Business Process Management Journal

Application of a Methodological Approach to Compare Ontologies

Ontologies are formal, well-defined, and flexible representations of knowledge related to a specific domain. They provide the base to develop efficient and interoperable solutions. Hence, a proliferation of ontologies in many domains is unleashed. Then, 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. In a previous study, 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 is based in a *golden-standard*, it can be *customized* to compare ontologies in any domain. In this work, we sought the integration of the OQuaRE quality model to the developed methodology. Using these metrics and the quality model from OQuaRE, we are incorporating a standard of software engineering at the quality validation into the Semantic Web. To show the suitability of our proposal, we apply our methodological approach to conduct comparative studies of ontologies in two different domains, one in the robotic area, in particular for the Simultaneous Localization and Mapping (SLAM) problem; and the other one, in the cultural heritage domain. With these cases of study, 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 domains.

Keywords: Ontologies; SLAM; Semantic Web; Cultural Heritage; Evaluation of Ontologies.

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 (Prestes et al. 2013; Burroughes and Gao 2016). 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 (Sathiya and Geetha 2019). 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 (Maedche and Staab 2002; Brank, Grobelnik, and Mladenic 2005; Burton-Jones, Storey, et al. 2005; Tartir et al. 2005; Pak and Zhou 2009; Duque-Ramos et al. 2011; Poveda-Villalón et al. 2012; Hlomani and Stacey 2014; Lantow 2016; Degbelo 2017; McDaniel and Storey 2019; J. M. Keil and Schindler 2019; Xue et al. 2021). 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 a previous work we propose a methodological process to qualitatively and quantitatively compare ontologies at *Lexical, Structural*, and *Domain Knowledge Modeling* levels, considering *Correctness* and *Quality* perspectives (Cardinale et al. 2020). Our methodological process represents a systematic guideline to perform comparison of ontologies, based in a *golden-standard*, which allows being *customized* to compare ontologies in any domain. 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 this work, we extend the methodological comparison process, by integrating the OQuaRE quality model (Duque-Ramos et al. 2011; Roldan-Molina et al. 2020). Using these metrics and the quality model from OQuaRE, we are incorporating a standard of software engineering at the quality validation into the Semantic Web. 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. In order to show the suitability of our proposal, we apply our methodological process to compare most recent and relevant ontologies in two domains: robotics and cultural heritage.

In the robotic domain, particularly for autonomous robots, their main tasks are mapping an environment and localize themselves, which conform the Simultaneous Localization and Mapping (SLAM) problem (Thrun 2003). SLAM deals with the necessity of building a map of an environment, while simultaneously determining the location of the robot within this map. In a previous study (Cornejo-Lupa et al. 2020), we surveyed the most popular and recent SLAM ontologies, classifying them according to the type of knowledge modeled. In this work, we evaluate and compare them, including our proposed ontology for SLAM (OntoSLAM), following our extended methodological process.

Concerning the cultural heritage domain, it is known that its diffusion and preservation are being supported by technology on the Web, leading to a unleashed creation of cultural heritage ontologies, covering different aspects, such as urban tourism, museums, tourist point of interest. In another previous work, we performed a revision of most recent ontologies for urban tourism and propose our own solution of a cultural heritage ontology, called CURIOCITY (Pinto-De la Gala et al. 2021).

In this work, we assess them following our extended comparative evaluation approach. 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 and urban tourism ontologies, for our cases of study).

In summary, the contribution of this work in three-fold: (i) a *customizable* methodological process extended with the OQuaRE model to conduct a comparative evaluation of ontologies in any domain; it is not an exhaustive list of metrics, but a systematic guideline; (ii) a comparative evaluation of SLAM and cultural heritage 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 the knowledge in a specific domain, there is a lack of a standard arrangement and generic ontology covering the full aspects of such knowledge; hence, an approach to identify the most appropriate ontology(ies) to use, combine, integrate, or extend is a great support in this decision making.

The remainder of this work is structured as follows. Section 2 presents an analysis of evaluation methods for comparison of ontologies. Section 3 draws up our proposed methodological process to compare and evaluate ontologies. Section 4 summarizes the categorization of the knowledge needed for both, to solve the SLAM problem and to represent the cultural heritage domain, in order to use them as *golden-standards* in their corresponding domains. The application of the methodological process to compare ontologies in the cultural heritage domain and to resolve the SLAM problem in robotics is presented in Section 5 and Section 6, respectively. Section 7 sketches out the conclusions of this work.

2. Related Work

The increasing use of ontologies in any domain has inspired the proposal of different metrics to evaluate them in different aspects, both qualitative and quantitative. Several studies have surveyed works describing such metrics (Brank, Grobelnik, and Mladenic 2005; Burton-Jones, Storey, et al. 2005; Tartir et al. 2005; Hlomani and Stacey 2014; Degbelo 2017; Bandeira et al. 2016) and others have focused on evaluating the correct construction of ontologies (McDaniel and Storey 2019; J. M. Keil and Schindler 2019; Thiéblin et al. 2020; Drakopoulos, Voutos, and Mylonas 2020). 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 comparative or evaluation criteria without classify them.

Regarding SLAM ontologies, we only find studies surveying the use of ontologies in the robotic domain (Zander et al. 2016; Crespo et al. 2020), but they do not consider evaluative or comparative issues, beyond of presenting robotic ontologies. We also found a work aimed at evaluating the performance of robotic ontologies, called PERK (Schlenoff, Foufou, and Balakirsky 2012). 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. Concerning the cultural heritage domain, we found some studies that evaluate the usability of ontologies in this domain (Freire and Valk 2019; Freire and Proença 2020). We also found a study to evaluate bibliographic ontologies analysing their classes and properties (Biagetti 2018). The studies related to cultural heritage evaluate specific criteria among ontologies.

More related to our research are the studies that propose a classification of criteria in order to evaluate ontologies at several levels. On each level, different evaluation methods can be applied to measure the corresponding criteria (Brank, Grobelnik, and Mladenic 2005; Hlomani and Stacey 2014), such as:

- *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 (Ulanov, Shevlyakov, et al. 2010);
- *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 (Brewster et al. 2004);
- *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;
- *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 (Porzel and Malaka 2004);
- *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 (Maedche and Staab 2002) 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, Grobelnik, and Mladenic 2005, 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 2009 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. 2011 propose the validation of the *Quality* of ontologies based on SQuaRE[a], a Software Engineering standard. They propose a framework, called OQuaRE[b], which considers two components: a *Quality Model* and *Quality Metrics*. The *Quality Model* considers the following characteristics: *Structural*, *Functional Adequacy*, *Reliability*, *Operability*, *Compatibility*, *Transferability*, and *Maintainability*. Sub-characteristics are specified on each characteristic 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 characteristics (and their sub-characteristics), they propose to score the *Quality Metrics*, from 1 (low quality) to 5 (high quality).

Hlomani and Stacey 2014 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, consis-

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

tency, 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! (Poveda-Villalón et al. 2012) and Ontometrics (Lantow 2016). 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).

Jan Martin Keil 2020 perceives the ontology evaluation and comparison through several correctness and completeness criteria. From these criteria, it is proposed ABECTO, an ABox evaluation and comparison tool of ontologies in the same domain. This framework implements five processors: (i) *Sources* to load ontologies; (ii) *Transformation* to add deduced axioms; (iii) *Maping* to map the ontology resources; (iv) *Comparison* to provide ontologies measurements; and (v) *Evaluation* to identify ontologies mistakes. ABECTO is mainly focused on evaluating the correctness of ontologies using their processors to evaluate their properties as categories. These processors help to evaluate different levels of the ontologies correctness, but not completely because only the ontologies properties are considered.

We believe that for a fair comparison among ontologies both *Correctness* and *Quality* perspectives (Hlomani and Stacey 2014), 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 2009, approaches the three levels evaluation, 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 do in this work.

Table 1: Comparative summary

Reference	**Perspective**	**Lexical**	**Structural**	**Domain Knowld.**	**Comparat. Approach**
Maedche & Staab, 2002	*Quality*	Partial	Partial	Partial	-
Brank et al., 2005	*Correctness & Quality**	Yes	Yes	Partial	-
Pak & Zhou, 2009	*Correctness & Quality**	Yes	Yes	Yes	-
OQuaRE, 2011	*Quality*	Partial	Partial	Partial	-
Hlomani & Stacey, 2014	*Correctness & Quality**	Partial	Partial	Partial	-
OOPS!, 2012	*Correctness*	-	Yes	-	-
Ontometrics, 2016	*Correctness*	-	Yes	-	-
ABECTO, 2020	*Correctnes*	Yes	Partial	Partial	Yes
Our proposal	***Correctnes Quality***	**Yes**	**Yes**	**Yes**	**Yes**
* Not at all levels					

3. Methodological Process for a Comparative Evaluation of Ontologies: Our Proposal

In a previous study (Cardinale et al. 2020), we propose a methodological process to compare ontologies considering *Correctness* and *Quality* perspectives (Hlomani and Stacey 2014) at three levels on the ontology, that group the majority of levels proposed by prior works: *Lexical*, *Structural*, and *Domain Knowledge* levels. Figure 1 shows how these perspectives are considered on each ontology evaluation level. We also suggest *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.

3.1. *Lexical Level*

At this level, the evaluation considers to include linguistic, vocabulary, and syntactic aspects. *Quality* and *Correctness* are measured by analyzing the concepts and the

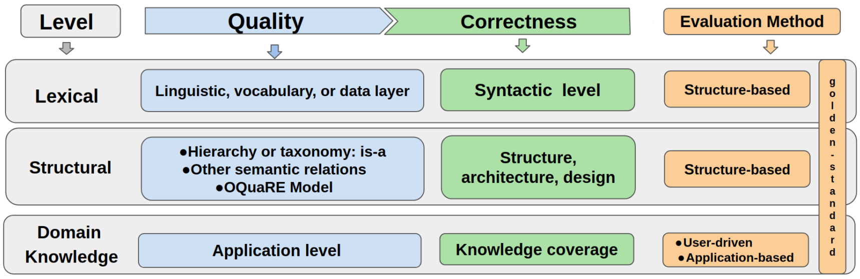

Figure 1: Perspectives, levels, and methods for a comparative study of ontologies

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.

