Analysis of Ontology Quality Dimensions, Criteria and Metrics

R. S. I. Wilson[1,2]([✉]) [ID], J. S. Goonetillake[2] [ID], W. A. Indika[3] [ID], and Athula Ginige[4] [ID]

[1] Uva Wellassa University, Badulla, Sri Lanka
shyama@uwu.ac.lk
[2] University of Colombo School of Computing, Colombo 07, Sri Lanka
jsg@ucsc.cmb.ac.lk
[3] University of Ruhuna, Matara, Sri Lanka
waindika@dcs.ruh.ac.lk
[4] Western Sydney University, Penrith, NSW, Australia
a.Ginige@westernsydney.edu.au

Abstract. Ontology quality assessment needs to be performed across the ontology development life cycle to ensure that the ontology being modeled meets the intended purpose. To this end, a set of quality criteria and metrics provides a basis to assess the quality with respect to the quality requirements. However, the existing criteria and metrics defined in the literature so far are messy and vague. Thus, it is difficult to determine what set of criteria and measures would be applicable to assess the quality of an ontology for the intended purpose. Moreover, there are no well-accepted methodologies for ontology quality assessment as the way it is in the software engineering discipline. Therefore, a comprehensive review was performed to identify the existing contribution on ontology quality criteria and metrics. As a result, it was identified that the existing criteria can be classified under five dimensions namely syntactic, structural, semantic, pragmatic, and social. Moreover, a matrix with ontology levels, approaches, and criteria/metrics was presented to guide the researchers when they perform a quality assessment.

Keywords: Ontology quality assessments · Quality criteria · Quality metrics · Quality dimensions · Ontology evaluation

1 Introduction

Ontology is "*a formal, explicit specification of a shared conceptualization*" [1] that has numerous capabilities such as *analyzing domain knowledge, making available implicit knowledge explicit, sharing a common understanding of the structure of information among people or software agents* [2]. Thus, ontologies are incorporated in information systems as a component to manage heterogeneous information and high-volume data in domains like medicine, agriculture, defense, and finance. This supports information consumers or software agents to make the right decisions to tackle practical problems. However, the right decisions depend on the quality of the information provided, in the case of ontology-based information systems, the quality of the ontology.

Ontology quality is a promising research area in the semantic web that has been discussed under ontology evaluation [3]. Ontology quality assessment is useful for ontology consumers to select a suitable ontology from a set of ontologies or assess the fitness of an ontology for an intended purpose [4, 5]. Moreover, quality assessment should not be limited to evaluate the product at the final stage. Thus, an ontology is also needed to be evaluated across the entire ontology life cycle. An ontology consists of levels namely syntactic, vocabulary (i.e., terminology), architecture, semantic, and context, also named as layers [6, 7]. The evaluation of ontology can be considered with respect to each ontology level/layer to reduce the complexity of the overall ontology quality assessments.

However, building a good quality ontology is not straightforward as it requires to consider several aspects such as logic, reasoning, structure, domain knowledge to be modeled concerning the specified tasks [8, 9]. To this end, ontology quality criteria are important to assess the components of an ontology that may have several measures (i.e., metrics) which provide an objective and a quantitative basis insight for ontology developers. Then, they can understand the areas to be revised to achieve a good quality of an ontology. Firstly, the authors of the article [6] have proposed a significant set of ontology quality criteria (i.e., characteristics) such as *correctness, soundness, consistency, completeness, conciseness, expandability,* and *sensitiveness.* Later, many more criteria have been added to this list by scholars from different points of view [3, 10–14]. For instance, the scholars of the article [3] have proposed a set of criteria by considering ontology as a software artifact. They have adopted the standard ISO/IEC 25000:2005 titled SQuaRE [15] and have suggested the criteria: *functional adequacy, reliability, performance efficiency, operability, maintainability, capability, transferability,* and *structural.* OntoQA is an approach that describes eleven (11) quality measures that can be used to evaluate the quality of an ontology at schema and data (i.e., knowledge bases) levels [14]. Schema level measures are *the richness of relationships, inheritance,* and *attributes.* The measures: *Class richness, average population, cohesion, fullness, connectivity* are classified as data level measures. Furthermore, OntoMetric is another web-based tool that assesses ontology quality under five criteria: *basic, schema, graph, knowledge,* and *class* which include 160 measures [13].

At present, many quality criteria and measures have been defined to assess the quality of an ontology from different perspectives such as ontology perspectives (i.e., inherent ontology quality), real-world perspectives, and users' perspectives [8, 13, 14, 16]. However, all these criteria and measures defined in the literature so far are messy and vague. For example, it is difficult to understand the quality criteria and thus the quality measures relevant to a given criterion in most cases as the terminology has been used inattentively. There is no distinction made between the two concepts quality criteria and quality measure from an ontology quality point of view. Moreover, the ontology has many levels as described in Sect. 4.1 and the quality of each level is significant for the overall quality of the ontology. Even if the quality of a certain level has been discussed in the literature in a very ad hoc way none of the existing definitions or approaches have defined quality criteria or measures in a methodical way for all of the levels. Consequently, no proper guidelines exist so far for ontology quality evaluation as the way it is with software engineering. For instance, when an ontology is evaluated through

an application-based approach, it is necessary to understand what quality criteria to be adopted at the contextual level, semantic level, and structural level. Currently, ontology researchers and practitioners limit the quality evaluation only to a certain set of criteria namely *expressiveness* and *usefulness* due to the nonexistence of proper guidelines [16].

