A Methodological Approach to Compare Ontologies: Proposal and Application for SLAM Ontologies

Yudith Cardinale
Universidad Católica San Pablo, Arequipa, Perú
Universidad Simón Bolívar, Caracas, Venezuela
ycardinale@usb.ve

María A. Cornejo-Lupa
Universidad Católica San Pablo
Arequipa, Perú
maria.cornejo.lupa@ucsp.edu.pe

Regina Ticona-Herrera
Universidad Católica San Pablo
Arequipa, Perú
rtherrera@ucsp.edu.pe

Dennis Barrios-Aranibar
Universidad Católica San Pablo
Arequipa, Perú
dbarrios@ucsp.edu.pe

ABSTRACT

Representation of the knowledge related to any domain with flexible and well-defined models, such as ontologies, provides the base to develop efficient and interoperable solutions. Hence, a proliferation of ontologies in many domains is unleashed. It is necessary to define how to compare such ontologies to decide which one is the most suitable for specific needs of users/developers. Since the emerging developing of ontologies, several studies have proposed criteria to evaluate them. Nevertheless, there is still a lack of practical and reproducible guidelines to drive a comparative evaluation of ontologies as a systematic process. In this paper, we propose a methodological process to qualitatively and quantitatively compare ontologies at *Lexical, Structural*, and *Domain Knowledge* levels, considering *Correctness* and *Quality* perspectives. Since the evaluation methods of our proposal are based in a *golden-standard*, it can be *customized* to compare ontologies in any domain. To show the suitability of our proposal, we apply our methodological approach to conduct a comparative study of ontologies in the robotic domain, in particularly for the Simultaneous Localization and Mapping (SLAM) problem. With this study case, we demonstrate that with this methodological comparative process, we are able to identify the strengths and weaknesses of ontologies, as well as the gaps still needed to fill in the target domain (SLAM for our study case).

CCS CONCEPTS

• **Information Systems** → **Ontologies**; • **Computer systems organization** → *Robotic autonomy*.

KEYWORDS

Ontology; Ontologies Evaluation; SLAM; Autonomous and Mobile Robots.

iiWAS '20, November 30-December 2, 2020, Chiang Mai, Thailand

ACM ISBN 978-1-4503-8922-8/20/11…$15.00
https://doi.org/10.1145/3428757.3429091

ACM Reference Format:
Yudith Cardinale, María A. Cornejo-Lupa, Regina Ticona-Herrera, and Dennis Barrios-Aranibar. 2020. A Methodological Approach to Compare Ontologies: Proposal and Application for SLAM Ontologies. In *The 22nd International Conference on Information Integration and Web-based Applications & Services (iiWAS '20), November 30-December 2, 2020, Chiang Mai, Thailand.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3428757.3429091

1 INTRODUCTION

The need of developing systems and applications that require access to real-world knowledge is in permanent growing. Semantic Web technologies, such as ontologies, seem to be clear solutions to offer standard and well-defined models for capturing the knowledge of any domain, from which to take their organizational and relational capacity [5, 33]. Domain ontologies, which formalize the terms used in a discipline, make possible the interoperability among systems. Hence, many works in specific domains are leveraging on such as formal knowledge representation.

However, there is no perfect way to develop an ontology, there will always be trade-offs, for example, between the value of high expressiveness and the cost of computation [37]. As a consequence, a proliferation of ontologies in the same domain is unleashed. Therefore, it is necessary to define how to evaluate existing ontologies to decide which one is the most appropriate in a specific domain, for specific research needs. Since the emerging developing of ontologies, several studies have proposed sets of criteria to evaluate them [1, 3, 6, 13, 18, 42]. Nevertheless, there is still a lack of practical and reproducible guidelines or methodologies to drive a comparative evaluation of ontologies as a systematic process.

In order to overcome these limitations, in this paper, we propose a methodological process to qualitatively and quantitatively compare ontologies at *Lexical, Structural*, and *Domain Knowledge Modeling* levels, considering *Correctness* and *Quality* perspectives. Although we do not pretend to offer an exhaustive list of metrics to measure each level, our methodological process represents a systematic guideline to perform comparison of ontologies in the same domain. Besides the support in the comparative evaluation process, our proposed methodological process is also aimed at identifying the gaps in the knowledge representation of a specific domain. Our methodological approach can be *customized* to compare ontologies in any domain, since the evaluation methods are based in a *golden-standard*. By defining the *golden-standard*, either as an ontology, as a corpus in the domain, or as the expert knowledge, the ontologies to evaluate are delimited.

In order to show the suitability of our proposal, we apply our methodological process to compare most recent and relevant ontologies in the robotic domain, in particularly for autonomous robots. The main tasks of autonomous robots are mapping an environment and localize themselves, which conform the Simultaneous Localization and Mapping (SLAM) problem [44]. SLAM deals with the necessity of building a map of an environment (i.e., navigation), while simultaneously determining the location of the robot within this map (i.e., positioning). Due to the evolution of mobile technologies and sensors, the complexity of the behaviors that robots are expected to perform is growing. Naturally, this trend involves the use of increasingly complex knowledge as well as the need for multi-robots and humans-robots collaboration. Thus, many researchers have focused on formally representing, with ontologies, some aspects related to the knowledge managed in SLAM. In a previous study, we surveyed the most popular and recent SLAM ontologies, classifying them according to the type of knowledge modeled [10]. We demonstrate that with this methodological comparative evaluation process, we are able to identify the strengths and weaknesses of ontologies, as well as the gaps not yet covered in the target domain (SLAM ontologies, for our study case).

In summary, the contribution of this work in three-fold: (i) a *customizable* methodological process to conduct a comparative evaluation of ontologies in any domain, which is not an exhaustive list of metrics, but a systematic guideline; (ii) a comparative evaluation of SLAM ontologies, based on the proposed methodological process; and (iii) a demonstration of the suitability of the methodological evaluation process, from which we pin down that although there exist several ontologies to represent SLAM knowledge, there is a lack of a standard arrangement and generic ontology covering the full aspects of such as knowledge.

This paper is structured as follows. Section 2 presents an analysis of evaluation methods for comparison of ontologies. Section 3 draws up the approach of this work. Section 4 summarizes the categorization of the knowledge needed to solve the SLAM problem in order to use it as a *golden-standard*. Section 5 presents a study case based on the comparison of most recent and relevant SLAM ontologies. Section 6 sketch out the conclusions of this work.

2 RELATED WORK

The increasing use of ontologies in any domain has inspired a lot of research in metrics to evaluate them in different aspects, both qualitative and quantitative. Since their emerging developing, several studies have surveyed works proposing such as metrics [1, 3, 6, 13, 18, 42] and works focused on evaluating their correct construction [20, 27]. However, none of these studies propose a general and methodological process to conduct a comparative evaluation of ontologies in the same domain. Moreover, they present only lists of criteria without classify them.

Regarding SLAM ontologies, we only find studies surveying the use of ontologies in the robotic domain [11, 48], but they do not consider evaluation or comparative issues, beyond of presenting robotic ontologies; and a work to evaluate performance of robotic ontologies, called PERK [38]. In [38], authors discuss the knowledge related to robotics, in a general way. Afterward, a set of qualitative criteria usually considered to evaluate ontologies are described and five previous studies (from 1995 to 2007) proposing metrics to evaluate ontologies are compared in terms of this set of criteria.

More related to our research are the studies that propose a classification of criteria in order to evaluate ontologies at several levels [3, 15, 18, 21, 24, 29, 32]. On each level, different **evaluation methods** can be applied to measure the corresponding criteria [3, 18], such as: (i) *golden-standard* based evaluation, in which ontologies are compared against a reference ontology, according to string matching, precision, recall measures, for example; the *golden-standard* could be another ontology, a corpus of documents of the domain, or the knowledge of domain experts [45]; (ii) *data-driven* evaluation intended to perform comparisons with datasets from the domain covered by the ontology, by applying, for example, clustering of concepts and measuring precision, recall [4]; (iii) *structure-based* evaluation, that evaluates ontologies according to their own structure and taxonomy –i.e., they do not demand extra information, as *golden-standard* and *data-driven* methods, besides the ontologies themselves; (iv) *application-based* evaluation focused on evaluating the ontologies according to their application and the results; it measures if the use of the ontology improves the performance or the quality of results of the application (a software program or a use-case scenario), in which the ontology is used [31]; and (v) *user-driven*, in which the evaluation is done by humans who concentrate their efforts on evaluating how well the ontologies meet a set of predefined criteria, standards, or requirements; this method involves evaluating the ontology through users' experiences.