At this level, the evaluation is based on similarity metrics that allow analyzing the proximity of concepts and related vocabulary within the domain, from the ontologies evaluated. The proposed metric, called *Linguistic Similarity (LS)*, is based on string similarity ($StringSim(O_i, O_j)$) and document similarity ($DocSim(O_i, O_j)$) metrics, as shown in Eq. (1), where $\alpha + \beta = 1$, are user-defined parameters; and O_i and O_j are the ontologies been compared, by pairs.

$$LS(O_i, O_j) = \alpha \times StringSim(O_i, O_j) + \beta \times DocSim(O_i, O_j) \tag{1}$$

To calculate these similarity metrics, the ontology entities (i.e., classes, relationships, properties) have to be extracted from their RDF/XML language implementations. Thus, we have the lists of entities names of the ontologies. Then, the string similarity of entity names is calculated by pairs, comparing strings from the corresponding lists, according to Eq. (2), where $s1$ and $s2$ are strings which represent ontology entities; $ed(s1, s2)$ represents the *edit* distance between $s1$ and $s2$; and $s1.len$ and $s2.len$ denote each string length. From results obtained of Eq. (2), the *String Similarity* of two ontologies ($StringSim(O_i, O_j)$ can be computed using Eq. (3), where, S_E is how many times *twoStringSim* (Eq. (2)) is greater than a predefined threshold; N_{Oi}, N_{Oj} are the total number of entities of each ontology; and D_E represents the number of duplicated entities.

$$twoStringSim(s1, s2) = \frac{1}{\frac{ed(s1,s2)}{\exp\left[s1.len + s2.len - ed(s1,s2)\right]}} \tag{2}$$

$$StringSim(O_i, O_j) = \frac{S_E \times 100}{(N_{Oi} + N_{Oj}) - D_E} \tag{3}$$

To calculate the *Document Similarity* ($DocSim(O_i, O_j)$ in Eq. (1)), it is proposed to use the Vector Space Model (VSM) to evaluate linguistic similarity between two ontologies (Jian et al. 2005). In this sense, 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) extracted from the lists of 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 is presented in Eq. (4), where t is the number of times a term occurs in a document, T is the total terms in document, D is the total of documents to compare; and d denotes the number of documents where the term occurs at least once. Then, the *Document Similarity* between the two ontologies is calculated by taking the cosine dot product, as Eq. (5) shows, where VS_{O_*} are the term weighting vectors of the ontologies.

$$TermWeighting = TF \times IDF; TF = \frac{t}{T}; IDF = \frac{1}{2} \times (1 + \log 2\frac{D}{d}) \qquad (4)$$

$$DocSim(O_i, O_j) = \frac{VS_{O_i} \cdot VS_{O_j}^t}{\|VS_{O_i}\| \|VS_{O_j}\|} \qquad (5)$$

3.2. *Structural Level*

At this level, aspects related to taxonomy, hierarchy, relationships, architecture, and design are evaluated. 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, *structure-based* evaluation methods are also appropriate, such as *Graph Similarity* or *RDF Similarity* (Jian et al. 2005; Maillot and Bobed 2018). 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.

In particular, to measure *Quality* at this level, the OQuaRE metrics are appropriate (Duque-Ramos et al. 2011), that measure the number and type of relations and the length of paths among classes in the ontology, as well as the number of properties, attributes, and annotations. OQuaRE metrics are calculated according to Eqs. (6) to (18).

$$
\begin{aligned}
LCOMOnto &= \Sigma PathLength(C_{Thing}, Leaf_{C_i})/\Sigma PathLeafC_j &(6)\\
WMCOnto &= \Sigma PathLength(C_{Thing}, Leaf_{C_i})/\Sigma Leaf_{C_i} &(7)\\
DITOnto &= \max(PathLength(C_{Thing}, Leaf_{C_i})) &(8)\\
NACOnto &= \Sigma C_i \Sigma AncC_j/\Sigma Leaf_{C_j} &(9)\\
NOCOnto &= \Sigma C_i \Sigma SubC_j/\big(\Sigma C_i - \Sigma Leaf_{C_k}\big) &(10)\\
CBOnto &= \Sigma C_i \Sigma AncC_j/\big(\Sigma C_i - \Sigma CT_k\big) &(11)\\
RFCOnto &= \big(\Sigma C_i \Sigma ProC_j + \Sigma C_i \Sigma AncC_k\big)/\Sigma C_i &(12)\\
NOMOnto &= \Sigma C_i \Sigma ProC_j/\Sigma C_i &(13)\\
RROnto &= \Sigma C_i \Sigma SubC_j/\big(\Sigma C_i \Sigma SubC_j + \Sigma C_i \Sigma ProC_k\big) &(14)\\
AROnto &= \Sigma C_i \Sigma RestC_j/\Sigma C_i &(15)\\
INROnto &= \Sigma C_i \Sigma SubC_j/\Sigma C_i &(16)\\
ANOnto &= \Sigma C_i \Sigma ApC_j/\Sigma C_i &(17)\\
TMOnto2 &= \Sigma C_i \Sigma AncC_j/\Sigma C_i &(18)
\end{aligned}
$$

where,

- C_i: Ontology classes.
- R_{Ci}: Relations of class C_i.
- Pro_{Ci}: Properties of class C_i.
- Anc_{Ci}: Direct ancestor of class C_i.
- Sub_{Ci}: Direct subconcept of class C_i.
- C_{Thing} : Ontology root.

These metrics address the validation of the *Quality* of ontologies based on SQuaRE[c], a Software Engineering standard, and on a Quality Model[d]. Each metric is rated with a score ranging from 1 to 5, to measure the following characteristics: *Structural, Functional Adequacy, Compatibility, Reliability, Transferability, Operability* and *Maintainability*[e].

3.3. *Domain Knowledge Level*

It considers evaluating how effectively the knowledge has been covered and how the results of the application are improved by the use of the ontology. Using a *golden-standard* imples the presence of experts of the domain during the comparison process; thus, *application-based* and *user-driven* evaluation methods can be monitored by these experts.

3.4. *Step-by-step comparative evaluation*

The methodological comparative process proposed on (Cardinale et al. 2020) is summarized in the following phases:

[c]SQuaRE: SO/IEC 25000:2005 standard for Software product Quality Requirements and Evaluation
[d]`http://miuras.inf.um.es/evaluation/oquare/`
[e]`http://miuras.inf.um.es/oquarewiki/index.php5/Main_Page`

(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 *Linguistic Similarity* and *Structural Similarity*. An advantage of using *Linguistic Similarity* and *Structural Similarity* (Jian et al. 2005), is that they can be aggregated to get a more accurate metric (*SimLingStruc*), as it is shown in Eq. (19); where $\gamma + \sigma = 1$ are user-defined parameters.

$$SimLingStruct(O_i, O_j) = \gamma * LS(O_i, O_j) + \sigma * StructSim(O_i, O_j) \qquad (19)$$

(3) *Quality* evaluation at *Structural Level* can be complemented by using the OQuaRE model; thus, quality software metrics are adapted to the Semantic Web.
(4) 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.
(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.

In the previous study we show 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.

4. Definition of a ***golden-standard***

As the comparative methodological approach suggests, the first step is to define the *golden-standard*, which customize the domain. In this work, for both cases of study, we have identified the *golden-standard* as the categorization of the knowledge. Actually, with a categorization of the knowledge, the comparative process allows identifying the ontologies that better fit specific research requirements. Both *golden-standards* have been validated by experts and are already published in the research community.

4.1. *Golden-standard for Cultural Heritage Knowledge*

We adopt our *golden-standard* from the knowledge categorization of cultural heritage domain proposed in (Pinto-De la Gala et al. 2021). This work considers the following categories:

(1) **Temporal Item**
  (a) Event
  (b) Time-Span
(2) **Permanent Item**
  (a) Place
  (b) Actor
  (c) Physical Object
  (d) Material
  (e) Person Extended
(3) **Ranking**
(4) **Exhibition**
  (a) Digital representations
  (b) Digital Processing and Analysis
  (c) Collections
  (d) Narrative
(5) **Extended Cultural Heritage**:
  (a) Performance
  (b) Site as Cultural Heritage
  (c) Event as Cultural Heritage
  (d) Culinary Tradition
  (e) Music and Songs

Temporal Item category refers to elements related to events, periods of time (happening before being). **Permanent item** are concepts to describe elements with persistent identity like places, actors, artifacts, materials, and even concepts related to singular characteristic of the person like nicknames, gender or religion (Person Extended). **Ranking** refers to elements which permit to rank items of an exhibition from different criteria . **Exhibition** are concepts related to describe and setting up an exhibition. **Extended Cultural Heritage**, are needed elements to define the concept of cultural heritage in an extended way, thus, it is possible include cultural expressions like performing arts, music, or culinary.

4.2. *Golden-standard for SLAM Knowledge*

The *golden-standard* for the SLAM problem (Cornejo et al. 2020), is composed by four knowledge categories:

(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 1: A Comparative Evaluation of Cultural Heritage Ontologies

We have performed a preliminary revision of relevant and recent ontologies for cultural heritage and we have proposed our own solution, called CURIOCITY[f] (***C****ultural Heritage for* ***Ur****ban Tourism in* ***I****ndoor/****O****utdoor environments of the* ***CITY***) (Pinto-De la Gala et al. 2021). In this work, we compare CURIOCITY ontology and ERLANGEN-CRM[g], an implementation of the ISO Standard Conceptual Reference Model (CIDOC CRM) (Doerr 2005), at *Lexical* and *Structural* levels. At *Domain Knwoledge* level, we compare twelve ontologies based on the *golden-standard* defined by the knowledge categorization presented in Section 4.1.

ERLANGEN CRM is an interpretation of the text of specification of CIDOC CRM, which is proposed by the Committee for Documentation of the International Council of Museums (CIDOC) in order to provide definitions, structures, classes, and relationships for describing cultural heritage documentation. CIDOC CRM is constantly updated through extension proposals (e.g., CRMDig, CRMgeo).

CURIOCITY ontology is a proposal based on UNESCO's definition of cultural heritage concept, aimed at providing classes and relationships to extend capabilities of cultural heritage representation under the perspective of urban tourism, in order to permit the inclusion of cultural activities (e.g., culinary, festivals, music), outdoor points of interest (e.g., landscapes, monuments), and the application of related concepts to enrich the exhibition description (e.g., ranking, narrative).

For the evaluation of ontologies on each level for both study cases, we have

[f] `https://github.com/giulianodelagala/CURIOCITY`
[g] `http://erlangen-crm.org/`

developed several *parsers* and tools, that are available online[h].