Nevertheless, our effort to analyze the quality criteria and measures that have been identified in the previous studies and synthesize them to provide an overview with ontology levels and approaches in order to produce a good quality ontology. To achieve this aim, data quality theories have been adopted. This would guide ontology developers and researchers to understand what quality criteria are to be assessed in each level (i.e., layer) and what the possible approaches would be to evaluate the ontologies.

2   Related Work

We analyzed the existing survey studies which have focused on ontology evaluation criteria, metrics, and approaches. Among them, a countable number of survey studies [16–20] were reviewed the related works comprehensively or systematically. However, none of them have provided a model or matrix, or overview among quality criteria, approaches, and ontology levels. This has caused a difficulty for researchers to gain insight on what quality criteria to be considered when performing ontology evaluation and what criteria would be more appropriate to assess each level of an ontology.

The author of the article [17] has highlighted the important quality criteria: *consistency*, *completeness*, *conciseness*, *expandability*, and *sensitiveness* through theoretical analysis and based on her experience. Then, a set of errors that can be made by ontology developers, have been classified under each quality criterion except *expandability* and *sensitiveness*. Finally, it has presented the ways of detecting *inconsistency*, *incompleteness*, and *redundancy*. Moreover, the work has also highlighted the requirement of developing language-dependent evaluation tools and the importance of documenting ontology quality with criteria. The research article [18] has considered the automatic, domain- and task-independent ontology evaluation as the scope of the study and has also focused on a set of ontology quality criteria that have been explained in five articles including [11, 17]. Furthermore, the evaluation of each ontology level with related measures has been described. For instance, the structure level can be evaluated by considering sub-graphs: *depth*, *breadth*, and *fan-outness*, and the context level can be assessed with competency questions, or through unit tests. Nevertheless, there is no clear comparison or discussion of ontology criteria and how those can be associated in each level when evaluating an ontology.

There are fifty-one structural quality measures that have been explored in [19]. Since the definition in natural language can be interpreted by different researchers from different perspectives and the paper has constructed formal definitions for each measure of quality criteria to provide a common understanding. Thus, the authors of the article [19] have addressed that issue by introducing formal definitions based on the Ontology Definition Model (ODM) and have presented the formal definitions for the quality measures: *Richness*, *Cohesion*, *Class Importance*, *Fullness*, *Coupling, Class Connectivity*, *Class Readability*. These formal definitions support researchers to compare the definitions and intents of measures when evaluating the structure of the ontology.

As the ontology quality assessment is required in each stage of the ontology life cycle, it is vital to aware of what criteria to be considered in each stage. Thus, the researcher of the article [16], has explored the quality criteria which are relevant for the evaluation of design and implementation stages. For that, a systematic review has been performed by retrieving articles from two reputed journals: the Journal of Web Semantics and the Semantic Web Journal. As the author has explored, *accuracy*, *adaptability*, *cognitive adequacy*, *completeness*, *conciseness*, *consistency*, *expressiveness*, and *grounding* as relevant criteria for evaluating the ontology in the design stage. To evaluate the quality of an ontology in the implementation stage, the criteria: *computational efficiency*, *congruency*, *practical usefulness*, *precision*, and *recall* have been recommended. Moreover, it has been revealed that few quality criteria such as *expressiveness* and *practical usefulness* have been used in practice though there are many quality criteria defined in theoretical approaches [16].

As a diverse set of ontology quality criteria exist, it is difficult for researchers to find a suitable set of quality criteria for assessing a particular ontology based on the intended purpose. To mitigate this issue, scholars have adopted well-defined theories and standards in the software engineering discipline [3, 20]. In the article [20], the authors have conducted a systematic review to identify the ontology quality criteria and grouped the measures of quality criteria into categories namely Inherent and Inherent-System, which have been defined in ISO/IEC 25012 Data Quality Standard. The adapted inherent quality criteria from this standard are *accuracy*, *completeness*, *consistency*, and *currentness*. The inherent-system criteria are *compliance*, *understandability*, *availability*. Under these criteria, the ontology measures identified through the survey have been mapped. For instance, the *accuracy* criterion includes the measures: *Incorrect Relationship, Hierarchy Over-specialization, Class Precision, Number of Deprecated Classes* and *Properties*, etc. and the *completeness* criterion includes the measures: *Number of Isolated Elements, Missing Domain or Range in Properties, Class Coverage,* and *Relation Coverage*. However, this classification can be applied to compare the quality of two or more ontologies in a similar domain, but it is not sufficient to assess a single ontology to get an idea on which components (i.e., levels) of an ontology have good quality and which are needed to be improved.