Maedche and Staab [24] propose an approach to evaluate an ontology by measuring its similarity with a *golden-standard* ontology. The similarity is measured at *Lexical* (i.e., how terms are used to convey meanings) and *Conceptual* (i.e., what conceptual relations exist among terms) levels. The *Lexical* comparison level is performed by calculating a *String Matching* measure, which in turn is based on edit distance among two strings. The similarity at the *Conceptual* level is calculated in terms of the taxonomic structure (taxonomic overlap) and relationships (relation overlap), i.e., based on *structure-based* evaluation methods.

Brank et al. [3] propose to evaluate six aspects of ontologies: (i) *Lexical, Vocabulary, or Data*, considering concepts, instances, facts, and vocabulary used in the ontology; any evaluation method can be used to perform this assessment; (ii) *Hierarchy or Taxonomy* to evaluate `is-a` relations between concepts; all previously described evaluation methods can be applied for this assessment; (iii) *Other Semantic Relations* that the ontology contains, besides `is-a`, may be evaluated separately; all evaluation methods are appropriate at this assessment; (iv) *Context or Application* to measure how the use of the ontology impacts on the domain application results; *application-based* evaluation methods obviously fit in this level, as well as *user-driven* evaluation; (v) *Syntactic Level*, for ontologies that have been mostly constructed manually, it is important to evaluate their formal correctness; at this level, *golden-standard* and human assessment (*user-driven*) are the most appropriate evaluation methods; (vi) *Structure, Architecture, Design Level*, also, this level is relevant for ontologies constructed manually; it evaluates if the ontology meets certain pre-defined design principles or criteria, structural concerns, and suitability for further development; usually, *user-driven* evaluation methods are applied at this level.

Pak and Zhou [29] propose a framework to evaluate four dimensions of ontologies: (i) *Ontology Scopes*, which includes domain, conceptual, and technical scopes; (ii) *Ontology Layers*, comprised by lexical/vocabulary, structural/architectural, representation/semantic, and context/application layers; (iii) *Ontology Life Cycle*, to detect the absence of well-defined properties in specification, knowledge acquisition, conceptualization, and integration; and (iv) *Ontology Quality*, measured in terms of consistency, conciseness, completeness, and reusability.

Duque-Ramos et al. [15] propose the validation of the *Quality* of ontologies based on SQuaRE[1], a Software Engineering standard. They propose a framework, called OQuaRE[2], which considers two components: a *Quality Model* and *Quality Metrics*. The *Quality Model* considers the following categories: *Structural*, *Functional Adequacy*, *Reliability*, *Operability*, *Compatibility*, *Transferability*, and *Maintainability*. Subcategories are specified on each category to specialize the measures. *Quality Metrics* define a set of *structure-based* evaluation criteria: Lack of Cohesion in Methods (LCOMOnto), Weighted Method Count (WMCOnto), Depth of subsumption hierarchy (DITOnto), Number of Ancestor Classes (NACOnto), Number of Children (NOCOnto), Coupling between Objects (CBOOnto), Response for a class (RFCOnto), Number of properties (NOMOnto), Properties Richness (RROnto), Attribute Richness (AROnto), Relationships per class (INROnto), Class Richness (CROnto), Annotation Richness (ANOnto), and Tangledness (TMOnto).

To evaluate the *Quality* of categories (and their subcategories), they propose to score the *Quality Metrics* (from 1 to 5).

Hlomani and Stacey [18] perceive the ontology evaluation from two complementary perspectives: *Correctness* and *Quality*. From these perspectives, they propose a four-layered metric suite for ontology evaluation: (i) *Overall Ontology Evaluation* in terms of its re-use; (ii) *Perspective of Evaluation*, that might be *Correctness* or *Quality*; (iii) *Criteria to evaluate Correctness or Quality* (i.e., accuracy, adaptability, clarity, cohesion, completeness, computational efficiency, conciseness, consistency, coupling, coverage); and (iv) *Quantitative Measures*, which indicate the level of satisfaction of one criterion (e.g., precision, recall, coverage, number of terms with inconsistent meaning, number of external classes referenced, number of roots, number of leaves, number of word senses count); thus, to obtain the quantitative measures, all or a combination of the previously described evaluation methods can be applied.

On-line platforms for the calculation of metrics to evaluate ontologies, have been also proposed, such as OOPS! [32] and Ontometrics [21]. Both platforms are based on *structure-based* evaluation methods, since they only are able to consider the own ontologies and they are more focused on evaluating ontology *Correcteness*.

OOPS! evaluates four dimensions of ontologies: (i) *Human Understanding*, by considering if the ontology, for example, creates synonyms as classes, merges different concepts in the same class, misses annotations, uses different naming criteria; (ii) *Logical Consistency*, to identify, for example, wrong inverse relationships, cycles in the hierarchy, misusing `owl:allValuesFrom`, wrong transitive relationships; (iii) *Real World Representation*, by identifying missing basic information and missing disjointness; and (iv) *Modelling Issues*, related to evaluate structural aspects such as existence of synonyms as classes, use of relationship `is` instead of using `rdfs:subClassOf`, `rdf:type` or `owl:sameAs`, unconnected elements, wrong inverse relationships, cycles in the hierarchy, incorrect use of ontology elements, among others. Ontometrics is mainly focused on evaluating the *Conceptual Layer* of ontologies based on accuracy, understandability, cohesion, computational efficiency, and conciseness, according to four types of metrics: *Schema Metrics*, to evaluate the ontology structure (e.g., attributes richness, class/relation ratio, axiom/class ratio); (ii) *Graph Metrics*, to evaluate the taxonomy tree of the ontology (e.g., absolute, average, and maximal depth/breadth);

(iii) *Knowledgebase Metrics*, to assess the ontology structure and instances that populate the ontology (e.g., average population, class richness); and (iv) *Class Metrics*, focused on evaluating single classes (e.g., readability, children count, properties count).

We believe that for a fair comparison among ontologies, both *Correctness* and *Quality* perspectives [18], must be evaluated, at three levels: *Lexical*, *Structural*, and *Domain Knowledge*. Moreover, this comparative evaluation must follow a methodological process. In Table 1, we compare the studies more related to our proposal in terms of these aspects. None of these works consider both *Correctness* and *Quality* at each evaluation level. We consider that only the work presented by Pak and Zhou [29] approaches the three evaluation levels, but only from one perspective, whether *Correctness* or *Quality*, not necessarily both. These works are mainly focused on proposing some strategies to evaluate ontologies, but they do not suggest any methodological approach to perform a comparative evaluation among several ontologies on the same domain, as we propose in this work.

Table 1: Comparative summary

Reference	Perspective	Lexical	Structural	Domain Knowld.	Comparat. Approach
Maedche & Staab [24]	*Quality*	Partial	Partial	Partial	-
Brank et al. [3]	*Correctness & Quality**	Yes	Yes	Partial	-
Pak & Zhou [29]	*Correctness & Quality**	Yes	Yes	Yes	-
OQuaRE [15]	*Quality*	Partial	Partial	Partial	-
Hlomani & Stacey [18]	*Correctness & Quality**	Partial	Partial	Partial	-
OOPS! [32]	*Correctness*	-	Yes	-	-
Ontometrics [21]	*Correctness*	-	Yes	-	-

* Not at all levels

3 METHODOLOGICAL PROCESS FOR A COMPARATIVE EVALUATION OF ONTOLOGIES: OUR PROPOSAL

Our proposal consists on considering *Correctness* and *Quality* perspectives [18], at three levels on the ontology that group the majority of levels proposed by prior works [3, 15, 18, 21, 24, 29, 32]: *Lexical*, *Structural*, and *Domain Knowledge* levels. Figure 1 shows

[1]SQuaRE: SO/IEC 25000:2005 standard for Software product Quality Requirements and Evaluation
[2]http://miuras.inf.um.es/evaluation/oquare/

how these perspectives are considered on each ontology evaluation level. We also suggest the *Evaluation Methods* that can be used to approach the respective comparative evaluation. We propose to have a *golden-standard*, either as an ontology, as a corpus in the domain, or as the expert knowledge.

