complex relationships than those found in the works above to perform a more comprehensive analysis of LEED certification. Furthermore, previous studies were not based on a grounding ontology.

6 Methodology

This is applied research, as it generates knowledge to automate the evaluation of sustainable buildings. For the domain analysis, knowledge from previous studies and understanding of the criteria for LEED certification were used. As for the approach, this research is descriptive, prescriptive and qualitative.

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

The development of the artifact to automate the LEED certification involves elaborating a simplified ontology, based on UFO (Guizzardi, 2005) and based on the ifcOWL ontology to store BIM data and the criteria for certification. Datasets were used as a model for preparing the ontology. Because it involves the elaboration of a simplified ontology, the basis for method 101 stands out (Noy; McGuinness, 2001).

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

7 Inferences for sustainable projects

The development of the ontology is important for the conception of a workflow that starts with a file in IFC format of a BIM project converted into OWL. To integrate external data, open data, in Comma-Separated Values (CSV) format, are obtained, mapped to ontology concepts and converted into RDF graphs. The datasets are inserted into the Parliament triplestore to carry out queries and inferences in SPARQL. A final query is performed to present the result of the inferences about the scores obtained for the certification.

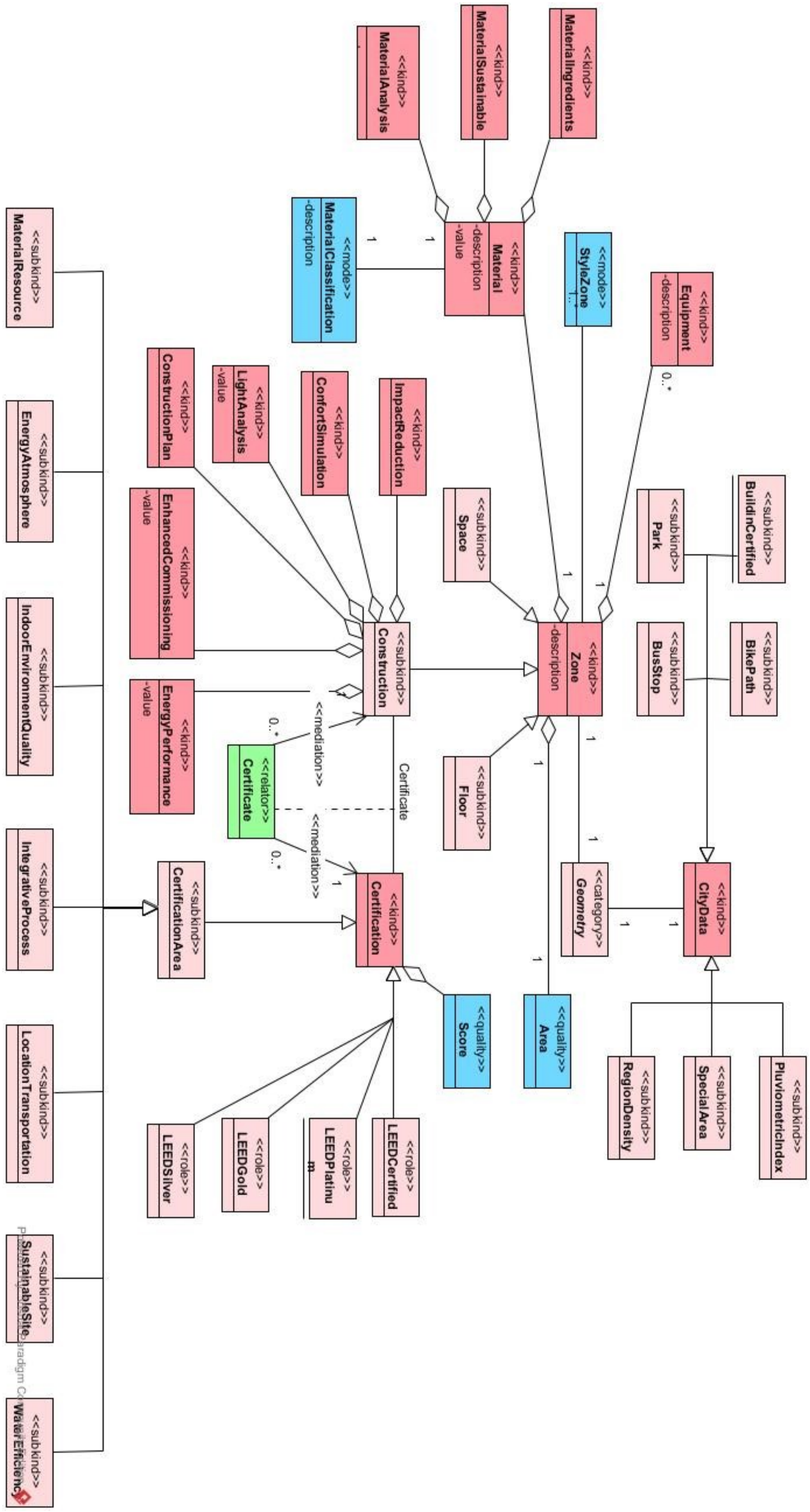
The Semantic BIM (SEBIM) ontology was created, reusing the ifcOWL (Pauwels; Terkaj, 2016), SimpleBIM (Pauwels; Roxin, 2016), BOT (Rasmussen *et al*., 2017) and BIMSO (Niknam; Karshenas, 2017) ontologies, to semantically organize the BIM data. Competency questions were based on LEED criteria on the use of materials, use of spaces, among others. Such criteria, in the context of ontology, support the inferences that assess the state of the building.