Moreover, the scholars [7, 21–25] have discussed several ontology evaluation approaches, criteria and ontology levels to be focused on when assessing ontology quality. Only the researchers of the articles [22, 23] have attempted to provide a comparison between the ontology quality approaches and criteria. However, the comparisons are abstract and difficult to interpret. The authors in [23] have stated that it is difficult to associate criteria with ontology approaches and ontology levels due to their diversity [23]. According to our study, a reason for having several criteria is due to the availability of different definitions to the same criteria, or vice versa (i.e., two or more closely related criteria may have the same definitions). This issue is further discussed in Sect. 4.2. Thus, we made an effort to carefully analyze these different definitions and to specify possible ontology quality criteria related to approaches, and levels (i.e., layers). To this end, a comprehensive theoretical analysis was conducted on ontology quality criteria and metrics (i.e., measures).

3 Methodology

To address the gap highlighted in Sects. 1 and 2, we performed a theoretical review by following the procedure proposed in [26]. To this end, the relevant background and gaps to be addressed have been explained in the previous sections. As the next step, the search terms to find the relevant papers from the databases: ACM Digital Library, IEEE Xplore Digital Library, Science Direct, Springer Link, and the search engine: Google Scholar, were defined. They are *ontology*, *ontology quality criteria*, *measures, metrics, quality assessment*, and *ontology evaluation.* Then, the general search strategy was developed to perform a search on the databases which are "*[ontology AND [Quality OR Evaluation OR Assessment] AND [Criteria OR Measures OR Metric]]*". At this stage, the articles were filtered purposefully by analyzing titles and abstracts as the study intention is not to explore the state-of-art in ontology quality assessment, but to analyze the ontology criteria which have been covered through the ontology levels, possibly with approaches. To reduce the searching results and to retrieve quality studies, the inclusion criteria such as;

- studies in English,
- studies published during (2010–2021),
- peer-reviewed,
- full-papers, and
- studies focused on quality assessment, criteria, and measures

have been applied. Finally, the relevant articles were downloaded through the reference management tool (i.e., Mendeley). Moreover, few potential articles were retrieved by looking up the references of the filtered articles. Thereafter, we selected the articles which are [3, 6, 10, 11, 27–35] for the analysis.

4 Data Analysis and Synthesis

4.1 Prerequisite

In previous studies, the following terms have been used interchangeably in their explanations of ontology quality. In this study, mostly we use the terms: criteria (i.e., characteristics), metrics (i.e., measures), dimensions, ontology levels, and ontology approaches to describe the theories in order to maintain consistency.

Criteria (i.e., Characteristics)
Ontology criteria (i.e., characteristics) describe a set of attributes. An attribute is a measurable physical or abstract property of an entity [36, 37]. In ontology quality, an entity can be a set of concepts, properties, or an ontology.

Metric (i.e., Measure)
Metric (i.e., measure) describes an attribute quantitatively or defines an attribute formally [38]. In other words, the ontology quality metric is used to measure the characteristic of an ontology that can be represented formally.

For instance, the conceptual complexity is a criterion (i.e., characteristic) that is used to evaluate an ontology (i.e., entity) and it can be quantitatively measured by using the metric: size, which may have measurements such as number of concepts and properties in the structure, number of leaf concepts and number of attributes per concepts.

Dimension (i.e., Aspects)

Dimensions (i.e., aspects) have been defined to classify several criteria/attributes based on different views. For example, if a dimension describes the content of an ontology, that may include a set of criteria related to the content assessments. The criteria like graph complexity, modularity, and graph consistency can be grouped into the structural dimension.

Therefore, similar to the software data quality, dimensions are qualitative and associate with several characteristics (i.e., attributes) that can be directly or indirectly measured through quantitative metrics.

Hereinafter, we use terms: *criteria* and *metric* instead of using the terms *ontology quality criteria* and *ontology quality metrics* respectively.

Ontology Levels (i.e., Layers)

In the ontology quality assessment, initially, three levels (also known as layers) to be focused on have been proposed in [6], namely: *content*, *syntactic* & *lexicon*, and *architecture*. Later, this was expanded by including *structural* and *context* [7, 18]. These levels/layers focus on different aspects of ontological information. The syntactic level considers the features related to the formal language that is used to represent the ontology. The lexicon level is also named vocabulary or data layer that takes into account the vocabulary that is used to describe concepts, properties, instances, and facts. The structural level/architectural layer focuses on the is-a relationship (i.e., hierarchical) which is more important in the ontology modeling against the other relations. Moreover, it considers the design principles and other structural features required to represent the ontology. Other non-hierarchical relationships and semantic elements are considered under the semantic level. The context level concerns the application level that the ontology is built for. It is important to assess whether the ontology confirms the real application requirements as a component of an information system or a part of a collection of ontologies [7, 39].

Ontology Evaluation Approaches (i.e., Methods, Techniques)

Mainly, ontology evaluation has been conducted under four approaches: application-based, data-driven-based, golden standard-based, and human-based [7]. In brief, the application-based approach: assesses the ontology when it is attached with the application and used in practice [39]. The data-driven approach: assesses the ontology against the data source (i.e., corpus) that is used for the ontology modeling. The golden-standard approach: compares the candidate ontology with the ontology that has the agreed quality or assesses the ontology with a benchmark/a vocabulary defined by experts. The human-based approach: assesses the ontology with the intervention of domain experts and ontology engineers based on the set of criteria, requirements, and standards [6, 7].