At *Lexical Level*, *Quality* and *Correctness* are measured by analyzing the concepts and the relative vocabulary to them and complemented with syntax revision of the entities modelled by the ontology. In agreement with other authors, the best way to evaluate this level is with the *golden-standard* technique. However, in case of not having the *golden-standard* available as an ontology, a *structure-based* comparison can be made since it depends only on the taxonomy of the ontologies to be compared/evaluated and can support the comparison against a *golden-standard* ontology. At *Structural Level*, both perspectives can be evaluated by reviewing the relationships between concepts; starting with the most common ones, like `is-a` and `has-*`, to later review more specific relationships of each ontology; and also by studying how these relationships are integrated, observing aspects such as architecture and design. As in the previous level, evaluating with the *golden-standard* is the ideal approach, but in case it is not available as an ontology, a *structure-based* evaluation is also appropriate.

At *Domain Knowledge Level*, we review the application level of the ontology and we care about the knowledge coverage, i.e., how well the *domain knowledge is covered* and how the *application results* are improved by the use of the ontology.

Using a *golden-standard* may imply the presence of experts on the domain during the comparison process; thus, *application-based* and *user-driven* evaluation methods can be monitored by these experts.

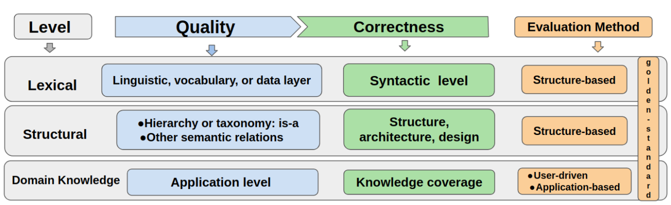

Figure 1: Perspectives, levels, and methods for a comparative study of ontologies

In the following, we illustrate how each level can be approached by showing examples of some appropriate metrics. They are not mandatory neither exhaustive. Researchers can adapt any other metric that fit on evaluating *Quality* or *Correctness* on each level.

3.1 Lexical Level

For the comparison of ontologies at *Lexical Level*, a metric of *structure-based* methods that can be used is the *Linguistic Similarity (LS)*.

It can be measured by *String Similarity* of entities, based on *lexical comparison* and a *statistic analysis* assessment [19].

For *lexical comparison*, *String Similarity* (*SS*) among names of entities (classes, properties, instances, etc.) can be applied. It can be based on edit distance, as Eq. (1) shows, where `s1` and `s2` are the considered strings (e.g., names of entities), `ed(s1,s2)` denotes the edit distance between strings `s1` and `s2`; `s1.len` and `s2.len` denote the length of the input strings `s1` and `s2`, respectively.

$$SS(s1,s2) = \frac{1}{e^{\frac{ed(s1,s2)}{|s1.len + s2.len - ed(s1,s2)|}}} \tag{1}$$

From *SS*, similarity between two ontologies can be calculated, using Eq. (2), where S_E is the number of times the *SS* is greater than ρ (a threshold defined by users); N_{O_i}, N_{O_j} are the total number of entities in ontologies O_i, O_j, respectively; and D_E is the number of duplicate entities in O_i, O_j (i.e., number of times that $SS = 1$).

$$StringSim(O_i, O_j) = \frac{S_E}{(N_{O_i} + N_{O_j}) - D_E} \tag{2}$$

Moreover, in *statistic analysis* many works propose to use the Vector Space Model (VSM) to evaluate linguistic similarity between two ontologies [19]. It consists on building an N-dimensional vector for each document, with the N terms in both documents. Vector components are the term weights assigned to the corresponding document for each of the N terms, calculated by a term weighting function. In our scenario, each ontology is represented as a document that consists of a *bag of terms* (conformed by the N terms that appear in any of the documents and each term is represented once in the vector) extracted from the entity's names, labels, and comments in the ontologies. The term weighting function to calculate each component in the N-dimensional vector for each ontology (let's called it VS_{O_i}), is defined in Eq. (3), where `t` is the number of times where one term occurs in a given document (i.e., ontology O_i); `T` denotes the maximum number of times the term appeared in a document; `D` denotes the number of documents in collection (i.e., the number of compared ontologies); and `d` denotes the number of documents where the given term occurs at least once. It is worthy to note that there are several preparing steps before calculating term weights, such as splitting words and removing stop words.

$$TermWeighting = TF * IDF; TF = \frac{t}{T}; IDF = \frac{1}{2} * (1 + log2\frac{D}{d}) \tag{3}$$

Thus, the *Document Similarity* (*DocSim*) between two ontologies is calculated by taking the vectors dot product, as Eq. (4) shows.

$$DocSim(O_i, O_j) = \frac{VS_{O_i} \cdot VS^t_{O_j}}{\|VS_{O_i}\| \|VS_{O_j}\|} \tag{4}$$

The two measures described above (*StringSim* and *DocSim*) are combined in Eq. (5) to determine *Linguistic Similarity (LS)*, where $\alpha + \beta = 1$, are parameters provided by users and they can be adjusted according to the needs of the researcher at time of application.

$$LS(O_i, O_j) = \alpha * StringSim(O_i, O_j) + \beta * DocSim(O_i, O_j) \tag{5}$$

This evaluation at *Lexical Level* can go beyond comparing words, since it is possible that two entities have the same meaning but they are denoted by two different words. Thus, the use of synonyms and Natural Language Processing (NLP), can improve this assessment.

After calculating this *Linguistic Similarity (LS)*, the idea is to elaborate a comparison matrix among the evaluated ontologies and with the *golden-standard*, if it is available as an ontology. Thus, the comparative evaluation at this level is performed with *structure-based* evaluation methods combined with a *golden-standard*. Table 2 shows an example of the comparison matrix, among two ontologies (O_1, O_2) and the *golden-standard* (*GS*). In the example in Table 2, O_1 and O_2 have 37% of *String Similarity* (*StringSim*) and 52% at *Document Similarity* (*DocSim*), which in turn result in a *Linguistic Similarity* (*LS*) of 40%. When the *golden-standard* (*GS*) is available, as in this example, each ontology is compared against it, obtaining 33% of *LS* between O_1 and *GS* and 27% of *LS* between O_2 and *GS*. This indicates that O_1 is more similar to *GS* than O_2; thus, it could be the most suitable to model the domain being analyzed.

If we do not have the *golden-standard* ontology available, we can perform the comparison among the ontologies. In this case, although this comparison does not indicate absolute *Correctness* or *Quality*, it offers a reference point to achieve a more complete analysis when we finish applying the comparison methodology proposal. Moreover, it can be complemented with other metric, such as the ones proposed in [3, 18, 24, 29] or the *Quality* metrics of OQuaRE and its score system [15]: LCOMOnto, DITOnto, NACOnto, NOCOnto, RFCOnto, NOMOnto, and AROnto.

Table 2: Comparison Matrix at Lexical Level evaluation

Pair	*StringSim*	*DocSim*	*LS* ($\alpha = \beta = 0,5$)
O_1–O_2	0,37	0,52	0,4
O_1–GS	0,23	0,45	0,34
O_2–GS	0,17	0,37	0,27

3.2 Structural Level

Since an ontology can be considered as a knowledge graph, in which an important issue is the relationships among entities, the comparative evaluation at the *Structural Level* is focused on evaluating relations. Thus, to evaluate the *Quality* and *Correctness* of the ontology structure, we suggest to review the `is-a` relationships, since they determine the hierarchy of entities modeled by the ontology. Another relationship that is always present in ontologies is `has-`$*$, where $*$ can be replaced by any name according to the specific domain that the ontology models (e.g., `has-measure`, `has-sensor`, `has-part`). It is also worthfull to review the structure of the ontologies, detecting the components that conform them – i.e., number of classes, relationships, properties, instances, and annotations – as shown in Table 3. In the example presented in Table 3, O_1 has almost twice as many classes as O_2 and therefore O_1 has more relationships than O_2. This is an indication of greater cohesion between the concepts of O_1. On the other hand, O_2 has more annotations than O_1, which has none. This is a favorable feature of O_2, because the annotations serve to better understand the ontology.