5.1. *Lexical Level*

To determine *Linguistic Similarity (LS)* as it is shown in Eq.(1), we first calculate *StringSim* according to Eq.(2), using edit distance. A *parser* was implemented to extract the ontologies' entities (concepts and properties) from their RDF/XML language implementations. To compute *DocSim* as it is shown in Eq.(5), we use the class TFIDFVectorizer of the scikit-sklearn library of Python[i]. With these two metrics, we get the *LS* of CURIOCITY and ERLANGEN-CRM ontologies, with $\alpha = 0.3$ and $\beta = 0.7$ for Eq. (1) (i.e., we consider greater relevance to *DocSim*), obtaining the results shown in Table 2.

Table 2: Metrics values for Lexical evaluation

Metric	Value
String Similarity (*StringSim*)	0.824
Document Similarity (*DocSim*)	0.760
Linguistic Similarity (*LS*)	0.779

The results obtained (i.e., 77% of similar terms) show an expected outcome, due to both ontologies are derived from the CIDOC CRM standard.

5.2. *Structural Level*

As suggested for the evaluation methodology, for *Quality* validation at *Structural* level, we follow the OQuaRE model. We compute all OQuaRE metrics and scored them according to the OQuaRE scale system (i.e., 1 means *not acceptable*, 3 is *minimally acceptable*, and 5 represents *exceeds the requirements*) (Duque-Ramos et al. 2011). All OQuaRE metrics are calculated automatically by means of a script, that in turns assigns scores according to OQuaRE charts and presents results data and graphics; thus, this script allows a better visual comparative analysis. The script was developed in Python 3.8, using libraries such as Rdflib for the management of the ontology graph and the corresponding queries for metrics calculation; Numpy for numerical processing, Pandas for generation of results tables, and Matplotlib for graph creation.

Table 3 shows all OQuaRE metrics and their corresponding score, for CURIOCITY and ERLANGEN CRM ontologies. At this evaluation level, all OQuaRE's characteristics (i.e., *Structural, Functional Adequacy, Compatibility, Reliability, Trans-*

[h] `https://github.com/Alex23013/ontoSLAM/tree/main/formal-validation`
[i] `https://scikit-learn.org/stable/modules/generated/sklearn.feature\_extraction.text.TfidfVectorizer.html`

ferability, Operability, and *Maintainability*) have been evaluated as presented in the following.

Table 3: OQuaRE Metrics Values

Metric	CURIOCITY		ERLANGEN-CRM	
	Value	Score	Value	Score
Lack of Cohesion in Methods (LCOMOnto)	7.057	2	6.948	2
Weigth method per class (WMCOnto)	10.580	3	15.980	1
Depth of subsumption hierarchy (DITOnto)	10	1	10	1
Number of Ancestor Concepts (NACOnto)	1.160	5	1.160	5
Number of Children Concepts (NOCOnto)	2.603	5	2.684	5
Coupling between Objects (CBOOnto)	1.149	5	1.184	5
Response for a concept (RFCOnto)	7.341	3	7.727	3
Number of properties (NOMOnto)	6.205	2	6.557	2
Relationship Richness (RROnto)	15.34%	1	15.02%	1
Attribute Richness (AROnto)	84.65%	5	79.54%	4
Relationships per concept (INROnto)	112.50%	5	115.90%	5
Annotation Richness (ANOnto)	192.04%	5	195.45%	5
Tangledness (TMOnto)	1.136	5	1.170	5

Structural characteristic evaluates ontology quality factors, such as Consistency, Formalization, and Entanglement. Table 4 links each of the sub-characteristics with the scores that OQuaRE assigns for the metrics obtained, both for CURIOCITY and ERLANGEN-CRM ontologies.

Table 4: Quality Evaluation of *Structural* characteristic

Ontology	CURIOCITY							ERLANGEN-CRM						
Sub-characteristic	RROnto	ANOnto	TMOnto	LCOMOnto	Formalization	Consistency	**Score**	RROnto	ANOnto	TMOnto	LCOMOnto	Formalization	Consistency	**Score**
Formal relations	1						1.00	1						1.00
Cohesion				2			2.00				2			2.00
Tangledness			5				5.00			5				5.00
Redundancy		5					5.00		5					5.00
Formalization					5		5.00					5		5.00
Consistency						5	5.00						5	5.00
Total Structural Score							**3.83**							**3.83**

Figure 2 represents the comparison of the two ontologies according to the *Structural* characteristic. In this case, both ontologies score the same for each sub-characteristic. The weakness of both ontologies is given in Cohesion, whose LCOMOnto metric shows there is a strong dependence among the components, mainly due to the complexity of the relationships between the concepts. The other sub-characteristic with a lower score is Formal Relations, linked to the RROnto metric, which indicates that ontologies present a lower number of sub-concepts versus the number of properties; not exactly a symptom of ontologies' weakness, but an indicator of how they are structured.

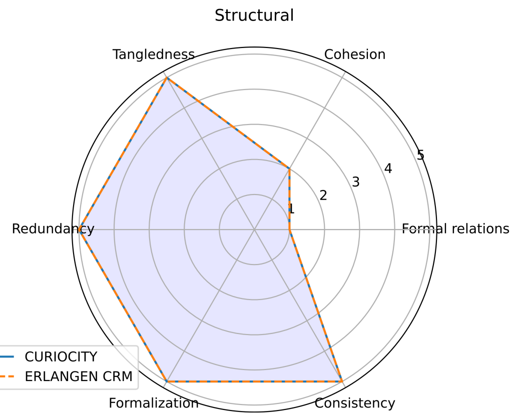

Figure 2: Quality Evaluation of *Structural* characteristic

Table 5 shows the scores of the sub-characteristic related to *Functional Adequacy* quality of both ontologies. Figure 3 represents the comparison of the two ontologies according to these scores. Both ontologies get equal scores for each sub-characteristic. The weakness of them is in Clustering and Similarity sub-characteristics, because a wide range of properties of each concept makes clustering process difficult. The other sub-characteristic with a low score is Results Representation, which indicates that the ontologies are complex; therefore, they have a degree of analysis difficulty in the results they offer.

Table 5: Quality Evaluation of *Functional Adequacy Characteristic*

Ontology	CURIOCITY										
Sub-characteristic	ANOnto	RROnto	AROnto	INROnto	NOMOnto	CROnto	TMOnto	LCOMOnto	Formalization	Consistency	**Score**
Controlled vocabulary	5										5.00
Schema and value reconciliation		1	4						5	5	3.75
Consistent search and query	5	1	4	5					5		4.00
Knowledge acquisition - representation	5	1			2						2.67
Clustering	5										5.00
Similarity		1	4								2.50
Indexing and linking		1	4	5							3.33
Results representation			4			1					2.50
Guidance			4	5							4.50
Decision trees			4	5			5				4.67
Knowledge reuse	5		4	5	2			2	5	5	4.00
Inference		1				1			5		2.33
CURIOCITY Functional adequacy Score											**3.73**
Ontology	ERLANGEN - CRM										
Controlled Vocabulary	5										5.00
Schema and value reconciliation		1	4						5	5	3.75
Consistent search and query	5	1	4	5					5		4.00
Knowledge acquisition - representation	5	1			2						2.67
Clustering	5										5.00
Similarity		1	4								2.50
Indexing and linking		1	4	5							3.33
Results representation			4			1					2.50
Guidance			4	5							4.50
Decision trees			4	5			5				4.67
Knowledge reuse	5		4	5	2			2	5	5	4.00
Inference		1				1			5		2.33
ERLANGEN-CRM Functional adequacy Score											**3.73**

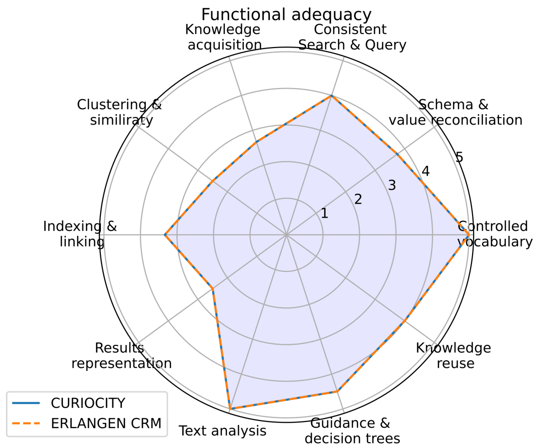

Figure 3: Quality Evaluation of *Functional Adequacy* Characteristic

Compatibility characteristic assesses the ability of ontologies of exchanging information. Table 6 relates the Replaceability sub-characteristic with the OQuaRE score for both for CURIOCITY and ERLANGEN CRM ontologies. CURIOCITY obtains a higher score, which indicates that it can be used instead of another ontology with the same purpose in the same environment.

Table 6: Quality Evaluation of *Compatibility* Characteristic

Ontology	CURIOCITY					ERLANGEN-CRM				
Sub-characteristic	WMCOnto	DITOnto	NOCOnto	NOMOnto	**Score**	WMCOnto	DITOnto	NOCOnto	NOMOnto	**Score**
Replaceability	3	1	5	2	2.75	1	1	5	2	2.25
Total Compatibility Score					**2.75**					**2.25**

Reliability characteristic evaluates the ability of ontologies to maintain their performance for a given period of time. Table 7 relates each sub-characteristic with the OQuaRE score obtained, both for CURIOCITY and ERLANGEN CRM ontologies.

In *Transferability* characteristic, ontology quality factors are evaluated in order to be transferred from one environment to another. Table 8 relates Adaptability sub-characteristic of *Transferability* with the corresponding OQuaRE score.

Operability characteristic evaluates ontology quality factors that indicate the effort needed for its use. Table 9 links Learnability sub-characteristic of *Operability* with the OQuaRE score assigned.

Figure 4 shows the comparison of the two ontologies according to the sub-characteristics corresponding to *Compatibility*, *Reliability*, *Transferability*, and *Operability* characteristics. CURIOCITY obtains a higher score for each of these characteristics, which may indicate an expected better performance. These scores are mainly due to a WMCOnto metric higher value; which denotes that CURIOCITY is less complex.