Table 1. The existing quality models for ontologies

Citation	Dimensions	Metric/attributes
Burton-Jones et al. 2005 [10]	Syntactic	Lawfulness, richness
	Semantic	Interpretability, consistency, clarity
	Pragmatic	Comprehensiveness, accuracy, relevance
	Social	Authority, history
Gangemi et al. 2006 [11]	Structural	Size, modularity, depth, breadth, tangledness, etc.
	Functional	precision, recall (coverage), Accuracy
	Usability-related	recognition, efficiency, and interfacing annotations
Zhu et al. 2017 [30]	Content	Correctness of entities, Semantic coverage, Vocabulary coverage
	Presentation	Size, relation, modularity: cohesion/coupling, non-redundancy
	Usage	Search efficiency, description complexity, definability, extendibility, tailorability, etc.

4.2 Ontology Quality Dimensions, Criteria, and Metrics

The ontology quality evaluation throughout the ontology life cycle ensures that good quality ontology is being developed. However, a major issue is the unavailability of an agreed methodology for it. As a result, several criteria and metrics have been defined without a strong theoretical foundation. According to our analysis, we were able to identify a set of criteria and metrics that were presented in Table 2. The related measurements have not been mentioned as all together hundreds of measurements are available. Thus, only the relevant citations have been provided for further references. Moreover, the metrics definitions similar to the ones in [11] and [3] can also be found in [12] and [40] respectively.

There are significant attempts in the literature to introduce generalized dimensions to classify criteria and metrics (see Table 1). The authors of the article [10] have introduced quality dimensions: *syntactic*, *semantic*, *pragmatic*, and *social* by adopting a semiotic framework named the semiotic metric suit. In the article [11], the researchers have classified metrics into three dimensions: *structural*, *functional*, and *usability-related*. In addition to that, the scholars in the research [30] have introduced a set of dimensions namely *presentation*, *content*, and *usage* considering the web service domain. When analyzing the metrics defined under each dimension, it has been recognized that the proposed dimensions are overlapping. For instance, the criteria: *modularity*, *size: concept/relations* defined under the structural dimension in [11] also appear in the presentation dimension in [30]. Moreover, the criteria in the dimensions: *semantic* [10], *functional* [11], and *content* [30] have been defined concerning the domain that the ontology being modeled. The *pragmatic* [10], *usability-related* [11], and *usage* [30] dimensions consider

the quality when an ontology is at the application level and to this end, criteria such as *functional accuracy*, *relevancy*, *adaptability*, *efficiency,* and *comprehensibility* have been considered. Based on that, we have identified five main distinguish dimensions that the criteria can be grouped such as *syntactic, structural, semantic, pragmatic,* and *social*. This can be seen as an extended version of the semiotic metric suit [10]. Then, taking this as a basis, the identified criteria and metrics in the literature were mapped by analyzing the given definitions (see Table 2).

Classification of Criteria and Metrics

After a thorough analysis, we identified fourteen main ontology criteria namely *syntactic correctness*, *cognitive complexity, conciseness, modularity, consistency, accuracy, completeness, adaptability, applicability, efficiency, understandability, relevance, usability,* and *accessibility* which are classified as follows (the possible metrics are in the italic format).

– Syntactic: describes the conformance to the rules of the language that the ontology is written [10, 30]

  • Syntactic correctness: *lawfulness, richness*

– Structural: describes the topological and logical properties of an ontology [11]

  • Cognitive Complexity: *size, depth, breadth, fan-outness*, Modularity: *cohesion, coupling*, Internal Consistency: *tangledness, circularity, partition*

– Semantic: describes the characteristics related to the semantic (meanings) of an ontology [10, 34].

  • Conciseness: *precision*, Coverage: *recall*, External Consistency: *clarity, interpretability*

– Pragmatic: describes the appropriateness of an ontology for an intended purpose/s

  • Functional Completeness: *competency questions, precision*, Accuracy, Adaptability, Applicability, Efficiency, Understandability, Relevance, Usability: *ease of use*, Accessibility

– Social: describes the characteristics related to ontology quality in use (user-satisfaction/ social acceptance)

  • *recognition, authority, history*

Few of the criteria can be further decomposed into sub-criteria related to different perspectives: inherent to the ontology (i.e., ontology perspective), domain-depend (i.e., real-world perspective), and user-depend (i.e., user perspective). For instance, there are two types of consistency: *internal consistency* and *external consistency* [30].