To get these measures, it is only needed to parse the owl files of ontologies, considering the OWL modeling primitives: classes refer to `owl:Class`, relationships refer to `owl:ObjectProperty`, properties refer to `owl:DatatypeProperty`, and individuals refer to `owl:Individual`. These values are indicators of the *Quality* and *Correctness* of ontologies: if the number of relations is proportional to the number of classes, it could be considered as an indicator of a well designed ontology, since it relates the classes that compose it. Otherwise, it could be an indicator of a limited modelling of knowledge of the domain. Moreover, this information can be used to calculate other metrics such as the OQuaRe criteria [15].

Table 3: Comparison Matrix by the Ontologies Structure

Ontology	Classes	Relationships			Properties	Instances	Annotations
		`is-a`	`has-*`	other			
O_1	55	23	10	0	23	0	0
O_2	27	15	10	2	0	12	12

To reinforce the evaluation of *Correctness* and *Quality* at *Structural Level*, *Graph Similarity* or *RDF Similarity* techniques can be used to measure the *Structural Similarity* between two ontologies (*StructSim*) [19, 25]. The comparison is performed by triples and the similarity comes from the accumulation of similarities of entities involved in the same role (subject, predicate, object) in the two triples being compared. That is, it is not enough for the entity to be repeated, as it is on the evaluation at *Lexical Level*, its role (subject, predicate, object) in the triplet must also be the same.

These metrics can be complemented or substituted by other *structure-based* metrics, such as the ones proposed in [3, 18, 24, 29], the *Quality* metrics of OQuaRE and its score system [15] (i.e., WMCOnto, CBOOnto, RROnto, INROnto, and TMOnto), or *Correctness* metrics of OOPS! [32] and Ontometrics [21].

Whatever the technique used to measure the *Structural Similarity*, in order to show the results of this comparative evaluation, it is proposed to use a comparison matrix. As in *Lexical Level*, the comparative evaluation at this level is performed with *structure-based* evaluation methods combined with a *golden-standard*, if it is available as an ontology.

3.3 Domain Knowledge Level

We propose to evaluate this level from two point of views: the impact of using the ontology in *application results* and the capacity of the ontology for *domain knowledge coverage*. To do so, domain experts propose questionnaires oriented to discover how the domain knowledge is represented by the ontology; as stated in the approach proposed by Bandeira et al. [1], who define a step-by-step approach to evaluate the *Quality* of an ontology through a questionnaire.

Questionnaires should be constructed based on the knowledge available in the area (*golden-standard*) and should contain a set of pertinent questions to determine if the knowledge of the domain is considered (to assess *Quality*) and if it correctly addressed (to check *Correctness*). Uses-case scenarios can be constructed by grouping related questions regarding one or more aspects of the domain knowledge and translate them into SPARQL queries. Thus, if the ontology can answer all the SPARQL queries, then the use of the ontology improves the performance or results of applications. A table summarizing the results of applying the SPARQL queries on ontologies can be elaborated, as the one shown in Table 4. If the ontology (O_i) answers to the query (Q_j), the cell is marked with a ✓. The last column determines the percentage of the queries answered. If the questionnaire was built based on a *golden-standard* (GS) it is expected that it answers 100% of the queries made.

Another way to measure the capacity of the ontology to model the *domain knowledge* is based on the *golden-standard*, by listing the main aspects of the knowledge that should be modelled (e.g.,

Table 4: Comparison Table by Questionnaires

	C_1			C_2			...	C_j		%
	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	...	Q_{n-1}	Q_n	
O_1	✓			✓	✓		...	✓		50
O_2	✓	✓		✓			...	✓	✓	60
GS	✓	✓	✓	✓	✓	✓	...	✓	✓	100

main classes of the *golden-standard* ontology, categorization of the domain knowledge). We propose another comparative table, as shown in Table 5, to show up if an ontology complies or not with a knowledge category. In Table 5(a), we show a numerical example, that represents the percentage covered on each category based on the number of questions answered in the previous step (Table 4). In Table 5(b), a ✓means that the corresponding ontology conceptualizes the subtopic of the respective category.

The comparative evaluation at this level is based on *application-based*, *user-driven*, and *golden-standard* evaluation methods, supported by questionnaires, use-case scenarios, and the participation of domain experts. It can be enriched with other *structure-based* metrics, sucha as the *Quality* metrics of OQuaRE and its score system [15] (i.e., CROnto and ANOnto).

Table 5: Comparison Table by Domain Knowledge Coverage

(a) Numerical view

	C_1	C_2	C_3
O_1	0	1.0	0.6
O_2	1.0	0.8	0.5
GS	1.0	1.0	1.0

(b) Binary view

	C_1		C_2		C_3	
	T_1	T_2	T_3	T_4	T_5	T_6
O_1		✓	✓		✓	✓
O_2	✓	✓	✓		✓	
O_3	✓	✓		✓		
GS	✓	✓	✓	✓	✓	✓

3.4 Step-by-step comparative evaluation

The methodological comparative process is summarized in the following phases:

(1) With the help of domain experts, define, find, or construct a *golden-standard*. This step marks the *customization* of this methodological comparative evaluation process. Recall the *golden-standard* can be represented by an ontology, a corpus, or a categorization of the domain knowledge. It defines the aspects of the knowledge needed to be represented for a research interest; it represents the specific domain knowledge. Thus, a successful comparative evaluation starts with a correctly specified *golden-standard*.

(2) Prepare the comparison matrices at *Lexical Level* and *Structural Level*, based on, for example, *Linguistic Similarity* (as shown in Eq. (5) and Table 2) and *Structural Similarity*. An advantage of using *Linguistic Similarity* and *Structural Similarity* is that they can be aggregated to get a more accurate metric [19] (*SimLingStruc*), as it is shown in Eq. (6); where $\gamma + \sigma = 1$ are user-defined parameters.

$$SimLingStruct(O_i, O_j) = \gamma * LS(O_i, O_j) + \sigma * StructSim(O_i, O_j) \quad (6)$$

(3) With the support of domain experts, prepare the questions and the SPARQL queries, evaluate each ontology with questions, SPARQL queries, or *golden-standard*, and create the comparison tables, as the ones shown in Table 4 and Table 5).

(4) To reinforce the above evaluations, the *Quakity* OQuaRe metrics can be calculated; we grouped these metrics by the three levels of comparison proposed (*Lexical, Structural*, and *Domain Knowledge*) in Table 6, as a reference.

(5) From the results of previous steps, elaborate a discussion to determine which ontologies are the most appropriate to use, extend, integrate. If the *golden-standard* (*GS*) is available as an ontology, the comparative analysis should result easy, since ontologies more similar to *GS* on each level, could denote better ones. In contrast, if *GS* is available as expert knowledge, we have a comparative evaluation of the considered ontologies against *GS* at *Domain Knowledge Level* and we can identify similar ontologies at *Lexical* and *Structural* levels. As a whole, this information represents a good input to make decisions.

Table 6: OQuaRe Metrics* grouped by Comparison Levels

Lexical Level	Structural Level	Domain Knowledge Level
LCOMOnto, DITOnto, NACOnto, NOCOnto, RFCOnto, NOMOnto, and AROnto	WMCOnto, CBOOnto, RROnto, INROnto, and TMOnto	CROnto and ANOnto

* All equations to calculate them are available at http://miuras.inf.um.es/evaluation/oquare/Metrics.html

4 DEFINITION OF A *GOLDEN-STANDARD*

Since the comparative methodological process is based on a *golden standard*, a referential ontology in the domain must be selected. It can be an ontology proposed as a standard for the appropriate organization (e.g., W3C, the Robotics and Automation Working Group). However, that ontology could not exist, be unavailable, or could cover partially the knowledge that users need to represent. Thus, a better practice is to build the *golden standard* based on a categorization of such as needed knowledge. Actually, with a categorization of the knowledge, the comparative process allows identifying the ontologies that better fit specific research requirements.