Maintainability characteristic evaluates ontology quality factors related to the capacity to be modified by changes in the environment, by requirements or in functional specifications. Table 10 links each of *Maintainability* sub-characteristics with

Table 7: Quality Evaluation of *Reliability* Characteristic

Ontology	CURIOCITY							ERLANGEN-CRM						
Sub-characteristic	WMCOnto	DITOnto	NOCOnto	RFCOnto	NOMOnto	LCOMOnto	**Score**	WMCOnto	DITOnto	NOCOnto	RFCOnto	NOMOnto	LCOMOnto	**Score**
Recoverability	3	1			2	2	2.00	1	1			2	2	1.50
Availability	3		5	3		2	3.25	1		5	3		2	2.75
Total Reliability Score							**2.63**							**2.13**

Table 8: Quality Evaluation of *Transferability* Characteristic

Ontology	CURIOCITY					ERLANGEN-CRM				
Sub-characteristic	WMCOnto	DITOnto	RFCOnto	CBOnto	**Score**	WMCOnto	DITOnto	RFCOnto	CBOnto	**Score**
Adaptability	3	1	3	5	3.00	1	1	3	5	2.50
Total Transferability Score					**3.00**					**2.50**

Table 9: Quality Evaluation of *Operability* Characteristic

Ontology	CURIOCITY							ERLANGEN-CRM						
Sub-characteristic	WMCOnto	LCOMOnto	RFCOnto	NOMOnto	CBOnto	NOCOnto	**Score**	WMCOnto	LCOMOnto	RFCOnto	NOMOnto	CBOnto	NOCOnto	**Score**
Learnability	3	2	3	2	5	5	3.33	1	2	3	2	5	5	3.00
Total Operability Score							**3.33**							**3.00**

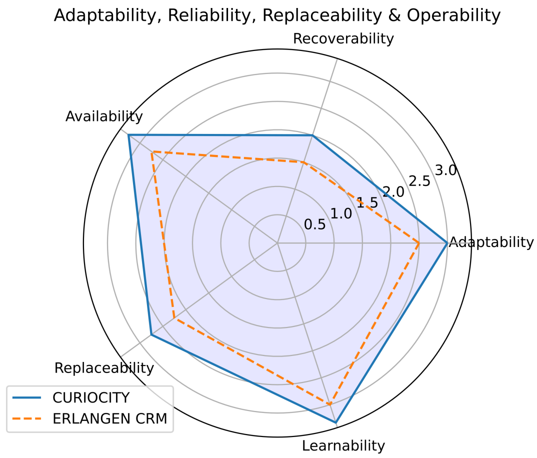

Figure 4: Quality Evaluation of *Compatibility*, *Reliability*, *Transferability*, and *Operability* Characteristics

the corresponding OQuaRE scores, both for CURIOCITY and ERLANGEN CRM ontologies. Figure 5 shows *Maintainability* ontologies comparison. CURIOCITY shows higher scores, in all the sub-characteristics. Weakest scores are in Analysability and Testability sub-characteristics, which can be understood there is a degree of difficulty in deficiencies diagnosis and validation.

After obtaining each characteristic score, we make a comparison between both ontologies and identify ontologies weaknesses and strengths. Table 11 shows the OQuaRE evaluation summary for CURIOCITY and ERLANGEN CRM ontologies.

Table 10: Quality Evaluation of *Maintainability* Characteristic

Ontology	CURIOCITY								ERLANGEN-CRM							
Sub-characteristic	WMCOnto	DITOnto	NOCOnto	RFCOnto	NOMOnto	LCOMOnto	CBOOnto	**Score**	WMCOnto	DITOnto	NOCOnto	RFCOnto	NOMOnto	LCOMOnto	CBOOnto	**Score**
Modularity	3						5	4.00	1						5	3.00
Reusability	3	1	5	3	2		5	3.16	1	1	5	3	2		5	2.83
Analysability	3	1		3	2	2	5	2.66	1	1		3	2	2	5	2.33
Changeability	3	1	5	3	2	2	5	3.00	1	1	5	3	2	2	5	2.71
Modification stability	3		5	3		2	5	3.60	1		5	3		2	5	3.20
Testability	3	1		3	2	2	5	2.66	1	1		3	2	2	5	2.33
Total Maintainability Score								**3.18**								**2.73**

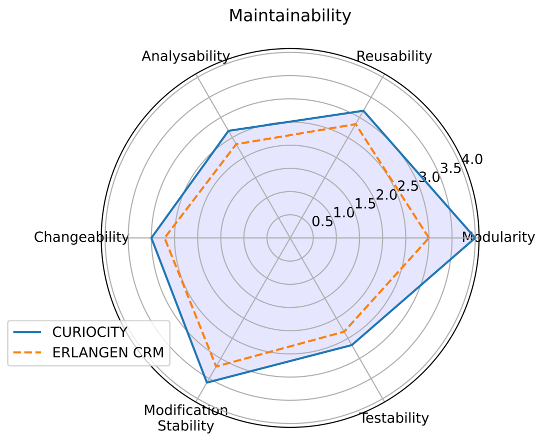

Figure 5: Quality Evaluation of *Maintainability* Characteristic

Figure 6 shows the comparison summary of the two ontologies according to the OQuaRE evaluation. CURIOCITY obtains an overall higher score. Both ontologies present equal scores in *Structural* and *Functional Adequacy* characteristics, at the same time these are the characteristics that present greater strength. CURIOCITY shows a better score in *Functional Adequacy*, *Compatibility*, *Reliability*, *Transferability*, *Operability*, and *Maintainability*, due to CURIOCITY has less complexity, since it has taken only a necessary portion of concepts offered by CIDOC CRM; which allows CURIOCITY to be easier to adapt and maintain.

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,

Table 11: OQuaRE Evaluation Summary

Ontology	CURIOCITY	ERLANGEN-CRM
Characteristic	Score	Score
Structural	3.83	3.83
Functional adequacy	3.73	3.73
Compatibility	2.75	2.25
Reliability	2.63	2.13
Transferability	3.00	2.50
Operability	3.33	3.00
Maintainability	3.18	2.73
Mean Score	**3.21**	**2.88**

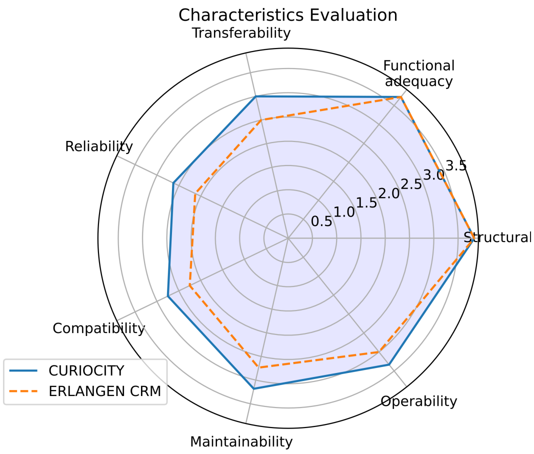

Figure 6: OQuaRE Evaluation Summary

by elaborating questionnaires and SPARQL queries based on the *golden-standard*. Thus, we elaborate questionnaires related to each category of the defined *golden-standard* and a corresponding SPARQL query to each ontology, as follows:

(1) **Temporal Item**

- (1a) Does the ontology represent events?
- (1b) Does the ontology represent time periods?

(2) **Permanent Item**

- (2a) Does the ontology represent places?
- (2b) Does the ontology represent person or groups?
- (2c) Does the ontology represent physical artworks?
- (2d) Does the ontology define kind of material from which artworks are made?
- (2e) Does the ontology represent some particular characteristics of a person?

(3) **Ranking**

(3) Can the ontology store some valuation or score given to an item?

(4) **Exhibition**

(4a) Does the ontology allow store some information about the digital representation of artifacts?
(4b) Does the ontology allow to store some information about digital analysis of artifacts?
(4c) Does it allow to describe additional information about an exhibition?
(4d) Does it allow to store some useful information to be used as curatorial narrative?

(5) **Extended Cultural Heritage**

(5a) Does it allow to represent performing arts?
(5b) Does the ontology allow to represent outdoor environments as cultural heritage?
(5c) Does the ontology represent events as cultural heritage?
(5d) Does it represent food with cultural interest?
(5e) Does it represent traditional music?

We have approached the query formulation based on filters for keywords related to each question; then, we inspect the relevance of the results to the given question. The SPARQL queries have the form:

```
PREFIX ecrm: <http://erlangen-crm.org/170309/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX time: <http://www.w3.org/2006/time#>
PREFIX cit: <http://curiocity.org/>
#Search Coincidence by keywords
SELECT DISTINCT ?s ?p ?o
WHERE {  ?s ?p ?o .
  FILTER (regex(str(?s) , "keyword", "i")) }
```

Some examples of this validation process are given in the following. In case of question (1b), using ”time” as the keyword for the query, we can obtain ontology elements linked to the representation of time periods. The SPARQL query returns 144 related triples for CURIOCITY, for instance:

```
time:Instant rdfs:subClassOf time:TemporalEntity
time:Interval rdfs:subClassOf time:TemporalEntity
cit:SP10_Declarative_Time-Span rdfs:subClassOf ecrm:E52_Time-Span
```

while for ERLANGEN-CRM, 90 triples are obtained. For this domain category, both ontologies share many concepts and properties.

In question (4a), we use ”digital” as keyword. CURIOCITY reports seven related triples, which are adopted from the CIDOC CRM extension CRMDig (Doerr and Theodoridou 2014), for instance:

```
cit:D11_Digital_Measurement_Event rdfs:subClassOf ecrm:E16_Measurement
cit:D1_Digital_Object rdfs:subClassOf ecrm:E73_Information_Object
```

We can also apply the filter to the object of the triples, with the purpose of locating related concepts as indicated by their comments or notes. According to this kind of query, in ERLANGEN-CRM is possible to identify four triples, among these:

```
ecrm:P138_represents rdfs:comment "...digitisation is here seen as a process..."
ecrm:E90_Symbolic_Object rdfs:comment "...represented by a serialized digital..."
```

Table 12: Domain Knowledge level - questionnarie

Ontologies	Temporal Item		Permanent Item					Rank-ing
	1a	1b	2a	2b	2c	2d	2e	3
ERLANGEN CRM	✓	✓	✓	✓	✓	✓		
CURIOCITY	✓	✓	✓	✓	✓	✓	✓	✓

Ontologies	Exhibition				Extended Cultural Heritage				
	4a	4b	4c	4d	5a	5b	5c	5d	5e
ERLANGEN CRM	✓		✓	✓					✓
CURIOCITY	✓	✓	✓	✓	✓	✓	✓	✓	✓

Ontologies	Questions answered
	%
ERLANGEN CRM	58.2
CURIOCITY	100.0

In some cases, ontology inspection can be complicated to validate coverage in a given category. Using a multiple option filter query can give hints of coverage, however, a review of the ontology documentation may be necessary for confirmation. For the case of question (5b) we can use as a query filter:

```
FILTER (regex(str(?o), "monument", "i") || regex(str(?o), "landscape", "i"))
```

from this query we do not obtain related triples for ERLANGEN-CRM, for CURIOCITY we obtain the triples:

```
cit:Natural_Landscape rdfs:label "Natural landscape"
cit:Cemetery rdfs:comment "This class..."
cit:Monument rdfs:label "Monument"
```

which indicates CURIOCITY has concepts related to a more specific representation of places regarded as cultural heritage.