Table 2. Associated ontology quality dimensions of criteria, and metrics

Dimensions	Criteria (i.e., characteristics)	Metrics	Citations	*Levels
Syntactic	Correctness: syntactic correctness	Lawfulness/well-formedness	[10, 30]	SY
Syntactic	Correctness: syntactic correctness	Richness	[10]	SY
Structural	Cognitive complexity: comprehensiveness	Size	[27, 29]	ST, L
Structural	Modularity	Coupling	[30, 35]	ST
Structural	Modularity	Cohesion	[3, 30, 35]	ST, S
Structural	Cognitive complexity	Depth	[12]	ST
Structural	Cognitive complexity	Breadth	[12]	ST
Structural	Cognitive complexity	Fan-outness	[12]	ST
Structural	Internal consistency	Tangledness	[12]	ST, S
Structural	Internal consistency	Circularity	[12, 17, 27]	ST, S
Structural	Internal consistency	Partition error	[6]	ST, S
Semantic	Conciseness	Precision	[3, 30, 33]	ST, S
Semantic	External consistency: semantic correctness	Clarity	[10, 17, 34]	S
Semantic	External consistency: semantic correctness	Interpretability	[17, 33]	S
Semantic	Coverage: semantic completeness	Inferred vocabulary coverage (recall)	[30]	S
Semantic	Coverage: syntactic completeness	Vocabulary coverage (recall)	[30]	L
Pragmatic	Functional completeness	Competency questions/precision	[34]	C
Pragmatic	Accuracy	–	[10, 34]	C
Pragmatic	Adaptability	–	[30, 34]	C
Pragmatic	Applicability	–	[30]	C
Pragmatic	Efficiency	–	[30]	C
Pragmatic	Understandability	–	[30]	C
Pragmatic	Relevance	–	[10, 34]	C
Pragmatic	Usability	Ease of use	[3, 34]	C
Pragmatic	Commercial accessibility: fitness	-	[18]	

(*continued*)

Table 2. (*continued*)

Dimensions	Criteria (i.e., characteristics)	Metrics	Citations	*Levels
Social	–	Recognition	[34]	C
Social	–	Authority	[10]	C
Social	–	History	[10]	C

*Levels are represented in symbols: syntactic - SY, structural - ST, lexical - L, semantic - S and context – C

Internal consistency is an inherent characteristic of ontologies that considers whether there is any self-contradiction within the ontology (i.e., ontology perspective) [18, 30]. In the article [17], the authors have classified three inconsistencies in this regard as *circularity errors, partition errors*, and *semantic errors*. Circularity and partition errors describe the logical inconsistencies related to the structure and the relations of an ontology, thus, both are inherent to the ontology. Tangledness that has been described in [11, 12] also considers as a measure of internal consistency of the structure as tangledness occurs when a class has multiple parent classes.

To determine the *semantic correctness* (i.e., semantic errors), it is necessary to consider the domain knowledge that the ontology used to specify the conceptualization. Thus, it comes under the *external inconsistency* that considers the consistency from the real-world perspective. Furthermore, *clarity* and *interpretability* can also be considered as metrics of *semantic correctness* [33].

Moreover, the definitions are given for the criteria: *completeness*, *coverage*, and *expressiveness* are closely related (see Table 3). Based on the *completeness* definition given in [41, 42] for Data Quality, we identified that ontology completeness also can be further decomposed as *coverage: syntactic completeness, coverage: semantic completeness*, and *functional completeness* considering different perspectives. For instance, *coverage* (or the semantic completeness) describes the completeness from the real-world perspective that determines the degree of covered entities (i.e., concepts, relation, attribute, instances) in the domain [17, 28, 33]. Moreover, the measures: *missing instances, missing properties, isolated relations*, and *incomplete formats* are also used to assess *coverage* from the ontological inherent point of view, which can be detected without domain knowledge. Thus, it is named *syntactic completeness*. In addition to that, we defined the criteria: *functional completeness* concerning the user perspective, in which the completeness is measured considering whether the ontology provides complete answers for the users' queries (i.e., competency questions), which is more subjective and difficult to measure. In the case of data-driven ontologies, functional completeness is measured against the corpus that the ontology to be covered.

Table 3. The definition is given for the criteria and metrics in the previous studies.

Criteria (C)/Metrics (M)	Definition
Coverage (C)	"*Measures how well the candidate ontology covers the terms extracted from the corpus*" [28]
Correctness (C)	"*Whether the ontology accurately represents the knowledge of a subject domain*" [30]
Syntactically correct (C)	"*Evaluates the quality of the ontology according to the way it is written*" [33]
Completeness (C)	"*Whether the ontology represents all of the knowledge of a subject domain*" [30] "*Refers to the extension, degree, amount or coverage to which the information in a user-independent ontology covers the information of the real world*" [17]. "*the level of granularity agreed*" [33]
Compatibility (C)	"*Whether the ontology contains junk, i.e., contents not in the subject domain*" [30]
Internal consistency (C)	"*Whether there is no self-contradiction within the ontology. (cannot measure)*" [30] "*the formal and informal definition have the same meaning*". [6]
External consistency (C)	"*Whether the ontology is consistent with the subject domain knowledge*" [30]
Syntactic completeness (C)	"*How much the vocabulary of the ontology matches exactly that of the standard*" [30]
Semantic completeness (C)	"*How much the vocabulary of the standard can be derived from the ontology*" [30]
Expressiveness (C)	"*Refers to an ontology's degree of detail*" [9]
Well formedness (M)	"*Syntactic correctness with respect to the rules of the language in which it is written*" [30]
Conciseness (C)	"*The key attribute for this is the lack of redundancy within the ontology*" [30] "*All the knowledge items of the ontology must be informative, so that non-informative items should be removed*" [3] "*An ontology is concise if it does not store any unnecessary or useless definitions if explicit redundancies do not exist between definitions, and redundancies cannot be inferred using other definitions and axioms*" [17, 33]
Structural complexity (C)	The cardinality of ontology elements: classes, individuals, properties

(*continued*)

Table 3. (*continued*)