As an illustrative example, in this section we show how a *golden standard* is constructed for the SLAM problem. In a previous study, we surveyed the most popular and recent SLAM ontologies, classifying them according to the type of knowledge modelled [10]:

1. Robot Information:
(a) Robot kinematic information
(b) Robot sensory information
(c) Robot pose information
(d) Robot trajectory information
(e) Robot position uncertainty

2. Environment mapping:
(a) Geographical information
(b) Landmark basic information (position)
(c) Landmark shape information
(d) Landmark position uncertainty

3. Timely Information:
(a) Time information of robots and objects
(b) Mobile objects

4. Workspace:
(a) Dimensions of mapping and localization
(b) Specific domain information

Robot Information refers on modelling all aspects related to the presence of a robot in a map that is being built, while the robot navigates on it. Then, it is important to specify the *kinematic information* of the robot, which defines the structure of the robot in terms of geometric information and allows to model the actions the robot can perform and its relations with the space it occupies in the world at every moment. Since the robot needs to use its sensors for mapping (to perceive the world), another important information to model is the *sensory information*, which refers to what sensors the robot has on-board. To perform localization, it is a must to model the *pose* of the robot each moment, this information is intrinsically related to its kinematic information in terms of instantiating it in a specific moment in the space being mapped. A solution of a SLAM problem is based on the interaction of the robot with its world, while it is navigating on the space; thus, the *trajectory* the robot follows must be modeled and stored. During the solution of the SLAM problem, the robot is acting and sensing its own position and trajectory, thus the information becomes uncertain in terms that the robot does not completely know where is it on each moment; hence, *position uncertainty* must be modelled.

Environment Mapping concerns to information that is being mapped by the robot, in terms of *geographical information* of the environment, considering its nature (indoor or outdoor) and the geographical relation between all its components. Also, the model must include relations between different environments of different natures in order to compose maps that are contiguous.

Like occurs with the position of the robot, it is important to model the information about *landmarks* in the environment, which are elements that the robot uses to localize itself and to find relations between elements of is world in order to map it accurately. Thus, for a landmark it is important to model its position (basic information) and also its *shape*, that refers to its geometrical information, and will permit the robot to recognize complex objects and its presence in the environment. Therefore, since the recognition and tracking of landmark position changes over time while the robot navigates and depends on perception, robot self localization uncertainty and the algorithm for inferring positions, it is important to model and storage the *landmark position uncertainty*.

A solution to the SLAM problem must be continuous in time (not a mere stage of the control architecture to be executed in a fixed time), that means, it does not finish while the robot is still working on its environment. Hence, **Time Information** related to the position of robots and objects is needed, because they will change over time depending on actions and sensing of robots. Also, objects could be *mobile*, thus, it is important to differentiate and model those objects in order to avoid miss-correlation of localization between positions of landmarks and robots.

Finally, the nature of environments does not only refer to be indoor or outdoor, but to the **Workspace** of the robot in terms of its *dimensions* (for instance two dimensional or three dimensional) and details about specific *domain information*, that permits to have accurate models and algorithms, considering a priori information about what kind of landmarks the robot could find on it and the semantics of all the objects and situations the robot will find during its navigation. Therefore, it is important to model the complete nature of the environment in order to increase accuracy to the solutions of SLAM problem.

5 STUDY CASE: A COMPARATIVE EVALUATION OF SLAM ONTOLOGIES

In a previous work, we have surveyed the most recent and relevant SLAM ontologies [10]. In order to compare them and illustrate the applicability of the proposed evaluation strategy, we select three available, open-source ontologies – i.e, POS [7], the one proposed by Fortes-Rey [17] (we call it FR2013 for a short reference), and KnowRob [43] – for *Lexical* and *Structural* evaluation, and 24 SLAM ontologies, for *Domain Knowledge* evaluation. POS and FR2013 ontologies extend SUMO [16], an Upper Ontology destined to serve as the basis for computerized information processing systems, and complement Core Ontology [33], which is designed for the domain of robotics and automation. KnowRob ontology is inspired by the Upper Cyc ontology [26]. All of these ontologies are designed for the mobile robots domain, including notions related to position, orientation, and posture. Additionally KnowRob has classes to describe the environment and the objects in it. We compare these three ontologies by pairs.

5.1 Lexical Level

To determine *Linguistic Similarity (LS)*, we first calculate *StringSim*, as in Eq. (2). To compute *DocSim*, we use the class TFIDFVectorizer of the scikit-sklearn library of Python[3]. With these two metrics, we get the *LS* of the selected ontologies, with $\alpha = 0.5$ and $\beta = 0.5$ for Eq. (5), obtaining the results shown on Table 7. As expected, FR2013 and POS have the highest percentage of similarity, since both are extensions of the same ontologies (SUMO and Core).

To complement this measure, we also have performed some of the OQuaRE *Quality* metrics. For *Lexical Level* we consider `NOMOnto`, which is the Number of properties per class (see Table 10).

5.2 Structural Level

Following the proposal, in Table 8 we show the analysis of the components of each ontology. To get these measures, we use WebProtégé[4] and a script in Python, to parse the OWL documents of each ontology and count the components that form them. Having the elements of the ontologies in an accountable way, allows visualizing the result of this comparison. For example, we can deduce that: (i) all ontologies mostly relate their classes as subclasses, as most relationships are `is-a` relations; (ii) KnowRob shows the greatest cohesion, since it has the greatest number of relations; and (iii) POS has the best legibility, because it has the greatest number of annotations.

Also at this level, we can compare the ontologies according to some OQuaRE *Qualitymetrics*. We have compute three of them: (i) `WMCOnto`: Mean number of properties and relationships per class; (ii) `RROnto`: Number of properties defined in the ontology divided by the number of relationships and properties; and (iii) `INROnto`: Mean number of relationships per class (see Table 10).

We do not calculate *Graph Similarity* among the ontologies. Instead, we use Falcon-AO [19], a tool focused on aligning ontologies, which evaluates *Lexical* and *Structural Similarity* together, that is *SimLingStruc* from Eq. (6), with γ=σ=0.5, according to the FALCON-AO documentation. This tool combines *Document Similarity*, based

[3]https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html
[4]https://webprotege.stanford.edu/

Table 7: Lexical Level comparison

Pair	StringSim	DocSim	LS
FR2013/POS	0,37	0,52	0,45
POS/KnowRob	0,11	0,44	0,34
FR2013/KnowRob	0,16	0,55	0,30

Table 8: Comparison Matrix by the Ontologies Structure

Ontology	Classes	Relationships			Properties	Instances	Annotations
		is-a	has-*	other			
FR2013	46	41	0	16	2	15	0
POS	29	29	11	14	0	9	105
KnowRob	252	182	117	76	62	2	3

Table 9: Linguistic/Structural Similarity (from Falcon-AO)

Pair	SimLingStruc
FR2013/POS	0,19
POS/KnowRob	0,05
FR2013/KnowRob	0,07

on *String Similarity*, with *Graph Similarity*. Results are shown oin Table 9. As for *Lexical Level*, POS and FR2013 are the most similar.

5.3 Domain Knowledge Level

Following the proposed methodological process, to evaluate the *Domain Knowledge Level*, we consider two aspects: *application results* and *domain knowledge coverage*. *Application results* can be evaluated with the support of experts in the domain, by elaborating questionnaires and SPARQL queries based on the *golden-standard*. In the following we show the questionnaires related to each category of the *golden-standard* and some examples of the SPARQL queries applied to POS, FR2013, and KnowRob.

(1) **Robot Information:** For SLAM solutions, it is important that the robot knows its location in a map and the pose, because according to that the robot could act differently with its environment. Additionally, other characteristics of the robot, including its geometry and the reference system are needed. As well as the possibility of passing from one system to another, in order to be able to share the acquired knowledge. In order to evaluate if the ontology fulfills the aspects on this category, the following questions were elaborated:
  (a1) Does the ontology store the geometry of the robot?
  (a2) Does the ontology define a referential system for each robot articulation?
  (a3) Does the ontology recognize types of articulations?
  (a4) Does the ontology allow transformations between referential systems?
  (b1) Does the ontology define an own reference systems for each sensor?
  (c1) Does the ontology represent the pose of a robot?
  (c2) Can it represent the relative position of a robot to the objects around it?
  (d1) Does it allow to store a path of the robot and query it?
  (e1) Does the ontology conceptualizes the uncertainty of the robot position?