Table 12 summarizes the results of applying questionnaires to both ontologies. CURIOCITY has a better coverage of the domain, because it has been modeled taken into consideration the proposed *golden-standard*. CURIOCITY do not cover in depth all the categories, just identifies some relevant concepts and properties, which allows extending the knowledge to other domains.

In the *knowledge coverage* evaluation, besides CURIOCITY and ERLANGEN CRM ontologies, we evaluate 10 more ontologies, related to the cultural heritage knowledge representation. Table 13 shows the analyzed cultural heritage ontologies using the proposed *golden-standard* (Pinto-De la Gala et al. 2021). The coverage analysis of each ontology was conducted according to the available documentation of each ontology.

5.4. *Discussion*

CURIOCITY and ERLANGEN CRM ontologies have been evaluated at three levels: *Lexical, Structural*, and *Domain Knowledge*.

Table 13: Comparison of Ontologies related to Cultural Heritage

Ontology	Temporal Item	Permanent		Exhibition			Extended Cultural Heritage					Ranking
	Event. Time-Lapse	Physical Object, Place, Actor	Person Extension	Collection	Digit. Repr. & Analysis	Curatorial Narr.	Performing Arts	Site as CH	Event as CH	Culinary Trad.	Music & Songs	
CURIOCITY (Pinto-De la Gala et al. 2021)	●	●	◑	●	●	●	●	●	●	●	●	◑
ERLANGEN CRM - CIDOC CRM (Doerr 2005)	●	●		◑	◑						◑	
Finto* (FINTO 2015)	●	●	◑	●	◑		●	◑	◑	◑	●	
MUSSEUM FINLAND (Hyvönen et al. 2005)	●	●		●								
EDM (Doerr, Gradmann, et al. 2010)	●	●	●	●	●							
SCULPTURE (Addis et al. 2003)	●	●			●							
CURATE (Mulholland, Wolff, and Collins 2012)	●	●			◑	●						
MOM (Hajmoosaei and Skoric 2016)	●	●	●	◑	◑							
OntoMP (Araújo et al. 2017)	●	●	●		◑							
Marchenkov et al. (Marchenkov et al. 2017)	●	●		●	●							◑
TOMS (Chanhom and Anutariya 2019)	●	●		●	◑							
Lo Turco et al. (Turco, Calvano, and Giovannini 2019)	●	●			●							

* Arts and Culture Category: MAO/TAO/KULO/MUSO/SEKO/VALO
◑ Contains related concepts

Lexical level evaluation results show that both ontologies keep a similarity of concepts with CIDOC CRM standard base ontology, which is an indication that a core has been preserved to allow interoperability with the standard and a glossary of common terms.

Structural level evaluation followed the OQuaRE methodology as a guideline. Based on metrics calculated for both CURIOCITY and ERLANGEN CRM, different characteristics were validated. For *Structural* characteristic, a score of 3.83 was obtained for both ontologies, which is considered acceptable, having as a weakness the Cohesion sub-characteristic, an indicator of strong dependence between the ontology's formal relations. *Functional Adequacy* score for both ontologies is 3.73, which is considered an acceptable value. The weaknesses identified are Clustering and Similarity, both indicators of difficulty in the process of identifying groups with similar characteristics in the ontology; another weakness is Results Representation, an indicator of difficulty to analyze results. CURIOCITY achieves a slightly better score for *Compatibility* characteristic (score 2.75) compared to ERLANGEN CRM (score 2.25), an indicator of how easily can be adapted to different environments and also of presenting a lower complexity among relationships. *Reliability* characteristic shows CURIOCITY (score 2.63) an expected better performance over a given period of time than ERLANGEN CRM (score 2.13). Likewise for the *Transferability* characteristic, CURIOCITY (score 3.00) scores higher versus ERLANGEN CRM (score 2.50), this characteristic reinforces the notion of adaptability. In the case of *Operability* characteristic, CURIOCITY scores 3.33, compared to score 3.0 for ERLANGEN. It means that CURIOCITY has slightly better conditions to be learned, due to a less complexity of its concepts and relationships. The last characteristic evaluated is *Maintenance*, which also shows that CURIOCITY (score 3.18) has better conditions than ERLANGEN CRM (score 2.73). Both ontologies show Analysability and Testability as the weakest sub-characteristics.

Having reviewed the characteristics evaluated, it can be concluded that CURIOCITY shows more favorable conditions compared to ERLANGEN CRM. This should not be understood as meaning that one is better than the other, which would be false; instead, these metrics help to identify the ontologies' strengths and weaknesses. For both ontologies, a point to take into consideration is their complexity, which translates into maintenance, learning, and analysis issues.

Domain Knowledge level evaluation shows a broader coverage of the cultural heritage domain by CURIOCITY, fulfilling its objective of being able to characterize elements of cultural heritage in indoor and outdoor environments.

6. Study Case 2: A Comparative Evaluation of SLAM Ontologies

In a previous work, we have surveyed the most recent and relevant SLAM ontologies (Cornejo et al. 2020). In order to compare them and illustrate the applicability of the proposed evaluation strategy, we select three available, open-source ontologies – i.e., POS (Carbonera et al. 2013), the one proposed by Fortes-Rey (Fortes 2013) (we call it FR2013 for a short reference), and KnowRob (Tenorth and Beetz 2009) – and our own solution, called OntoSLAM [j], for *Lexical* and *Structural* evaluation. For *Domain Knowledge* evaluation, we consider other twenty one SLAM ontologies.

POS and FR2013 ontologies extend SUMO (Eid et al. 2007), an Upper Ontology destined to serve as the basis for computerized information processing systems, and complement CORA (Prestes et al. 2013), which is designed for the domain of robotics and automation. KnowRob ontology is inspired by the Upper Cyc ontology (Matuszek et al. 2006). 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. OntoSLAM is or proposal to model SLAM information, considering SLAM as a dynamic process. Hence, this ontology considers the temporality associated with the position of the robot and the position of the landmarks in the environment, as well as, the uncertainty associated with these positions. This ontology arises from the extension of FR2013, KnowRob, and ISRO. We compare these four ontologies by pairs at *Lexical* and *Structural* levels.

6.1. *Lexical Level*

Following a similar process from the previous study case and using the same *parser and tools*, we compute *Linguistic Similarity (LS)* from *StringSim* and *DocSim* calculations. The summary of the results are presented in Table 14. The highest value

[j] https://github.com/Alex23013/ontoSLAM/tree/main/finalOntology

Table 14: Lexical Level comparison

Pair	*StringSim*	*DocSim*	*LS*
FR2013/POS	0.374	0.591	0.482
FR2013/KnowRob	0.165	0.505	0.335
FR2013/OntoSLAM	**0.434**	0.618	**0.526**
POS/KnowRob	0.119	0.582	0.351
POS/OntoSLAM	0.422	**0.634**	**0.528**
KnowRob/OntoSLAM	0.177	0.616	0.397

on *StringSim* is between FR2013 and OntoSLAM, since many classes of OntoSLAM were inspired by classes of FR2013. For *DocSim*, the highest similarity is between POS and OntoSLAM, this is because both are extensions of the same ontologies. Thus, according to *LS*, OntoSLAM keeps 52% of similarity with FR2013 and POS.

6.2. *Structural Level*

Following the methodological process, in Table 15 we show the analysis of each ontology components. To get these measures, we use a Python script, in the same way of study case 1, to parse the OWL documents of each ontology and count their components. Having the ontologies elements in an accountable way, allow us to visualize the comparison results For instance, we can deduce that: (i) all of the compared ontologies mostly relate their classes as subclasses, i.e., most relationships are `is-a` relations; (ii) KnowRob shows the greatest cohesion, since it has the greatest number of relations and classes; (iii) OntoSLAM has the best legibility, because it has the greatest number of annotations; and (iv) FR2013 shows a good reproducibility since it has the highest number of instances.

Table 15: 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	7
KnowRob	**252**	**182**	117	76	62	2	3
OntoSLAM	69	86	34	13	14	0	**33**

We do not calculate *Graph Similarity* among the ontologies. Instead, we use Falcon-AO (Jian et al. 2005), a tool focused on aligning ontologies, which evaluates *Lexical* and *Structural Similarity*, that is *SimLingStruc* from Eq. (19), with γ=σ=0.5, according to the FALCON-AO documentation. This tool combines *Document Similarity*, based on *String Similarity*, with *Graph Similarity*. Results are

Table 16: Linguistic/Structural Similarity (from Falcon-AO)

Pair	SimLingStruc
FR2013/POS	0.19
FR2013/KnowRob	0.08
FR2013/OntoSLAM	**0.24**
POS/KnowRob	0.05
POS/OntoSLAM	0,17
KnowRob/OntoSLAM	0,12

shown in Table 16. At this level, the ontologies with the greatest similarity, according to *SimLingStruc* metric, are FR2013 and OntoSLAM, since FR2013 is one of the base ontologies for the development of OntoSLAM; therefore, some classes and their relations are similar (24%). These low similarity values among these ontologies, reflects that they take base ontologies, such as CORA and SUMO, that are extended with classes and relations according to specific needs.

To complement the evaluation and comparison of these ontologies in terms of *Quality*, we also follow the OQuaRE model, as suggested by the methodology. We compute all OQuaRE metrics and scored them, according to the model, in which 1 means *not acceptable*, 3 is *minimally acceptable*, and 5 is *exceeds the requirements*. All OQuaRE metrics are applied on the compared ontologies. The values and scores obtained are shown on Table 17.