The conceptual modeling (Figure 3) was carried out in Visual Paradigm with the OntoUML plugin that allows the linking of stereotypes of the UFO foundation ontology. After the validation runs of the diagram, the conversion to gUFO was performed and new verifications were performed in Protégé 5.5. The components of the construction project were chosen to meet the analysis of the criteria chosen for prototype trials. Because it is a simplification of the IFC format, new classes can be included according to other analyses, following the same method.

A built-environment Zone is a generalized kind in Building, Space, and Floor subkinds. This class organization makes it possible to interconnect floors and other spaces, such as an apartment consisting of bedrooms, bathroom, living room and kitchen. The zone style, stereotyped as mode, is an important classification to organize and identify relevant zones for environmental assessment, such as green areas, parking for green vehicles, among others.

As many analyzes involve using materials and equipment, kind classes were considered to abstract this data, and the inclusion relationship in a Zone. The use of equipment involves the measurement of data and the internal location of sensors, air conditioning equipment, photovoltaic panels, among others. The relationship between zones and materials seeks to identify compliance with sustainability requirements. The materials are related to their composition of raw materials and classification, according to the type of use.

Figure 3 - Conceptual modeling



Source: Created by the authors.

The attributes correspond to the storage of numerical values, such as the obtained score, alphanumeric descriptions, dates and geospatial data. The latter corresponds to the import of the GeoSPARQL ontology to link open data about the surroundings of the built environment.

Publishing open data on the Web is an action of city halls that promotes the smart city (Isotani; Bittencourt, 2015). On the Belo Horizonte City Hall website, it is possible to find sets of data on services and locations, but still without using the potential of the Semantic Web. The amount of data is too much for human interpretation and makes linking data costly and error prone.

Open data on means of transport and parks are used to evaluate criteria of the Location and Transport class. These data have geometric coordinates of areas and points of the city, following the standardization of the Open Geospatial Consortium (OGC), which defines a vocabulary to represent geospatial data. This data format is necessary to perform queries, applying the GeoSPARQL functions, which return triples according to proximities and relations between coordinates (Perry; Herring, 2011).

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

Given the options for converting tabular data to RDF (Ding *et al.,* 2011; Tandy; Herman; Kellogg, 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 accurate and machine-readable manner (Rashid *et al*., 2020).

The RDF graph generated by executing the ssd2rdf script contains the formalization of the data and favors integration for queries about city locations using spatial data. Thus, data about bike lanes, parks and bus stops are converted into RDF graphs.