(2) **Environment Mapping:** The aim with this category is to evaluate the robot's ability to describe the environment in which it is located. Although it is known that the names that objects can take may vary according to the environment, what is sought to know is if the ontology is able to represent other objects than robots. This capability is what opens up the possibility of a more complex SLAM, since if robots are able to differentiate objects from their environments, they have the ability to locate itself either quantitatively or qualitatively with respect to them. In addition, objects could be mobile as in the case of a door, which can be open or closed according to the angle with respect to its point of origin or have subareas of interest such as the knob on the door. To evaluate that, the ontology has to answer questions as the following:
  (a1) Does it allow to store empty spaces and their coordinates?
  (b1) Does it differentiate objects around the robot in terms of their name and characteristics?
  (b2) Does it allow to represent the pose of an object in the robot environment?
  (b3) Does it allow to know the relative position between objects?
  (c1) Does it allow storing the geometry of objects in the robot environment?
  (c2) Does it allow to store sub-objects of interest in larger objects?
  (c3) Does it register objects (other than robots) with joints?
  (d1) Does it model the uncertainty of objects position?

(3) **Timely Information:** This third group of questions are intended to know if the ontology is capable of modeling a path of the robot, representing the its movements: where it has moved and for how long it has remained in that movement or position. Therefore, we need to know if the ontology is able to answer questions as the following:
  (a1) Does it allow to store the different poses of a robot in time?
  (b1) Does it allow to store the different poses of objects in time?

(4) **Workspace Information:** This information is related to general characteristics of environment being mapped, such as its dimensional space as well as the capacity to model entities that belong only to its knowledge domain. Thus, the ontology should answer questions like:
  (a1) Does it clearly indicate the dimensions of the work space in which it works?
  (b1) Does it allow to model specific information of the application domain?

All these questions were translated into SPARQL queries to be applied to POS, FR2013, and KnowRob ontologies. We show only few examples for limitations in the space. The following query shows the main characteristics of the class `Pose` in POS ontology, answering question (c1) of **Robot Information**.

```
PREFIX POS: <http://purl.org/ieee1872-owl/pos#>
SELECT DISTINCT ?predicate ?object
    WHERE { POS:Pose ?predicate ?object }
```

From the result of this SPARQL, we identified the following properties of class `Pose`:

```
<POS:Pose,type,ObjectProperty>
<POS:Pose,subPropertyOf, measure>
<POS:Pose,domain,Object>
<POS:Pose,range,PoseMeasure>.
```

In KnowRob ontology the class `Pose` is also found:

```
<owl:Class rdf:about="http://knowrob.org/kb/knowrob.owl#Pose">
    <rdfs:subClassOf
    rdf:resource="http://www.ease-crc.org/ont/EASE-OBJ.owl#
    6DPose"/>
</owl:Class>
```

For FR2013, a more complex query is needed, since it does not has the class `Pose` defined explicitly. But the pose of the robot can be represented with a vector of points with their respective cartesian `PositionPoints`, each one associated with a representative part of the robot:

```
PREFIX FR2013:
   <http://www.semanticweb.org/ontologies/2013/7/RobotsAutomation.owl#>
SELECT ?subject ?dev ?artifact ?ref ?ofCS
WHERE { FR2013:RobotPart rdfs:subClassOf ?dev ?subject .
                    rdfs:subClassOf ?dev ?dev .
                    rdfs:subClassOf ?artifact ?artifact .
        FR2013:ref rdfs:domain ?ref .
        FR2013:ofCS rdfs:range ?ofCS .
        FR2013:CartesianPositionPoint rdfs:subClassOf ?ofCS }
```

We identified that `FR2013:RobotPart` is subclass of `SUMO:RobotPart`, `SUMO:Device`, `SUMO:Artifact` classes, and related with `CoordinateSystem` and `PositionPoint`.

In FR2013 ontology, it is possible to check how is given the relationship between time and position (i.e., question (a1) related

Table 10: OQuaRE Quality Metrics

Ontologies	Lexical Level		Structural Level						Domain Knowledge Level			
	NOMOnto	score	WMCOnto	score	RROnto	score	INROnto	score	CROnto	score	ANOnto	score
FR2013	0,043	5	0,39	5	3,2 %	1	34,8 %	2	32,6 %	2	0 %	1
POS	0	5	0,48	5	0 %	1	48,3 %	3	31,0 %	2	362 %	5
KnowRob	0,246	5	0,62	5	17,8 %	1	38,1 %	2	0,8 %	1	1,2 %	1

to **Timely Information**). However, since POS is a synchronous ontology, it does not present an approximation for modeling time. To perform the query on FR2013 ontology, the following SPARQL query was done:

```
PREFIX SUMO:<http://www.semanticweb.org/ontologies/2013/7/
                 RobotsAutomation.owl#SUMO:>
SELECT  ?measure ?unit
    WHERE { ?measure rdfs:subClassOf SUMO:PhysicalQuantity .
            ?unit rdfs:subClassOf ?measure }
```

From the answer on this SPARQL, we corroborate that FR2013 ontology models time and position, both as `PhysicalMeasures` in the units of `PositionPoint`, `PositionRegion`, associated to `SUMO:TimePoints` and SUMO:TimeIntervals.

In KnowRob, it is clear the association of the robot pose with a specific time, with the `StampedPose` class and its subclass `SpatioTemporalRegion`. Furthermore, from the queries we got classes associated to time like: `TimeInterval, TimePoint, startTime`, and `endTime`. Table 11 shows the results of applying the questionnaires in the three ontologies.

In the *Knowledge Coverage* evaluation, besides POS, FR2013, and KnowRob ontologies, we consider 21 more ontologies, related to the knowledge representation for the SLAM problem. We present in Table 12 all analyzed ontologies using our *golden-standard*, represented as the categorization of SLAM knowledge. The ontologies are classified taking into account each category and subcategory of the *golden-standard*. Complementing the evaluation described above, the following OQuaRE metrics were computed to strengthen *Quality* evaluation: (i) `CROnto`: Mean number of instances per class and (ii) `ANOnto`: Mean number of annotations per class. All OQuaRE metrics are shown in Table 10.

5.4 Discussion

From *Lexical Level* comparison (see Table 7), although FR2013 and POS ontologies extend SUMO and Core ontologies, they only have 45% of *Linguistic Similarity (LS)*; the low *String Similarity (StringSim)*, 37%, reflects that they do not keep the same entity names, from SUMO and Core. *Document Similarity (DocSim)* takes into account all words in the ontologies, including common expressions on all OWL documents (e.g., `<owl:Class>`, `<owl:ObjectProperty>`), meanwhile *StringSim* only considers the names of entities (i.e., names of classes, properties, instances). Thus, *DocSim* is higher than *StringSim* in all cases.

At *Structural Level* comparison, besides results shown in Table 8, that reflect some characteristics of the three ontologies (already mentioned in the Section 5.2), we use Falcon-AO tool to obtain the combined *Linguistic/Structural Similarity (SimLingStruc)*, as shown in Table 9. This measure is more precise, since it considers not only the similarity of the words but all relations, namely the graph with which the class and its subclasses are expressed. FR2013 and POS are 19% similar, according to the Falcon-AO metric; it means that their RDF graphs are quite different (even though, the *LS* is 45%). As for *LS*, FR2013 and POS have low *SimLingStruc* with KnowRob (i.e., 7% and 5%, respectively).