Table 17: OQuaRE Metric values for SLAM Ontologies

Metric	FR2013		POS		Knowrob		OntoSLAM	
	Value	Score	Value	Score	Value	Score	Value	Score
Lack of Cohesion in Methods (LCOMOnto)	5.550	3	2.500	4	1.200	5	3.744	4
Weigth method per class (WMCOnto)	5.842	4	0.384	5	0.035	5	3.577	5
Depth of subsumption hierarchy (DITOnto)	6.000	3	2.000	5	1.000	5	5.000	3
Number of Ancestor Concepts (NACOnto)	1.000	5	1.000	5	0.858	1	1.044	5
Number of Children Concepts (NOCOnto)	1.692	5	1.666	5	2.141	5	2.708	5
Coupling between Objects (CBOOnto)	0.955	1	1.000	5	0.7222	1	1.043	5
Response for a concept (RFCOnto)	1.600	5	1.793	5	0.845	1	2.144	5
Number of properties (NOMOnto)	0.644	5	0.793	5	0.123	5	1.101	5
Relationship Richness (RROnto)	60.27%	4	17.85%	1	85.44%	5	46.99%	3
Attribute Richness(AROnto)	11.11%	1	31.03%	5	5.15%	1	2.89%	1
Relationships per concept (INROnto)	97.77%	5	17.24%	1	72.22%	4	94.20%	5
Annotation Richness (ANOnto)	71.11%	4	110.34%	5	12.69%	1	46.37%	3
Tangledness (TMOnto)	0.955	1	1.000	5	0.722	1	1.043	5

For the *Structural* characteristic evaluation (shown in Table 18) POS and OntoSLAM are the ones with the highest score of 4.16. Followed by FR2013 and KnowRob, with scores of 3.83 and 3.66 respectively. The analysis at sub-characteristic levels, shown in Figure 7, reveals that in Consistency and Formalization all the ontologies have a score of 5. What makes OntoSLAM and POS outperform the rest is their score of 5 in Tangledness. Also, we can note the inversely proportional performance of the ontologies for the Redundancy and Formal relations sub-characteristics.

Table 18: Quality Evaluation of *Structural* Characteristic for SLAM Ontologies

Ontology	POS							FR2013						
Sub-characteristic	RROnto	ANOnto	TMOnto	LCOMOnto	Formalization	Consistency	**Score**	RROnto	ANOnto	TMOnto	LCOMOnto	Formalization	Consistency	**Score**
Formal relations	1						1.00	4						4.00
Cohesion				4			4.00				3			3.00
Tangledness			5				5.00			5				5.00
Redundancy		5					5.00		1					1.00
Formalization					5		5.00					5		5.00
Consistency						5	5.00						5	5.00
Total Structural Score							**4.16**							**3.83**
Ontology	KnowRob							OntoSLAM						
Sub-characteristic	RROnto	ANOnto	TMOnto	LCOMOnto	Formalization	Consistency	**Score**	RROnto	ANOnto	TMOnto	LCOMOnto	Formalization	Consistency	**Score**
Formal relations	5						5.00	3						3.00
Cohesion				5			5.00				4			4.00
Tangledness			1				1.00			5				5.00
Redundancy		1					1.00		3					3.00
Formalization					5		5.00					5		5.00
Consistency						5	5.00						5	5.00
Total Structural Score							**3.66**							**4.16**

The analysis of the *Functional Adequacy* characteristic is calculated with the same metrics used in Table 5. Figure 8 shows the results of this analysis. The highest score of all ontologies is for the Text Analysis subfeature, which means that the structure presented by the ontologies helps to associate the words with the concepts they model. A special case is for the Controlled Vocabulary subfeature, in which the ontologies analyzed have different scores. This subfeature refers to the ability of entity labels to avoid heterogeneity. Having a low score, like KnowRob, means that there is no homogeneity in the entity labels. This heterogeneity is more difficult

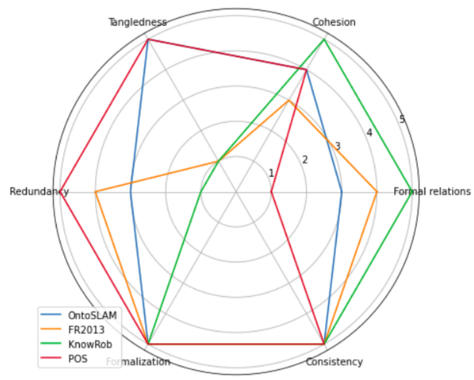

Figure 7: Quality Evaluation of *Structural* Characteristic for SLAM Ontologies

to avoid in large ontologies such as KnowRob. In the remain subfeatures, FR2013 and OntoSLAM get scores greater than 3, i.e., they satisfy this characteristic of *Functional Adequacy*; while KnowRob and POS, with scores less than 3, do not reach this feature.

Figure 8: Quality Evaluation for *Functional Adecuacy* Characteristic for SLAM ontologies

Unlike *Functional Adequacy*, the *Compatibility* characteristic has only the sub-feature of Replaceability. To evaluate this sub-characteristic the following metrics are taken into account: WMCOnto, DITOnto, NOCOnto and NOMOnto. Table 19 shows the characteristic evaluation, where POS and KnowRob achieved a perfect score of 5, followed by OntoSLAM with 4.5 and FR2013 with 4.25.

Table 19: Quality Evaluation of *Compatibility* Characteristic for SLAM ontologies

Sub-characteristic	Replaceability				
Ontology	WMCOnto	DITOnto	NOCOnto	NOMOnto	**Score**
POS	5	5	5	5	5.00
FR2013	4	3	5	5	4.25
KnowRob	5	5	5	5	5.00
OntoSLAM	5	3	5	5	4.50

Transferability characteristic has only Adaptability sub-characteristic to be evaluated. For this sub-characteristic, metrics WMCOnto, DITOnto, RFCOnto, and CBOnto are considered. In the evaluation of Adaptability, POS achieves a perfect score of 5, followed by OntoSLAM (score of 4.5). POS achieves a high score in this characteristic because being a small ontology it is easily adaptable. Unlike KnowRob which is a large ontology.

Table 20: Quality Evaluation of *Transferability* Characteristic for SLAM ontologies

Sub-characteristic	Adaptability				
Ontology	WMCOnto	DITOnto	RFCOnto	CBOnto	**Score**
POS	5	5	5	5	5.00
FR2013	4	3	5	1	3.25
KnowRob	5	5	1	1	3.00
OntoSLAM	5	3	5	5	4.50

Table 21 shows the ontologies performance when evaluated on the characteristic of *Operability*. In the same way as in *Compatibility*, OntoSLAM and POS are the ones that obtain the highest scores (4.83). They are followed by KnowRob and FR2013, with scores of 3.83 and 3.66, respectively.

Figure 9 shows the related sub-characteristics of four characteristics (*Compatibility*, *Reliability*, *Transferability*, and *Operability*). OntoSLAM (score 4.5) and POS (score 5.00) have the highest score for *Adaptability* sub-characteristic, same behaviour they show in *Learnability* sub-characteristic with score 4.83, which ensures that they both can be easily learned by new users. For Availability sub-characteristic, the highest score was obtained by KnowRob (score 5) and the lowest by FR2013 (score 3). In the remaining sub-characteristics (Replaceability and Recoverability) all ontologies present good results, KnowRob and POS (score 5) have the highest, while FR2013 and POS present an acceptable score.

Table 21: Quality Evaluation of *Operability* Characteristic for SLAM ontologies

Sub-characteristic	Learnability						
Ontology	WMCOnto	LCOMOnto	RFCOnto	NOMOnto	CBOnto	NOCOnto	**Score**
POS	5	4	5	5	5	5	4.83
FR2013	4	3	5	5	1	5	3.83
KnowRob	5	5	1	5	1	5	3.66
OntoSLAM	5	4	5	5	5	5	4.83

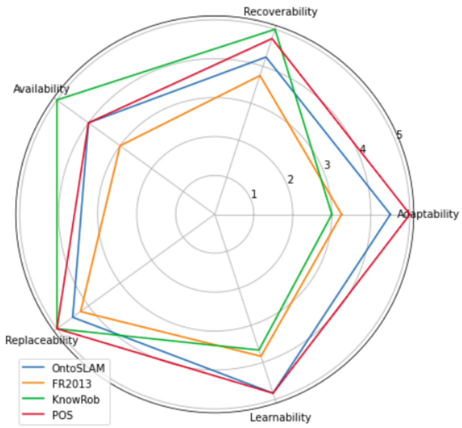

Figure 9: Quality Evaluation of *Compatibility*, *Reliability*, *Transferability*, and *Operability* Characteristics for SLAM ontologies

Last characteristic to be evaluated is *Maintainability*, shown in Figure 10. The scores were calculated with the metrics from Table 10. POS and OntoSLAM have scores higher than 4 in all the sub-characteristics; while KnowRob and FR2013 present scores lower than 4 in all the sub-characteristics, showing a weakness in the Redundancy sub-characteristic.

Table 22 shows a summary of all the characteristics evaluated in OQuaRE *Quality* model. After the analysis, we obtain results that shown that POS and OntoSLAM are the ontologies with the highest mean scores, 4.45 and 4.27, respectively. Although POS has the highest score, it is not as consistent as OntoSLAM, which has scores of more than 3 for all characteristics. In the case of KnowRob, despite having perfect scores in two characteristics (*Reliability* and *Compatibility*), the low scores in the others decrease its mean score.

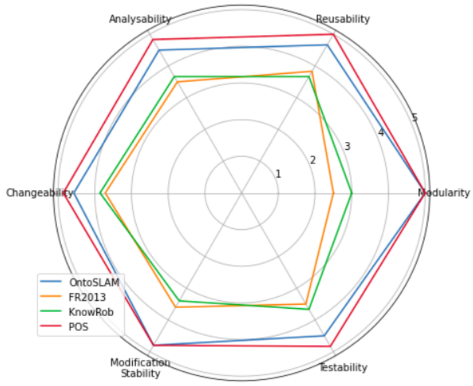

Figure 10: Quality Evaluation of *Maintainability* Characteristic for SLAM ontologies

Table 22: OQuaRE Evaluation Summary for SLAM Ontologies

Ontology	**POS**	**FR2013**	**KnowRob**	**OntoSLAM**
Characteristic	Score	Score	Score	Score
Structural	4.16	3.67	3.67	4.16
Functional adequacy	2.89	3.42	2.99	3.10
Reliability	4.38	3.38	5.00	4.13
Operability	4.83	3.83	3.66	4.83
Compatibility	5.00	4.25	5.00	4.50
Transferability	5.00	3.25	3.00	4.50
Maintainability	4.89	3.44	3.54	4.67
Mean Score	**4.45**	**3.61**	**3.84**	**4.27**

6.3. *Domain Knowledge Level*

As in the previous case of study, at this level we evaluate *application results* and *domain knowledge coverage*. For the *application results* validation, we elaborate questionnaires, with the support of experts, related to each category of the *golden-standard*. Some examples of the SPARQL queries applied to POS, FR2013, KnowRob, and OntoSLAM are shown in the following.