The inferences of this work were performed in SPARQL, applying an INSERT to insert the inferred triple in the data set, directing the score to the class corresponding to the criterion. Query in Figure 4 displays the inference made to evaluate and score the installation for bicycles. This criterion corresponds to the quantitative analysis of spaces for bicycles according to the number of units (apartments) in the built environment.

Figure 4 - Inference to installation for bicycles

```
PREFIX sebim: <http://www.semanticweb.org/SEBIM#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX gufo: <http://purl.org/nemo/gufo#>

INSERT {?w <http://www.semanticweb.org/SEBIM#score> “1”^^xsd:decimal}
WHERE {
   ?bike rdf:type sebim:Material.
   sebim:RackBike sebim:isMaterialClassificationOf ?bike.
   ?v gufo:inheresIn ?bike.
   ?v gufo:hasQualityValue ?n.
   {SELECT (count(?s) as ?place) WHERE {?unidade rdf:type sebim:Zone.
                                        sebim:Unit rdf:isStyleZone ?unidade.
                                        ?x gufo:inheresIn ?unidade.}  }
}
GROUP BY ?bike ?v ?n ?place
HAVING (?place <= ?n)
```

Source: Created by the authors.

Queries for analysis of the environment require using GeoSPARQL functions to capture the geospatial coordinates of the environment using Well-Known-Text (WKT) representations of polygons, 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; Herring, 2011). Query in Figure 5 displays the inference that analyzes the proximity of bus stops and creates the triple with the score. This criterion analyzes a minimum distance to a bus stop from a built environment.

Figure 5 - Inference to analyze the distance

```
PREFIX sebim: <http://www.semanticweb.org/SEBIM#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX gufo: <http://purl.org/nemo/gufo#>
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 {?w <http://www.semanticweb.org/SEBIM#Score> "1"^^xsd:decimal }
WHERE {
  ?bus rdf:type sebim:BusStop.
  ?local gufo:inheresIn ?bus.
  ?local sebim:asWkt ?g.
  ?build rdf:type sebim:Construction.
  ?locbuild gufo:inheresIn ?build.
  ?locbuild sebim:asWkt ?gb.
   FILTER (geof:distance(?locbuild, ?gb, uom:metre) <= 800)
}
```

Source: Created by the authors.

The process of inserting new triples via SPARQL automates the evaluation of LEED certification criteria. The queries were grouped into data internal to the BIM project, data from simulations and data external to BIM. The initial experimentation was based on the creation of fictitious instances and attributes to validate the inferences of the criteria for evaluating a constructive project. As the LEED provides criteria by type of construction (new construction, maintenance, hospitals and warehouses), the evaluation of a new construction was adopted certification. Given the complexity of the information needed for all criteria, those that represented a different type of extraction were selected. Thus, out of the 57 criteria for a new build, 7 criteria were implemented, as explained in the next section.

8 Developed artifact

To experiment with the methods employed in the artifact, an academic project designed by students of the architecture course at Fundação Mineira de Educação e Cultura (FUMEC University) was chosen. The modeling was created in

Autodesk Revit software and contains 27-story glass buildings. Revit enables three-dimensional modeling in BIM, supports design, enables simulations of building analysis, and exports data for integration with other tools (Autodesk, 2022).

In the first attempt to export data to integrate them with other external data, it was decided to generate the IFC file of the project in Revit. With the file in hand, we tried to convert the IFC format into an RDF graph. The tools used were IFCtoRDF (Pauwels; Terkaj, 2016) to convert data using the ifcOWL ontology, and IFCtoLDB (Bonduel *et al.*, 2018) to convert data according to the BOT ontology. Both convert data from the IFC file to their respective ontologies, annotating them. However, due to the limitations of the tools, 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 insufficiency of the result obtained, the second attempt to export data involved the generation of tabular files in Revit, instead of exporting in IFC format. This attempt proved to be more effective, since it allowed the use of the sdd2rdf tool, linking the data to the SEBIM ontology, developed in this work.