Table 10 shows the OQuaRE metrics that we compute at the three level of comparison, reinforcing the criterion of *Quality*. According to the OQuaRE score system – 1 means *not acceptable*, 3 is *minimally acceptable*, and 5 is *exceeds the requirements*– the metric taken into account at *Lexical Level*, `NOMOnto` (Number of properties), has the best score for the three ontologies. At the *Structural Level*, the metric `WMCOnto` (Weighted Method Count) also exceeds expectations for the three ontologies; however, for `RROnto` all three ontologies obtained a score of 1, which means that their level of Properties Richness is unacceptable. The third metric applied at this level, is `INROnto`, which is related to the Number of Relationships per Class; POS reached a score of 3 which means minimally acceptable, while the other two reached only a score of 2. At the *Domain Knowledge Level*, the metric `CROnto` evaluates Class Richness, FR2013 and POS scored 2, while KnowRob scored only 1; for `ANOnto` metric, that evaluates Annotation Richness, POS obtained a score of 5 showing its superiority in relation to KnowRob and FR2013, which reached a score of 1.

At *Domain Knowledge Level*, based on our *golden-standard* and questionnaires, we conclude that concerning **Robot Information**, the three ontologies model well robot *kinematic* and *pose* information, but still missing concepts related to *sensors* and *pose uncertainty*. We note an advantage of FR2013 over POS, since it can model the robot's path. Also, it is notable the superiority of KnowRob to describe types of articulations and the representation of the robot position relative to the objects around. In the **Environment Mapping** category, there is a clear limitation on POS and FR2013 for modeling the environment that surrounds robots. Although they have inherited from SUMO the `Object` class and from Core Ontology the definitions of `Robots` and their types, it is only possible that a `Robot` recognizes an `Object` (only another robot) with respect to its position. However, it is not possible to define an `Object` that is not a `Robot`. This is no the case of KnowRob, because it has classes to represent `Objects` with their dimensions, properties like color and also the relative position among them. For **Timely Information**, FR2013 and KnowRob model better the domain of SLAM, since robots can define their position and poses in relation to time. Finally, for the **Workspace**, the three ontologies describe their dimensional workspace, but only KnowRob is able to model specific domain entities at the application level.

Regarding the *knowledge coverage* evaluated on 24 ontologies, we can observe that few ontologies are based only in one category of our *golden-standard*, as ontologies presented in [30, 35, 39], that focus specifically on the final result of the process of solving the

Table 11: Domain Knowledge level - questionnarie

Ontologies	Robot Information									Environment Mapping								Timely Inform.		Workspace Inform.		Questions answered
	a1	a2	a3	a4	b1	c1	c2	d1	e1	a1	b1	b2	b3	c1	c2	c3	d1	a1	b1	a1	b1	%
POS	✓	✓		✓		✓				✓										✓		28.57
FR2013	✓	✓		✓		✓		✓		✓								✓		✓		38.09
KnowRob	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	85.71

Table 12: Analyzed Ontologies under Golden Standard

Name	Ref.	Robot Information					Environment Mapping				Timely Inf.		Workspace	
		(a)	(b)	(c)	(d)	(e)	(a)	(b)	(c)	(d)	(a)	(b)	(a)	(b)
Robot Ontology, 2005	[39]	✓	✓											
Burroughes and Gao, 2017	[5]	✓	✓	✓	✓		✓	✓			✓	✓	✓	
OASys, 2012	[30]	✓	✓											
Core Ontology, 2013	[33]	✓					✓	✓						
POS, 2013	[7]	✓		✓			✓						✓	
SUMO, 2007	[16]		✓	✓			✓	✓			✓		✓	
ISRO, 2020	[9]	✓	✓	✓	✓		✓	✓	✓		✓	✓		✓
ADROn, 2018	[35]	✓	✓											
PROTEUS, 2011	[14]	✓	✓	✓	✓	✓	✓	✓						
RoboEarth, 2015	[36]	✓	✓	✓			✓	✓	✓	✓		✓	✓	✓
OUR-K, 2011	[23]	✓	✓	✓			✓	✓	✓		✓		✓	
Martinez et al., 2007	[28]			✓			✓	✓						✓
Uncertain Ontology, 2011	[34]		✓				✓	✓	✓	✓			✓	✓
ROSPlan, 2015	[8]		✓	✓				✓						
KnowRob, 2012	[43]	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓
Li et al., 2013	[22]			✓			✓	✓					✓	✓
OMRKF, 2007	[40]		✓				✓	✓			✓	✓	✓	✓
Wu et al., 2014	[47]		✓	✓		✓	✓	✓	✓	✓			✓	
Deeken et al., 2018	[12]		✓	✓			✓	✓					✓	✓
Sun et al., 2019	[41]	✓	✓	✓	✓		✓	✓		✓			✓	✓
Crespo et al., 2020	[11]		✓	✓			✓	✓	✓				✓	✓
Space Ontology, 2010	[2]						✓							
Wang and Chen, 2011	[46]						✓	✓	✓				✓	✓
Fortes-Rey, 2013	[17]	✓		✓	✓		✓				✓		✓	

SLAM problem. Since these ontologies only provide **Robot Information** in two subcategories: *kinematic information*, to model the actions of the robots in different ways, and *sensory information* that must be gathered with environment sensors or robot sensors. Further, we identify the Space Ontology [2] that takes into account only the *geographical information* to represent spacial knowledge, related partially to the *Environment Mapping* category.

Other ontologies combine partially several categories, such as: (i) **Robot Information** and **Environment Mapping** covered partially by ontologies proposed in [8, 14, 33]; only PROTEUS [14] covers all the subcategories of **Robot Information**; (ii) **Environment Mapping** and **Workspace** represented partially by the ontology proposed in [46]; and (iii) some subcategories of **Robot Information**, **Environment Mapping**, and **Workspace** considered in [11, 12, 22, 28, 34, 41, 47] and POS [7]. As we shown in the analysis more than a half of the selected ontologies do not consider **Temporal Information**, disregarding dynamic environments for SLAM solutions. Finally, we identify the ontologies presented in [5, 9, 16, 23, 36, 40], KnowRob [43], and FR2013 [17] that model partially aspects in all categories. We can conclude that none of the analyzed ontologies comply our *golden-standard*, showing in this way some limitations to solve the SLAM problem. Even though there exist several ontologies to represent such knowledge, there is a lack of a standard arrangement and generic ontology covering the full aspects of the SLAM knowledge.

Performing these evaluations allow researchers to know better the possibilities, they have when it comes to choosing an ontology to provide semantics to an application.

6 CONCLUSIONS

We present a methodological comparative evaluation strategy, based on *golden standard*, *structure-based*, *application-based*, and *user-driven* evaluation methods. We propose to evaluate *Quality* and *Correctness* of ontologies at *Linguistic, Structural*, and *Domain Knowledge* levels. To demonstrate the suitability of our approach, we evaluate three available ontologies in the robotic domain, in particular, for the SLAM problem (POS, FR2013, and KnowRob) at *Linguistic* and *Structural* levels and 24 ontologies at *Domain Knowledge* level. The methodological comparative process can be *customized* for specif research interests with an appropriate *golden standard*, allowing to determine the gaps in the domain. We have pointed out the gaps that should be considered to propose a wider SLAM ontology. Additionally, the methodology allows selecting the most appropriate metrics, according to the researchers' preferences. We are working on developing an ontology able to model all required knowledge in SLAM solutions.

ACKNOWLEDGEMENTS

The research has received funding from FONDO NACIONAL DE DESARROLLO CIENTÍFICO, TECNOLÓGICO Y DE INNOVACIÓN TECNOLÓGICA - FONDECYT as executing entity of CONCYTEC under grant agreement no. 01-2019-FONDECYT-BM-INC.INV in the project RUTAS: Robots for Urban Tourism, Autonomous and Semantic web based (Robots para centros Urbanos Turísticos Autónomos y basados en Semántica).