(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?
(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, OntoSLAM, 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>
```

OntoSLAM shows the class `Pose` :

```
 <owl:Class rdf:about="OS:Pose">
    <rdfs:subClassOf rdf:resource="OS:AbstractThing"/>
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    It is composed by the state of the DOF of the Physical Thing
</rdfs:comment>
</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 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`. In OntoSLAM we have the relationship

`PosAtTime`. This relation associates an `OS:Position` with an `ISRO:TimePoint`, whose definititon is:

```
<owl:ObjectProperty rdf:about="OS:posAtTime">
    <rdfs:domain rdf:resource="OS:Position"/>
    <rdfs:range rdf:resource="ISRO:TimePoint"/>
</owl:ObjectProperty>
```

Table 23 shows the results of applying the questionnaires in the four ontologies. OntoSLAM is the ontology that best models the SLAM knowledge domain because it can model all the subcategories of the *golden-standard*. OntoSLAM uses FR2013 and KnowRob as reference; thus, it does not have their limitations in terms of the SLAM knowledge domain.

Table 23: Domain Knowledge level - questionnarie

Ontologies	Robot Information								Environment Mapping								Timely Inform.		Workspace Inform.		Questions answered
	a1	a2	a3	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
OntoSLAM	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%

In the *knowledge coverage* evaluation, besides POS, FR2013, OntoSLAM, and KnowRob ontologies, we consider 21 more ontologies, related to the knowledge representation for the SLAM problem. Table 24 presents all analyzed ontologies using our *golden-standard*, represented as the categorization of SLAM knowledge. The first four rows in Table 24, represent the four ontologies that were compared in the previous levels; afterward, are the other ontologies in descending order respect their publication date. There was no increasing development over time, each ontology seeks to model specifically the problem for which it was designed but does not follow any standard.

6.4. *Discussion*

From *Lexical* level comparison (see Table 14), the highest similarities are between POS/OntoSLAM with 52.8% and FR2013/OntoSLAM with 52.6% of *Linguistic Similarity (LS)*. This is because the three ontologies are extensions of the same ontologies, SUMO and CORA. However, FR2013/OntoSLAM pair has a higher *StringSim* of 43%, which indicates that they keep the same entity names, from SUMO and CORA. While for *DocSim* there is a 63% similarity between POS/OntoSLAM. *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.

Table 24: Analyzed Ontologies under *golden-standard*

Ontology	**Robot Information**					**Environment Mapping**				**Timely Inf.**		**Workspace**	
	(a)	**(b)**	**(c)**	**(d)**	**(e)**	**(a)**	**(b)**	**(c)**	**(d)**	**(a)**	**(b)**	**(a)**	**(b)**
OntoSLAM	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fortes-Rey 2013	✓		✓	✓		✓				✓		✓	
POS, Carbonera et al. 2013	✓		✓			✓						✓	
KnowRob, Tenorth and Beetz 2009	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓
ISRO, Chang et al. 2020	✓	✓	✓	✓		✓	✓	✓		✓	✓		✓
Crespo et al. 2020		✓	✓			✓	✓	✓				✓	✓
Sun, Zhang, and J. Chen 2019	✓	✓	✓	✓		✓	✓		✓			✓	✓
ADROn, Ramos, Vázquez, et al. 2018	✓	✓											
Deeken, Wiemann, and Hertzberg 2018		✓	✓			✓	✓					✓	✓
Burroughes and Gao 2016	✓	✓	✓	✓		✓	✓			✓	✓	✓	
RoboEarth, Riazuelo et al. 2015	✓	✓	✓			✓	✓	✓	✓		✓	✓	✓
ROSPlan, Cashmore et al. 2015		✓	✓				✓						
Wu, Tian, et al. 2014		✓	✓		✓	✓	✓	✓	✓			✓	
Li et al. 2013			✓			✓	✓					✓	✓
CORA, Prestes et al. 2013	✓					✓	✓						
OASys, Paull, Severac, et al. 2012	✓	✓											
PROTEUS, Dhouib et al. 2011	✓	✓	✓	✓	✓	✓	✓						
Wang and Q. Chen 2011						✓	✓	✓				✓	✓
Uncertain Ontology, Pronobis and Jensfelt 2011		✓				✓	✓	✓	✓			✓	✓
OUR-K, Lim et al. 2011	✓	✓	✓			✓	✓	✓		✓		✓	
Space Ontology, Belouaer et al. 2010						✓							
SUMO, Eid et al. 2007		✓	✓			✓	✓			✓		✓	
OMRKF, Suh et al. 2007		✓				✓	✓			✓	✓	✓	✓
Mozos et al. 2007			✓			✓	✓						✓
Robot Ontology, Schlenoff and Messina 2005	✓	✓											

At *Structural* level comparison, besides results shown in Table 15, that reflect some characteristics of the four ontologies (already mentioned in the Section 6.2), we use Falcon-AO tool to obtain the combined *Linguistic/Structural Similarity (SimLingStruc)*, as shown in Table 16. 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 OntoSLAM are 24% similar, according to the Falcon-AO metric; it means that their RDF graphs are quite different (even though, the *LS* is 52%). The second highest *SimLingStruc* is FR2013/POS pair with *LS* of 48% although the pair POS/OntoSLAM has a bigger *LS* of 52%.

Regarding the OQuaRE *Quality* evaluation, OntoSLAM and POS are the ones that showed a better performance with mean scores higher than 4 (see Table 22). It means that these ontologies comply with the requirements of this OQuaRE *Quality* model. The other two ontologies, KnowRob and FR2013, comply in a minimally acceptable way. This is because KnowRob is an ontology that models information very specific to the problem it solves (teleoperation in indoor environments for service robots), so it has a very low score in the characteristic of *Functional Adequacy*.

At *Domain Knowledge* level, based on our *golden-standard* and questionnaires, we conclude that concerning **Robot Information**, the four ontologies model well robot *kinematic* and *pose* information, but three of them 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. But only OntoSLAM can answer all the questions of this category. 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 CORA 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 and OntoSLAM, because they have classes to represent `Objects` with their dimensions, properties like color and also the relative position among them. For **Timely Information**, KnowRob and OntoSLAM model better the domain of SLAM, since robots can define their position and poses in relation to time. Finally, for the **Workspace**, the four ontologies describe their dimensional workspace, but only KnowRob and OntoSLAM are able to model specific domain entities at the application level.

Regarding the *knowledge coverage* evaluated on 24 ontologies, we identify the ontologies presented in Burroughes and Gao 2016; Eid et al. 2007; Chang et al. 2020; Riazuelo et al. 2015; Lim et al. 2011; Suh et al. 2007, KnowRob, and FR2013 model partially aspects in all categories. But only the newest OntoSLAM can model all the categories of the *golden-standard* defined. 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.

7. 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. The proposed metrics can be complemented with the OQuaRE model that allows evaluating Software Engineering characteristics in the Semantic Web domain. To demonstrate the suitability of our approach, we evaluate available ontologies in the cultural heritage and robotic domains. Thus, we show that the methodological comparative process can be *customized* for specif research interests with an appropriate *golden-standard*, allowing to determine the gaps in the domain. Additionally, the methodology allows selecting the most appropriate metrics, according to the researchers' preferences. For example, if the objective of researchers is to find or develop the best ontology in terms of quality at the structural level, he/she should use the OQuaRE metrics; however, if he/she wants to compare two existing ontologies and see how similar they are, he/she might use graph matching for this level. Our proposed methodology can be applied in the areas of ontology integration, ontology matching, and to choose the most appropriate ontology to solve a specific problem. However, the low diffusion of the ontologies source code in some areas, limits its application. Thus, it is a must to encourage researchers to make available their ontologies.

We are currently working on extending this methodology by incorporating more detailed empirical evaluation. Besides a formal evaluation, as we propose with this methodology, an empirical evaluation will provide guidelines to test the targeted ontologies in real scenarios in their respective domains.

References

Addis, Matthew et al. (2003). "SCULPTEUR: Towards a New Paradigm for Multimedia Museum Information Handling". In: *Lncs*. Springer Berlin Heidelberg, pp. 582–596.

Araújo, Cristiana et al. (June 2017). "Annotated Documents and Expanded CIDOC-CRM Ontology in the Automatic Construction of a Virtual Museum". In: *Developments and Advances in Intell. Syst. and Applications*. Springer International P., pp. 91–110.

Bandeira, J. et al. (2016). "FOCA: A methodology for ontology evaluation". In: *arXiv preprint arXiv:1612.03353*.

Belouaer, L. et al. (2010). "Ontology Based Spatial Planning for Human-Robot Interaction". In: *Symp. on Temporal Representation and Reasoning*, pp. 103–110.

Biagetti, Maria Teresa (2018). "A comparative analysis and evaluation of bibliographic ontologies". In: *Challenges and Opportunities for Knowledge Organization in the Digital Age*. Ergon-Verlag, pp. 499–510.

Brank, J., M. Grobelnik, and D.a Mladenic (2005). "A survey of ontology evaluation techniques". In: *Conf. on Data Mining and Data Warehouses*, pp. 166–170.

Brewster, C. et al. (2004). "Data driven ontology evaluation". In: *Language Resources and Evaluation*. Web & Internet Science, South. Wireless Group, pp. 1–4.

Burroughes, G. and Y. Gao (2016). "Ontology-Based Self-Reconfiguring Guidance, Navigation, and Control for Planetary Rovers". In: *J. Aerospace Inf. Sys.* 13.

Burton-Jones, A., V. C. Storey, et al. (2005). "A semiotic metrics suite for assessing the quality of ontologies". In: *Data & Knowledge Eng.* 55.1, pp. 84–102.

Carbonera, J. L. et al. (2013). "Defining positioning in a core ontology for robotics". In: *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 1867–1872.

Cardinale, Yudith et al. (2020). "A Methodological Approach to Compare Ontologies: Proposal and Application for SLAM Ontologies". In: *22nd Int. Conf. on Information Integration and Web-based Applications & Services*, pp. 223–233. DOI: 10.1145/3428757.3429091.

Cashmore, M. et al. (Jan. 2015). "Rosplan: Planning in the robot operating system". In: 2015, pp. 333–341.

Chang, D. S. et al. (2020). "Ontology-based knowledge model for human-robot interactive services". In: *ACM Symp. on Applied Computing*, pp. 2029–2038.

Chanhom, Weeraphan and Chutiporn Anutariya (2019). "TOMS: A Linked Open Data System for Collaboration and Distribution of Cultural Heritage Artifact Collections of National Museums in Thailand". In: *New Generation Computing* 37.4, pp. 479–498.

Cornejo, M. et al. (2020). “A categorization of simultaneous localization and mapping knowledge for mobile robots”. In: *ACM Symp. on Applied Computing*, pp. 956–963.

Cornejo-Lupa, María A et al. (2020). “A Survey of Ontologies for Simultaneous Localization and Mapping in Mobile Robots”. In: *ACM Computing Surveys (CSUR)* 53.5, pp. 1–26.

Crespo, J. et al. (2020). “Semantic Information for Robot Navigation: A Survey”. In: *Applied Sciences* 10.2, p. 497.