Criteria (C)/Metrics (M)	Definition
Modularity (C)	"*How well the ontology is decomposed into smaller parts, to make it easier to understand, use, and maintain*" [30]
Cohesion (M)	"*The degree of interaction within one module or ontology*". [30]
Coupling (M)	"*The degree of cross-referencing*" [30]
Tangledness (M)	"*This measures the distribution of multiple parent categories*" [3]
Cycles (M)	"*The existence of cycles through a particular semantic relation*" [3], "*occurs when a class is defined as a specialization or generalization of itself*" [6]
Applicability (C)	"*This relates to whether the ontology is easy to apply for a specific task, which in this case is the description of web service semantics*" [30]
Adaptability (C)	"*How easily the ontology can be changed to meet the specific purposes of developing a particular web service*" [30]
Efficiency (C)	"*How easily semantic information can be processed for various purposes*" [30]
Comprehensibility (C)	"*Whether human readers can easily understand the semantic description*" [30]
Representation correctness (C)	"*Evaluate the quality of mappings of entities, relations, and features into the elements of the ontology were specified*" [33]
Structural accuracy (C)	"*It accounts for the correctness of the terms used in the ontology*" [3]
Lawfulness (M)	"*Correctness of syntax -the degree to which ontology language rules have complied*" [10]
Richness (M)	"*Refers to the proportion of features in the ontology language that have been used in an ontology*" [10]
Interpretability (M)	"*The knowledge provided by the ontology should map into meaningful concepts in the real world*" [33]
Consistency (C)	"*Whether terms have a consistent meaning in the ontology. E.g. if an ontology claims that X is a subclass_of Y, and that Y is a property of X, then X and Y have inconsistent meanings and are of no semantic value*" [10]
Clarity (M)	"*Whether the context of terms is clear. E.g.: in the context of academics, chair is a person, not furniture)*" [10]
Comprehensiveness (C)	"*Number of classes and properties - a measure of the size of the ontology*" [10]

(*continued*)

Table 3. (*continued*)

Criteria (C)/Metrics (M)	Definition
Accuracy (C)	*"Whether the claims an ontology makes are true"* [10]
Relevance (C)	*"Whether the ontology satisfies the agents specific requirements"* [10]
Authority (C)	*"Extent to which other ontologies rely on it (define terms using other ontology definitions)"* [10]
History (M)	*"Number of times ontology has been used"* [10]
Adaptability (C)	*"Whether the ontology provides a secure foundation that is easily extended and flexible enough to react predictably to small internal changes"* [34]
Ease of use (M)	*"Assess the level of documentation in the form of annotations included"* [34]
Recognition (M)	"*Assess the level of use it has received within its community*" [34]

5   Results and Discussion

When mapping the criteria with respect to the defined dimensions in our study, few deviations have been observed in the literature. For instance, in the article [34], the scholars have defined *structural complexity*: *the number of subclasses in the ontology* as a syntactic characteristic [34]. Although it describes the static property of the ontology structure, it is not a property that reflects the syntactic feature as defined in [34]. Importantly, the authors in [32] show that the structural metric: *cohesion* can be adopted to measure the semantics instead of the structural features of an ontology. This implies that the metrics do not strictly attach to one dimension and they can be measured in different ways to achieve the desired quality objectives. Thus, the measures would influence many dimensions through several criteria, which could be mapped with several ontology levels as shown in Table 2.

Moreover, in Table 4, we represented the metrics with levels and evaluation approaches that give an overview for researchers to identify possible quality metrics to be considered in each level with a suitable approach. Moreover, it has been mentioned that whether those metrics can be assessed manually or semi-/automatically. Based on our analysis, the metrics related to the structural, and syntactic levels can be automated as they are domain-independent. The metrics that come under the semantic and lexicon level also are automatable, however, need extra effort as those are domain depended. To assess the context level criteria, manual methods are mostly required as those are relative to the users and may not have specific quantitative metrics.

The metric: *formalism* describes the capabilities of the language that the ontology is written such as *machine understanding, reasoning*, and *defining required features* (i.e., entities, properties, relations). This has not been included under the proposed dimensions as it is considered before the ontology is modeled. However, the metric *formalism* is

Table 4. Matrix of ontology quality metrics, levels, and approaches.

Approaches	Layers				
	Context	Semantic	Lexicon	Structural	Syntactic
Application/Task/Users-based	Completeness: CQs (M) Accuracy (M) Ease of use (M) Relevancy (M) Efficiency (M) Applicability (M) Adaptability (M) Accessibility (M) History (M) Recognition (A)		*Clarity*		
Human-based (i.e., domain experts, ontology engineer)-	Accuracy (SA) Authority (M)	Conciseness: Precision (M) Comprehensibility (M)	*Cohesion (A)*	Size (A), Depth (A), Breadth (A), Fan-out (A), Tangledness (A), Circularity (A), Coupling (A), Cohesion (A), Partition (A), *Disjointsness*	Lawfulness (A) Richness (A) Formalization (M) *Flexibility*
Data-driven		Clarity (A) Interpretability (A)	Clarity (A)		
Golden-standard based	Accuracy (A)	Inferred Vocabulary coverage: (recall) (A), Clarity (A) Interpretability (A)	Vocabulary coverage(recall) (A)	Cohesion (A)	

*automate-A, semi-automate-SA, manual-M

noteworthy to consider when selecting a suitable ontology language for modeling. Thus, we included it in the matrix that can be assessed manually by ontology engineers.