REFERENCES

[1] J. Bandeira, I. Bittencourt, P. Espinheira, and S. Isotani. 2016. FOCA: A methodology for ontology evaluation. *arXiv preprint arXiv:1612.03353* (2016).
[2] L. Belouaer et al. 2010. Ontology Based Spatial Planning for Human-Robot Interaction. In *Symp. on Temporal Representation and Reasoning.* 103–110.
[3] J. Brank, M. Grobelnik, and D.a Mladenic. 2005. A survey of ontology evaluation techniques. In *Conf. on Data Mining and Data Warehouses.* 166–170.
[4] C. Brewster et al. 2004. Data driven ontology evaluation. In *Language Resources and Evaluation.* Web & Internet Science, South. Wireless Group, 1–4.
[5] G. Burroughes and Y. Gao. 2016. Ontology-Based Self-Reconfiguring Guidance, Navigation, and Control for Planetary Rovers. *J. Aerospace Inf. Sys.* 13 (2016).
[6] A. Burton-Jones, V. C. Storey, et al. 2005. A semiotic metrics suite for assessing the quality of ontologies. *Data & Knowledge Eng.* 55, 1 (2005), 84–102.
[7] J. L. Carbonera et al. 2013. Defining positioning in a core ontology for robotics. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems.* 1867–1872.
[8] M. Cashmore, M. Fox, D. Long, D. Magazzeni, et al. 2015. Rosplan: Planning in the robot operating system. 2015 (01 2015), 333–341.
[9] D. S. Chang et al. 2020. Ontology-based knowledge model for human-robot interactive services. In *ACM Symp. on Applied Computing.* 2029–2038.
[10] M. Cornejo, R. Ticona, Y. Cardinale, and D. Barrios. 2020. A categorization of simultaneous localization and mapping knowledge for mobile robots. In *ACM Symposium on Applied Computing.* 956–963.
[11] J. Crespo, J. C. Castillo, O. Mozos, and R. Barber. 2020. Semantic Information for Robot Navigation: A Survey. *Applied Sciences* 10, 2 (2020), 497.
[12] H. Deeken, T. Wiemann, and J. Hertzberg. 2018. Grounding semantic maps in spatial databases. *Robotics and Autonomous Systems* 105 (2018), 146–165.
[13] A. Degbelo. 2017. A snapshot of ontology evaluation criteria and strategies. In *Int. Conf. on Semantic Systems.* 1–8.
[14] S. Dhouib et al. 2011. Control architecture concepts and properties of an ontology devoted to exchanges in mobile robotics. In *Control Architectures of Robot.*
[15] A. Duque-Ramos, J. Fernandez-Breis, R. Stevens, and N. Aussenac-Gilles. 2011. OQuaRE: A SQuaRE-based approach for evaluating the quality of ontologies. *Journal of Research and Practice in Information Technology* 43 (05 2011), 159–176.
[16] M. Eid et al. 2007. A Universal Ontology for Sensor Networks Data. In *Computational Intelligence for Measurement Systems and Applications.* 59–62.
[17] V. Fortes. 2013. A positioning ontology for C-SLAM. In *Monograph – UFRGS - Curso de Bacharelado em Ciencia da Computacao.* 23–33.
[18] H. Hlomani and D. Stacey. 2014. Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey. *Semantic Web Journal* 1, 5 (2014).
[19] N. Jian, W. Hu, G. Cheng, and Y. Qu. 2005. Falcon-AO: Aligning Ontologies with Falcon. In *Workshop on Integrating Ontologies.* 87–93.
[20] J. M. Keil and S. Schindler. 2019. Comparison and evaluation of ontologies for units of measurement. *Semantic Web* 10, 1 (2019), 33–51.
[21] B. Lantow. 2016. OntoMetrics: Application of On-line Ontology Metric Calculation. In *BIR Workshops*, Vol. 1684. 1–12.
[22] R. Li et al. 2013. Multi-layered map based navigation and interaction for an intelligent wheelchair. In *Robotics and Biomimetics.* 115–120.
[23] G. H. Lim et al. 2011. Ontology-Based Unified Robot Knowledge for Service Robots in Indoor Environments. *Systems, Man, and Cybernetics* 41, 3 (2011).
[24] A. Maedche and S. Staab. 2002. Measuring Similarity between Ontologies. In *Knowledge Eng. and Knowledge Manag.: Ontologies and the Semantic Web.* Springer.
[25] P. Maillot and C. Bobed. 2018. Measuring structural similarity between RDF graphs. In *ACM Symp. on Applied Computing.* 1960–1967.
[26] C. Matuszek, M. Witbrock, J. Cabral, and J. DeOliveira. 2006. An introduction to the syntax and content of Cyc. *UMBC Computer Science and Electrical Eng. Depart. Collection* (2006).
[27] M. McDaniel and V. C. Storey. 2019. Evaluating Domain Ontologies: Clarification, Classification, and Challenges. *Comput. Surveys* 52, 4 (2019), 1–44.
[28] O. M. Mozos et al. 2007. From Labels to Semantics: An Integrated System for Conceptual Spatial Representations of Indoor Environments for Mobile Robots.
[29] J. Pak and L. Zhou. 2009. A framework for ontology evaluation. In *Workshop on E-Business.* Springer, 10–18.
[30] L. Paull, G. Severac, et al. 2012. Towards an Ontology for Autonomous Robots. In *Conf. on Intelligent Robots and Systems.* 1359–1364.
[31] R. Porzel and R. Malaka. 2004. A task-based approach for ontology evaluation. In *Ontology Learning and Population, Valencia, Spain.* Citeseer, 1–6.
[32] M. Poveda-Villalón et al. 2012. Validating ontologies with Oops!. In *Conf. on knowledge Eng. and Knowledge Management.* Springer, 267–281.
[33] E. Prestes et al. 2013. Towards a core ontology for robotics and automation. *Robotics and Autonomous Systems* 61, 11 (2013), 1193 – 1204.
[34] A. Pronobis and P. Jensfelt. 2011. Multi-modal semantic mapping. In *RSS Workshop on Grounding Human-Robot Dialog for Spatial Tasks.*
[35] F. Ramos, A. Vázquez, et al. 2018. Ontology based design, control and programming of modular robots. *Integrated Computer-Aided Eng.* 25, 2 (2018), 173–192.
[36] L. Riazuelo et al. 2015. RoboEarth Semantic Mapping: A Cloud Enabled Knowledge-Based Approach. *IEEE Trans. on Automation Science and Eng.* 12, 2 (2015), 432–443.
[37] B. Sathiya and T. Geetha. 2019. A review on semantic similarity measures for ontology. *J. of Intelligent & Fuzzy Systems* 36, 4 (2019), 3045–3059.
[38] C. Schlenoff, S. Foufou, and S. Balakirsky. 2012. Performance evaluation of robotic knowledge representation (PERK). In *Performance Metrics for IS.*
[39] C. Schlenoff and E. Messina. 2005. A Robot Ontology for Urban Search and Rescue. In *W. on Research in Knowld. Represent. for Autonomous Systems.* 27–34.
[40] I. H. Suh et al. 2007. Ontology-based multi-layered robot knowledge framework (OMRKF) for robot intelligence. In *Intelligent Robots and Sys.* 429–436.
[41] X. Sun, Y. Zhang, and J. Chen. 2019. High-Level Smart Decision Making of a Robot Based on Ontology in a Search and Rescue Scenario. *Future Internet* 11, 11 (2019), 230.
[42] S. Tartir, I. B. Arpinar, M. Moore, A. P. Sheth, and B. Aleman-Meza. 2005. OntoQA: Metric-based ontology quality analysis. *The Ohio Center of Excellence in Knowledge-Enabled Computing (Kno.e.sis) at CORE Scholar* (2005).
[43] M. Tenorth and M. Beetz. 2009. KNOWROB: knowledge processing for autonomous personal robots. In *RSJ on Intelligent Robots and Systems.* 4261–4266.
[44] S. Thrun. 2003. Exploring artificial intelligence in the New Millennium. Chapter Robotic Mapping: A Survey, 1–35.
[45] A. Ulanov, G. Shevlyakov, et al. 2010. Monte Carlo study of taxonomy evaluation. In *Database and Expert Systems Applications.* IEEE, 164–168.
[46] T. Wang and Q. Chen. 2011. Object semantic map representation for indoor mobile robots. In *Int. Conf. on System Science and Eng.* 309–313.
[47] H. Wu, G. Tian, et al. 2014. Spatial semantic hybrid map building and application of mobile service robot. *Robotics and Autonomous Systems* 62, 6 (2014), 923–941.
[48] S. Zander et al. 2016. A Survey about the Usage of Semantic Technologies for the Description of Robotic Components and Capabilities.. In *SAMI@ iKNOW.*