Deeken, H., T. Wiemann, and J. Hertzberg (2018). “Grounding semantic maps in spatial databases”. In: *Robotics and Autonomous Systems* 105, pp. 146–165.

Degbelo, A. (2017). “A snapshot of ontology evaluation criteria and strategies”. In: *Int. Conf. on Semantic Systems*, pp. 1–8.

Dhouib, S. et al. (2011). “Control architecture concepts and properties of an ontology devoted to exchanges in mobile robotics”. In: *Control Architectures of Robot*.

Doerr, Martin (2005). “The CIDOC CRM, an Ontological Approach to Schema Heterogeneity”. In: *Semantic Interoperability and Integration*. Ed. by Y. Kalfoglou et al. Dagstuhl Seminar Proceedings 04391. Dagstuhl, Germany: Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

Doerr, Martin, Stefan Gradmann, et al. (2010). “The europeana data model (EDM)”. In: *World Library and Information Congress: 76th IFLA general conf. and assembly*. Vol. 10, p. 15. URL: https://pro.europeana.eu/page/edm-documentation.

Doerr, Martin and Maria Theodoridou (2014). *CRMdig an extension of CIDOC-CRM to support provenance metadata*. Tech. rep. Technical Report 3.2. Heraklion: ICS-FORTH.

Drakopoulos, Georgios, Yorghos Voutos, and Phivos Mylonas (2020). “Recent Advances On Ontology Similarity Metrics: A Survey”. In: *2020 5th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conf. (SEEDA-CECNSM)*. IEEE, pp. 1–7.

Duque-Ramos, A. et al. (May 2011). “OQuaRE: A SQuaRE-based approach for evaluating the quality of ontologies”. In: *Journal of Research and Practice in Information Technology* 43, pp. 159–176.

Eid, M. et al. (2007). “A Universal Ontology for Sensor Networks Data”. In: *Computational Intelligence for Measurement Systems and Applications*, pp. 59–62.

FINTO (2015). *Finish Thesaurus and ontology service*. Accessed: 6-5-2020. URL: http://finto.fi/en/.

Fortes, V. (2013). “A positioning ontology for C-SLAM”. In: *Monograph – UFRGS - Curso de Bacharelado em Ciencia da Computacao*, pp. 23–33.

Freire, Nuno and Diogo Proença (2020). “RDF Reasoning on Large Ontologies: A Study on Cultural Heritage and Wikidata”. In: *IFIP Int. Conf. on Artificial Intelligence Applications and Innovations*. Springer, pp. 381–393.

Freire, Nuno and Sjors de Valk (2019). “Automated interpretability of linked data ontologies:: an evaluation within the cultural heritage domain”. In: *IEEE Int. Conf. on Big Data (Big Data)*. IEEE, pp. 3072–3079.

Hajmoosaei, Abdolreza and Petra Skoric (2016). “Museum Ontology-Based Metadata”. In: *Int. Conf. on Semantic Computing (ICSC)*. IEEE.

Hlomani, H. and D. Stacey (2014). “Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey”. In: *Semantic Web Journal* 1.5.

Hyvönen, Eero et al. (2005). “MuseumFinland—Finnish museums on the semantic web”. In: *Journal of Web Semantics* 2-3, pp. 224–241.

Jian, N. et al. (2005). “Falcon-AO: Aligning Ontologies with Falcon”. In: *Workshop on Integrating Ontologies*, pp. 87–93.

Keil, J. M. and S. Schindler (2019). “Comparison and evaluation of ontologies for units of measurement”. In: *Semantic Web* 10.1, pp. 33–51.

Keil, Jan Martin (2020). “ABECTO: An ABox Evaluation and Comparison Tool for Ontologies”. In: *European Semantic Web Conference*. Springer, pp. 140–145.

Lantow, B. (2016). “OntoMetrics: Application of On-line Ontology Metric Calculation”. In: *BIR Workshops*. Vol. 1684, pp. 1–12.

Li, R. et al. (2013). “Multi-layered map based navigation and interaction for an intelligent wheelchair”. In: *Robotics and Biomimetics*, pp. 115–120.

Lim, G. H. et al. (2011). “Ontology-Based Unified Robot Knowledge for Service Robots in Indoor Environments”. In: *Systems, Man, and Cybernetics* 41.3. ISSN: 1083-4427.

Maedche, A. and S. Staab (2002). “Measuring Similarity between Ontologies”. In: *Knowledge Eng. and Knowledge Manag.: Ontologies and the Semantic Web*. Springer. ISBN: 978-3-540-45810-4.

Maillot, P. and C. Bobed (2018). “Measuring structural similarity between RDF graphs”. In: *ACM Symp. on Applied Computing*, pp. 1960–1967.

Marchenkov, Sergey A et al. (Nov. 2017). “Smart museum of everyday life history in Petrozavodsk State University: Software design and implementation of the semantic layer”. In: *21st Conf. of Open Innovations Association (FRUCT)*. Vol. 562. IEEE, pp. 224–230. DOI: 10.23919/FRUCT.2017.8250186.

Matuszek, C. et al. (2006). “An introduction to the syntax and content of Cyc”. In: *UMBC Computer Science and Electrical Eng. Depart. Collection*.

McDaniel, M. and V. C. Storey (2019). “Evaluating Domain Ontologies: Clarification, Classification, and Challenges”. In: *ACM Computing Surveys* 52.4, pp. 1–44.

Mozos, O. M. et al. (2007). “From Labels to Semantics: An Integrated System for Conceptual Spatial Representations of Indoor Environments for Mobile Robots”. In:

Mulholland, Paul, Annika Wolff, and Trevor Collins (2012). "Curate and Storyspace: An Ontology and Web-Based Environment for Describing Curatorial Narratives". In: *Extended Semantic Web Conference*. Springer Berlin Heidelberg, pp. 748–762.

Pak, J. and L. Zhou (2009). "A framework for ontology evaluation". In: *Workshop on E-Business*. Springer, pp. 10–18.

Paull, L., G. Severac, et al. (2012). "Towards an Ontology for Autonomous Robots". In: *Conf. on Intelligent Robots and Systems*, pp. 1359–1364.

Pinto-De la Gala, Alexander et al. (2021). "Towards an Ontology for Urban Tourism". In: *Proc. of the 36th Annual ACM Symp. on Applied Computing*. SAC '21. Virtual Event, Republic of Korea, pp. 1887–1890. URL: https://doi.org/10.1145/3412841.3442142.

Porzel, R. and R. Malaka (2004). "A task-based approach for ontology evaluation". In: *Ontology Learning and Population, Valencia, Spain*. Citeseer, pp. 1–6.

Poveda-Villalón, M. et al. (2012). "Validating ontologies with Oops!" In: *Conf. on knowledge Eng. and Knowledge Management*. Springer, pp. 267–281.

Prestes, E. et al. (2013). "Towards a core ontology for robotics and automation". In: *Robotics and Autonomous Systems* 61.11, pp. 1193–1204. ISSN: 0921-8890.

Pronobis, A. and P. Jensfelt (2011). "Multi-modal semantic mapping". In: *RSS Workshop on Grounding Human-Robot Dialog for Spatial Tasks*.

Ramos, F., A. Vázquez, et al. (2018). "Ontology based design, control and programming of modular robots". In: *Integrated Computer-Aided Eng.* 25.2, pp. 173–192.

Riazuelo, L. et al. (2015). "RoboEarth Semantic Mapping: A Cloud Enabled Knowledge-Based Approach". In: *IEEE Trans. on Automation Science and Eng.* 12.2, pp. 432–443.

Roldan-Molina, Gabriela R et al. (2020). "Ontology fixing by using software engineering technology". In: *Applied Sciences* 10.18, p. 6328.

Sathiya, B. and T. Geetha (2019). "A review on semantic similarity measures for ontology". In: *J. of Intelligent & Fuzzy Systems* 36.4, pp. 3045–3059.

Schlenoff, C., S. Foufou, and S. Balakirsky (2012). "Performance evaluation of robotic knowledge representation (PERK)". In: *Performance Metrics for IS*.

Schlenoff, C. and E. Messina (2005). "A Robot Ontology for Urban Search and Rescue". In: *W. on Research in Knowld. Represent. for Autonomous Systems*, pp. 27–34. ISBN: 1-59593-202-X.

Suh, I. H. et al. (2007). "Ontology-based multi-layered robot knowledge framework (OMRKF) for robot intelligence". In: *Intelligent Robots and Sys.* Pp. 429–436.

Sun, X., Y. Zhang, and J. Chen (2019). "High-Level Smart Decision Making of a Robot Based on Ontology in a Search and Rescue Scenario". In: *Future Internet* 11.11, p. 230.

Tartir, S. et al. (2005). "OntoQA: Metric-based ontology quality analysis". In: *The Ohio Center of Excellence in Knowledge-Enabled Computing (Kno.e.sis) at CORE Scholar*.

Tenorth, M. and M. Beetz (2009). "KNOWROB: knowledge processing for autonomous personal robots". In: *RSJ on Intelligent Robots and Systems*, pp. 4261–4266.

Thiéblin, Elodie et al. (2020). "Survey on complex ontology matching". In: *Semantic Web* 11.4, pp. 689–727.

Thrun, S. (2003). "Exploring artificial intelligence in the New Millennium". In: chap. Robotic Mapping: A Survey, pp. 1–35.

Turco, Massimiliano Lo, Michele Calvano, and Elisabetta C Giovannini (Feb. 2019). "Data modeling for museum collections". In: *Int. Archives of the Photogrammetry, Remote Sensing and Spatial Inf. Sc.* XLII-2/W9, pp. 433–440. DOI: `10.5194/isprs-archives-XLII-2-W9-433-2019`.

Ulanov, A., G. Shevlyakov, et al. (2010). "Monte Carlo study of taxonomy evaluation". In: *Database and Expert Systems Applications.* IEEE, pp. 164–168.

Wang, T. and Q. Chen (June 2011). "Object semantic map representation for indoor mobile robots". In: *Int. Conf. on System Science and Eng.* Pp. 309–313.

Wu, H., G. Tian, et al. (2014). "Spatial semantic hybrid map building and application of mobile service robot". In: *Robotics and Autonomous Systems* 62.6, pp. 923–941.

Xue, Xingsi et al. (2021). "Optimizing ontology alignment through linkage learning on entity correspondences". In: *Complexity* 2021.

Zander, S. et al. (2016). "A Survey about the Usage of Semantic Technologies for the Description of Robotic Components and Capabilities." In: *SAMI@ iKNOW.*