Data from the BIM project to be evaluated as sustainable and data from the city of Belo Horizonte were semantically annotated using the semantic dictionary and converted seamlessly to the RDF standard (with the sdd2rdf tool) and, finally, inserted into the triplestore. The execution of inferences via SPARQL queries generated new triples containing the final score to verify whether the LEED certification criteria were met. The criteria analyzed in the project in question are presented in Table 1.

Table 1 - Score of the criteria used in the artifact

LEED Category	Criterion	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	Internal lighting	2	2

Souce: Created by the authors.

It can be concluded that only two criteria did not obtain the maximum score. It should also be noted that criteria with a maximum score above one may have scoring scales defined in the LEED manual.

During the research, the difficulties identified (Table 2) in the traditional (manual) process of organizing information for LEED certification are used to validate the relevance of applying the semantic technologies employed in the proposed prototype.

Table 2 - Solutions to the difficulties encountered

Difficulty / Research challenge	Solution
Extension of the BIM technology IFC format	Conversion from IFC to simplified ontology (SEBIM)
Integration with external data	Semantic annotation of tabular data
Organization of information	Knowledge graphs
Data analysis (LEED criteria)	SPARQL Inferences
Data recovery	SPARQL Queries

Source: Created by the authors.

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

In the first version of the artifact, 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 modeling of complex rules, obtaining quantities in the A-Box (as presented in Query 3) and ease of maintenance. Thus, the implementation of the rule-based reasoning model used SPARQL queries, generating new triples for inferences about the score of each criterion.

As the evaluation process covers the phases of a constructive project and requires reviewing and updating the project to meet the criteria, it is important to consider the historical organization of the information. In this way, uploading a new version of a BIM project allows comparing results and analyzing the evolution in meeting the criteria for certification.

Traditional assessment to assess the sustainability of a built environment involves surveying and organizing a variety of documents and data from a variety of sources. This task is still performed manually and prone to errors (Jiang; Wang; Wu, 2018). Thus, the quest to apply semantic technologies aims to meet the best practices for integrating, retrieving information, and generating new knowledge. The workflow proposed by this research allows implementing a friendly solution for specialists who are not familiar with semantic technologies. The prototype can be implemented in frameworks such as JenaSemanticWeb and VirtuosoOpenSource, which support GeoSPARQL functions and resources for inferences.

9 Conclusions

A discussion is presented on the organization of knowledge and Semantic Web technologies to make inferences to support decision-making. The paradigms for the use of reasoning engines involve semantic analysis, application of rules and queries in knowledge graphs. The use of SPARQL queries is highlighted as a more efficient solution, but the selection of the most appropriate modeling paradigm depends on the application's domain and needs.

The use of UFO metaphysical constraints to support the ontology allows analyzing and criticizing the quality of the elaborated conceptual model. In the philosophically realistic hypothesis that the real world is perceived in the same way by people, greater interoperability of the model is expected, since the developed ontology favors greater compatibility with other models.

In the context of domain analysis, it is highlighted that the focus of Information Science is on the domain of knowledge to elaborate databases that are used in Computer Science to computerize access to human knowledge. The knowledge organization process depends on the use of tools. Therefore, it is necessary to use techniques that preserve the relevance of the human expert, regardless of the change in access tools. This research reinforces that this is an important aspect, since it involves the use of technologies.

To demonstrate the use of the presented technologies, an artifact was developed to integrate data and generate inferences with Semantic Web technologies to automate the LEED certification assessment. BIM technology, a reference in the civil construction industry, is used as the basis of a simplified ontology to integrate data. Open data, in tabular format, are semantically annotated using a semantic data dictionary. With the data integrated into a triplestore, SPARQL queries provided inferences about meeting the criteria for certification. The results suggest that this solution promotes a semantic extension of the constructive elements in BIM, facilitates integration with other knowledge bases, and organizes the data for the retrieval of information and for the application of inferences to generate new knowledge.

Experimenting with the artifact was limited to an academic project with less information than a real professional project. Due to copyright restrictions, this work did not obtain designs of a real building. Thus, as future work, the development of a tool for commercial purposes may enable its use in real cases.