To measure each metric, at least one approach is available. If many approaches are applicable, a suitable approach could be selected based on the purpose and availability of resources (i.e., time, experts, type of users, standards). For example, accuracy can be assessed through expert interventions, application-based or golden standard-based approaches [10]. However, in most cases, the golden standard (i.e., standard ontology, vocabularies, rules) is not available for comparison, and definitely, it is necessary to go with one of the other two approaches. If it is hard to evaluate an ontology in a real environment (i.e., application-based) with naïve users then the experts-based methods are acceptable.

When identifying the possible metrics with respect to the levels, we ignored the metrics defined in [3] since they have been defined by assuming ontology is a software artifact. As a result, the provided metrics are more subjective, and it is hard to match them with the ontology levels except the metrics defined under the structural dimension. Moreover, the authors in [22], have provided a criteria selection framework, without differentiating the metrics and the criteria. However, we adopted few metrics from it and have been included them in the matrix according to their classification (i.e., which are in the italic format in Table 4).

6   Conclusion and Future Work

A comprehensive theoretical analysis was performed on the ontology quality assessment criteria with the aim of providing an overview of criteria and possible metrics. The outcome of this analysis can be used to assess each ontology level with possible approaches as this domain has not been covered in the previous studies. To this end, we analyzed the definitions provided in the research works and clarified the vaguely defined definitions by studying theories in [36–38, 41–44]. Consequently, we were able to identify fourteen ontology quality criteria namely *syntactic correctness, cognitive complexity, conciseness, modularity, consistency, accuracy, completeness, adaptability, applicability, efficiency, understandability, relevance, usability*, and *accessibility*. These criteria have been classified under five dimensions namely: *syntactic, structural, semantic, pragmatic,* and *social*. Finally, a matrix was constructed that presents the association among the ontology levels, approaches, and criteria/metrics (see Table 4). This would become useful to gain an insight for researchers when dealing with the ontology quality assessment. Moreover, the absence of empirical evidence on the ontology quality assessment has limited the use of criteria in practice, and finding a methodological approach to derive ontology quality criteria with respect to the users' requirements (i.e., fit for the intended purpose) remains an open research problem.

References

1. Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: principles and methods. Data Knowl. Eng. **25**, 161–197 (1998)
2. Natalya, F.N., Deborah, L.M.: What is an ontology and why we need it. https://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html. Accessed 09 May 2021
3. Duque-Ramos, A., Fernández-Breis, J.T., Stevens, R., Aussenac-Gilles, N.: OQuaRE: a SQuaRE-based approach for evaluating the quality of ontologies. J. Res. Pract. Inf. Technol. **43**, 159–176 (2011)
4. Staab, S., Gomez-Perez, A., Daelemana, W., Reinberger, M.-L., Noy, N.F.: Why evaluate ontology technologies? Because it works! IEEE Intell. Syst. **19**, 74–81 (2004)
5. Duque-Ramos, A., et al.: Evaluation of the OQuaRE framework for ontology quality. Expert Syst. Appl. **40**, 2696–2703 (2013)
6. Gómez-Pérez, A.: Towards a framework to verify knowledge sharing technology. Expert Syst. Appl. **11**, 519–529 (1996)
7. Brank, J., Grobelnik, M., Mladenić, D.: A survey of ontology evaluation techniques. In: Proceedings of the Conference on Data Mining and Data Warehouses, pp. 166–170 (2005)
8. Ontology Summit 2013. http://ontolog.cim3.net/OntologySummit/2013/index.html. Accessed 09 May 2021
9. McDaniel, M., Storey, V.C.: Evaluating domain ontologies: clarification, classification, and challenges. ACM Comput. Surv. **52**, 1–44 (2019)
10. Burton-Jones, A., Storey, V.C., Sugumaran, V., Ahluwalia, P.: A semiotic metrics suite for assessing the quality of ontologies. Data Knowl. Eng. **55**, 84–102 (2005)
11. Gangemi, A., Catenacci, C., Ciaramita, M., Lehmann, J.: Modelling ontology evaluation and validation. In: Sure, Y., Domingue, J. (eds.) ESWC 2006. LNCS, vol. 4011, pp. 140–154. Springer, Heidelberg (2006). https://doi.org/10.1007/11762256_13
12. Gangemi, A., et al.: Ontology evaluation and validation: an integrated formal model for the quality diagnostic task, **53** (2005)
13. Lantow, B.: OntoMetrics: putting metrics into use for ontology evaluation. In: KEOD, pp. 186–191 (2016)
14. Tartir, S., Arpinar, I.B., Moore, M., Sheth, A.P., Aleman-Meza, B.: OntoQA: metric-based ontology quality analysis (2005)
15. ISO/IEC 25000:2005. https://www.iso.org/standard/35683.html. Accessed 09 May 2021
16. Degbelo, A.: A snapshot of ontology evaluation criteria and strategies. In: Proceedings of the 13th International Conference on Semantic Systems, pp. 1–8. Association for Computing Machinery, New York (2017)
17. Gómez-Pérez, A.: Ontology Evaluation. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies. International Handbooks on Information Systems, pp. 251–273. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24750-0_13
18. Vrandečić, D.: Ontology evaluation. In: Staab, S., Studer, R. (eds.) Handbook on ontologies. IHIS, pp. 293–313. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-92673-3_13
19. Reynoso, L., Amaolo, M., Vaucheret, C., Álvarez, M.: Survey of measures for the structural dimension of ontologies. In: 2013 IEEE 12th International Conference on Cognitive Informatics and Cognitive Computing, pp. 83–92 (2013)
20. Mc Gurk, S., Abela, C., Debattista, J.: Towards ontology quality assessment. In: MEP-DaW/LDQ@ ESWC, pp. 94–106 (2017)
21. Pak, J., Zhou, L.: A framework for ontology evaluation. In: Sharman, R., Rao, H.R., Raghu, T. S. (eds.) WEB 2009. LNBIP, vol. 52, pp. 10–18. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-17449-0_2