The presented solution allows the use of more efficient and persuasive methods for project management that facilitate decision making, instead of a manual approach based only on the experience of a professional. The built ontology can be reused and expanded to include other project needs. Thus, it is expected that future similar applications will collaborate in the evaluation of constructive projects such as the construction authorization in the city halls, among other evaluations that require complex analyses.

References

ALMEIDA, J. P. A. *et al*. gUFO: a lightweight implementation of the Unified Foundational Ontology (UFO). **GitHub**, San Francisco, 2019.

AUTODESK. **Revit Family**. Autodesk, Inc., California, 26 Oct. 2022.

AZHAR, S. *et al*. Building information modeling for sustainable design and LEED® rating analysis. **Automation in Construction**, Netherlands, v. 20, n. 2,

E-ISSN 1808-5245

p. 217-224, 2011. Available in: https://doi.org/10.1016/j.autcon.2010.09.019. Accessed on: 15 May. 2021.

BASSILIADES, N. SWRL2SPIN: converting SWRL to SPIN. **CEUR Workshop Proceedings**, Germany, v. 2204, p. 1-15, 2018.

BAX, M. P. Design science: filosofia da pesquisa em ciência da informação e tecnologia. **Ciência da Informação**, Brasília, v. 42, n. 2, p. 298-312, 2013. Available in: https://doi.org/10.18225/ci.inf.v42i2.1388. Accessed on: 16 Oct. 2021.

BONDUEL, M. *et al*. The IFC to linked building data converter - current status. **CEUR Workshop Proceedings**, Germany, v. 2159, p. 34-43, 2018.

COPPENS, S. *et al*. Reasoning over SPARQL. **CEUR Workshop Proceedings**, Germany, v. 996, p. 1-5, 2013.

DING, L. *et al*. TWC LOGD: a portal for linked open government data ecosystems. **Journal of Web Semantics**, Netherlands, v. 9, n. 3, p. 325-333, 2011. Available in: https://doi.org/10.1016/j.websem.2011.06.002. Accessed on: 28 Sept. 2021.

FALBO, R. de A. SABiO: systematic approach for building ontologies. **CEUR Workshop Proceedings**, Germany, v. 1301, p. 1-14, 2014.

GEARON, P.; PASSANT, A.; POLLERES, A. (ed.) SPARQL 1.1 update. **W3C**, [*s.l.*], 21 Mar. 2013. W3C Recommendation.

GEYER, P. Advanced engineering informatics systems modelling for sustainable building design. **Advanced Engineering Informatics**, United Kingdom, v. 26, n. 4, p. 656-668, 2012. Available in: https://doi.org/10.1016/j.aei.2012.04.005. Accessed on: 15 May. 2021.

GUIZZARDI, G. **Ontological foundations for structural conceptual models**. University of Twente: Telematica Institut/CTIT, 2005.

GUIZZARDI, G. *et al*. Ontologias de fundamentação, modelagem conceitual e interoperabilidade semântica. **CEUR Workshop Proceedings**, Germany, v. 728, p. 1-10, 2011.

HAWKE, S. *et al*. SPARQL 1.1 entailment regimes. **W3C**, [*s.l.*],12 May 2011.

HJØRLAND, B. Domain analysis in information science: eleven approaches: traditional as well as innovative. **Journal of Documentation**, United Kingdom, v. 58, n. 4, p. 422-462, 2002. Available in: https://doi.org/10.1108/00220410210431136. Accessed on: 17 Feb. 2021.

HJØRLAND, B. Reviews of concepts in knowledge organization. **Knowledge Organization**, Germany, v. 44, n. 5, p. 349-379, 2017. Available in: https://doi.org/10.5771/0943-7444-2017-1-55. Accessed on: 17 Feb. 2021.

HJØRLAND, B. What is Knowledge Organization (KO)? **Knowledge Organization**, Germany, v. 35, n. 2/3, p. 86-101, 2008. Available in 10.5771/0943-7444-2008-2-3-86. Accessed on: 17 Feb. 2021.

HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in information science: domain-analysis. **Journal of the American Society for Information Science**, Leesburg, v. 46, n. 6, p. 400-425, 1995. Available in: https://doi.org/10.1002/(SICI)1097-4571(199507)46:6<400::AID-ASI2>3.0.CO;2-Y. Accessed on: 17 Feb. 2021.

HONG, S.-H.; LEE, S.-K.; YU, J.-H. Automated management of green building material information using web crawling and ontology. **Automation in Construction**, Netherlands, v. 102, p. 230-244, 2019. Available in: https://doi.org/10.1016/j.autcon.2019.01.015. Accessed on: 15 May. 2021.

HORROCKS, I. *et al*. OWL rules: a proposal and prototype implementation. **Web Semantics**, Netherlands, v. 3, n. 1, p. 23-40, 2005. Available in: https://doi.org/10.1016/j.websem.2005.05.003. Accessed on: 08 June 2021.

ISOTANI, S.; BITTENCOURT, I. I. **Dados abertos conectados**. São Paulo: Novatec, 2015.

JIANG, S.; WANG, N.; WU, J. Combining BIM and ontology to facilitate intelligent green building evaluation. **Journal of Computing in Civil Engineering**, Reston, v. 32, n. 5, p. 1-15, 2018. Available in: https://doi.org/10.1061/(ASCE)CP.1943-5487.0000786. Accessed on: 02 May. 2021.

KRIJNEN, T.; BEETZ, J. A. SPARQL query engine for binary-formatted IFC building models. **Automation in Construction**, Netherlands, v. 95, p. 46-63, 2018. Avaliable in: https://doi.org/10.1016/j.autcon.2018.07.014. Accessed on: 21 Aug. 2021.

LAWAN, A.; RAKIB, A. The semantic web rule language expressiveness extensions-a survey. **arXiv**, Ithaca, Mar. 2019. Available in: https://doi.org/10.48550/arXiv.1903.11723. Accessed on: 01 June 2021.

NIKNAM, M.; KARSHENAS, S. A shared ontology approach to semantic representation of BIM data. **Automation in Construction**, Netherlands, v. 80, p. 22-36, 2017. Available in: https://doi.org/10.1016/j.autcon.2017.03.013. Accessed on: 15 May. 2021.

NOY, N. F.; MCGUINNESS, D. L. Ontology Development 101: a guide to creating your first ontology. **Knowledge Systems Laboratory**, Stanford, v. 32, 2001.

PAUWELS, P; ROXIN, A. SimpleBIM: from full ifcOWL graphs to simplified building graphs. *In:* European Conference on Product and Process Modelling, 11th, 2016. **Proceedings** […]. [*S.l.*]: ECPPM, 2016.

PAUWELS, P.; TERKAJ, W. EXPRESS to OWL for construction industry: towards a recommendable and usable ifcOWL ontology. **Automation in Construction**, Netherlands, v. 63, p. 100-133, 2016. Available in: https://doi.org/10.1016/j.autcon.2015.12.003. Accessed on: 24 Feb. 2021.

PERRY, M.; HERRING, J. GeoSPARQL: a geographic query language for RDF data. **Open Geospatial Consortium**, Arlington, 2011.

RASHID, S. M. *et al*. The semantic data dictionary - an approach for describing and annotating data. **Data Intelligence**, Beijing, v. 2, n. 4, p. 443-486, 2020. Disponível em: https://doi.org/10.1162/dint_a_00058. Accessed on: 24 Feb. 2021.

RASMUSSEN, M. H. *et al*. Proposing a central AEC ontology that allows for domain specific extensions. *In*: Joint Conference on Computing in Construction, 3., Creta, 2017**. Proceedings** […]. Edinburg: Heriot-Watt University, 2017. p. 237-244.

SACKS, R. *et al*. **BIM handbook: a guide to building information modeling for owners, managers, designers, engineers, and contractors**. 3. ed. New Jersey: John Wiley & Sons, 2018.

SMIRAGLIA, R. P. **Domain analysis for knowledge organization: tools for ontology extraction**. Oxfordshire: Chandos Publishing, 2015.

SUCHÁNEK, M. OntoUML specification. **OntoUMLReadthedocs**, [*s.l.*], p. 1-123, 11 Sept. 2022.

TANDY, J.; HERMAN, I.; KELLOGG, G. (ed.). Generating RDF from tabular data on the web. **W3C**, [*s.l.*], 17 Dec. 2015. W3C Recommendation.

TERKAJ, W.; SOJIC, A. Ontology-based representation of IFC EXPRESS rules: an enhancement of the ifcOWL ontology. **Automation in Construction**, Netherlands, v. 57, p. 188-201, 2015. Avaliable in: https://doi.org/10.1016/j.autcon.2015.04.010. Accessed on: 22 Feb. 2021.

US GREEN BUILDING COUNCIL (USGBC). LEED v4.1: building design and construction. **US Green Building Council**, Washington, p. 259, 2020.

WONG, J. K. W.; ZHOU, J. Enhancing environmental sustainability over building life cycles through green BIM: a review. **Automation in Construction**, Netherlands, v. 57, p. 156-165, 2015. Avaliable in: https://doi.org/10.1016/j.autcon.2015.06.003. Accessed on: 24 Feb. 2021.

XU, Z. *et al*. Study on the evaluation method of green construction based on ontology and BIM. **Advances in Civil Engineering**, London, v. 2019, p. 1-20, 2019. Avaliable in: https://doi.org/10.1155/2019/5650463. Accessed on: 23 Feb. 2021.

ZHANG, D. *et al*. A semantic and social approach for real-time green building rating in BIM-based design. **Sustainability**, Basel, v. 11, n. 14, p. 1-16, 2019. Avaliable in: https://doi.org/10.3390/su11143973. Accessed on: 24 Feb. 2021.

Aplicação de inferências em uma modelagem ontológica

Resumo: A modelagem de dados apoiada em ontologias favorece à organização do conhecimento e à utilização de máquinas de inferências para obter novos conhecimentos sobre os dados e apoiar a tomada de decisões de especialistas. No entanto, é preciso discutir as técnicas para promover as inferências e explorar a semântica entre os dados. O problema de pesquisa é verificar como o emprego de tecnologias semânticas representaria uma alternativa vantajosa, em relação ao estado da arte, para analisar, sistematizar e organizar as informações de um domínio, acelerando um processo para tomada de decisões. Apresentam-se os conceitos por trás do desenvolvimento de um artefato que emprega tecnologias semânticas e uma ontologia de domínio para integrar dados de projetos *Building Information Modeling* e outros dados tabulares externos ao projeto para automatizar as tarefas da certificação *Leadership in Energy and Environmental Design*. Para demonstrar o artefato, dados são anotados semanticamente e um grafo de conhecimento é gerado no formato *Resource Description Framework*. Uma vez integrados no grafo, são aplicadas inferências para avaliar os critérios da certificação por meio de consultas. Os resultados sugerem que a aplicação destas tecnologias promove a extensão semântica, facilita a integração com outras bases de conhecimento e organiza conceitualmente os dados para melhor recuperar informações e gerar novos conhecimentos.

Palavras-chave: inferência; análise de domínio; interoperabilidade semântica

Recebido: 18/01/2023
Aceito: 10/07/2023

Declaração de autoria:

Concepção e elaboração do estudo: Cristiano Geraldo Teixeira Silva, Marcello Peixoto Bax.

Coleta de dados: Cristiano Geraldo Teixeira Silva.
Análise e interpretação de dados: Marcello Peixoto Bax, Cristiano Geraldo Teixeira Silva.
Redação: Marcello Peixoto Bax, Cristiano Geraldo Teixeira Silva.
Revisão crítica do manuscrito: Marcello Peixoto Bax

Como citar:

BAX, Marcello Peixoto; SILVA, Cristiano Geraldo Teixeira. Application of inferences in ontological modeling. **Em Questão**, Porto Alegre, v. 29, e-129483, 2023. DOI: https://doi.org/10.1590/1808-5245.29.129483.

Parecer de avaliação
O parecer de avaliação deste artigo está disponível em:

https://seer.ufrgs.br/index.php/EmQuestao/article/view/129483/89934