22. Hooi, Y.K., Hassan, M.F., Shariff, A.M.: Ontology evaluation: a criteria selection framework. In: 2015 International Symposium on Mathematical Sciences and Computing Research (iSMSC), pp. 298–303 (2015)

23. Raad, J., Cruz, C.: A survey on ontology evaluation methods: In: Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, pp. 179–186. SCITEPRESS - Science and and Technology Publications, Lisbon (2015)

24. Verma, A.: An abstract framework for ontology evaluation. In: 2016 International Conference on Data Science and Engineering (ICDSE), pp. 1–6 (2016)

25. Reyes-Pena, C., Tovar-Vidal, M.: Ontology: components and evaluation, a review. Res. Comput. Sci. **148**, 257–265 (2019)

26. Kitchenham, B.: Procedures for undertaking systematic reviews. Joint Technical report, Computer Science Department, Keele University (TR/SE-0401) and National ICT Australia Ltd. (0400011T.1) (2004)

27. Gavrilova, T.A., Gorovoy, V.A., Bolotnikova, E.S.: Evaluation of the cognitive ergonomics of ontologies on the basis of graph analysis. Sci. Tech. Inf. Process. **37**, 398–406 (2010)

28. Demaidi, M.N., Gaber, M.M.: TONE: a method for terminological ontology evaluation. In: Proceedings of the ArabWIC 6th Annual International Conference Research Track, pp. 1–10. Association for Computing Machinery, New York (2019)

29. Cross, V., Pal, A.: Metrics for ontologies. In: NAFIPS 2005 - 2005 Annual Meeting of the North American Fuzzy Information Processing Society, pp. 448–453 (2005)

30. Zhu, H., Liu, D., Bayley, I., Aldea, A., Yang, Y., Chen, Y.: Quality model and metrics of ontology for semantic descriptions of web services. Tsinghua Sci. Technol. **22**, 254–272 (2017)

31. Evermann, J., Fang, J.: Evaluating ontologies: towards a cognitive measure of quality. Inf. Syst. **35**, 391–403 (2010)

32. Ma, Y., Jin, B., Feng, Y.: Semantic oriented ontology cohesion metrics for ontology-based systems. J. Syst. Softw. **83**, 143–152 (2010)

33. Rico, M., Caliusco, M.L., Chiotti, O., Galli, M.R.: OntoQualitas: a framework for ontology quality assessment in information interchanges between heterogeneous systems. Comput. Ind. **65**, 1291–1300 (2014)

34. McDaniel, M., Storey, V.C., Sugumaran, V.: Assessing the quality of domain ontologies: metrics and an automated ranking system. Data Knowl. Eng. **115**, 32–47 (2018)

35. Oh, S., Yeom, H.Y., Ahn, J.: Cohesion and coupling metrics for ontology modules. Inf. Technol. Manag. **12**, 81 (2011)

36. IEEE Standard for a Software Quality Metrics Methodology: In IEEE Std 1061-1992, pp. 1–96 (1993)

37. Kitchenham, B., Pfleeger, S.L., Fenton, N.: Towards a framework for software measurement validation. IEEE Trans. Softw. Eng. **21**, 929–944 (1995)

38. Basili, V.R., Caldiera, G., Rombach, H.D.: The goal question metric approach. Encyclopedia Softw. Eng. **2**, 528–532 (1994)

39. Porzel, R., Malaka, R.: A task-based approach for ontology evaluation. In: Proceedings of the Workshop on Ontology Learning and Population at the 16th European Conference on Artificial Intelligence (ECAI), (2004)

40. OQuaRE: A SQUaRE based quality evaluation framework for Ontologies. http://miuras.inf.um.es/evaluation/oquare/. Accessed 09 May 2021

41. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. J. Manag. Inf. Syst. **12**, 5–33 (1996)

42. Mouzhi Ge., Markus, H.: A review of information quality research. In. Proceedings of the 2007 International Conference on Information Quality, pp. 76–91 (2007)

43. Uschold, M., Gruninger, M.: Ontologies: principles, methods and applications. Knowl. Eng. Rev. **11**, 93–136 (1996)
44. Nigel, B.: Human-computer interaction standards. In: Anzai, Y., Ogawa, K.(eds.) Proceedings of the 6th International Conference on Human Computer Interaction, vol. 20, pp. 885–890. Elsevier, Amsterdam (1995)