Ontology Quality Evaluation Methodology

R. Shyama I. Wilson[1,2(✉)] , Jeevani S. Goonetillake[2] , Athula Ginige[3] ,
and Walisadeera Anusha Indika[4]

[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] Western Sydney University, Penrith, NSW, Australia
a.Ginige@westernsydney.edu.au
[4] University of Ruhuna, Matara, Sri Lanka
waindika@dcs.ruh.ac.lk

Abstract. Lack of methodologies for ontology quality evaluation causes a challenge in producing good quality ontologies. Thus, we developed an iterative quality methodology to address this gap by analyzing the existing quality theories defined in ontology engineering, as well as, the theories in software engineering. Accordingly, this paper presents the developed methodology including how the other ontology quality theories get associated with it. Moreover, a use case in the agriculture domain has been demonstrated in order to provide an understanding of how the methodology can be applied in a real context. In the future, many experiments are expected to be carried out to fine-tune the methodology and to illustrate its usefulness.

Keywords: Ontology quality · Quality evaluation · Ontology · Agriculture

1 Introduction

Ontology engineering is a field that concerns the ontology development process (i.e., planning, designing, building, and maintaining) and studies methodologies, methods, tools and languages to produce a well-engineered ontology [1, 2]. However, it is debatable whether a well-engineered ontology is fit for the intended purpose, in other terms, it can be questioned whether a well-engineered ontology meets the requirements of the considered use case. This can be ensured only through a rigorous quality evaluation across the ontology development. However, the existing well-known ontology methodologies such as Uschold and King's method [3] TOVE [4], METHONTLOGY and On-to-Knowledge, have identified ontology evaluation as a phase of a methodology rather than considering it as an ongoing task which begins in the early stages of the developmental process. Moreover, these methodologies have not provided a detailed description of the quality evaluation. As a result, the attention to be provided for quality evaluation of ontologies got limited. Consequently, quality problems could not be detected until the ontology is used in an operational environment. Thus, in turn, it can adversely affect the overall quality of

a system in which the ontology is integrated. Moreover, solving those quality problems requires an extra effort and not only that, it can be expensive and time-consuming. Thus, quality evaluation of ontologies cannot be underestimated.

As explained in [5, 6], quality evaluation requires a proper understanding of the intended purpose of the ontology and the domain to be modeled. Moreover, it is an iterative process that should be initiated at the requirements analysis phase and should be carried out in parallel with the development of the ontology. In analyzing the existing works, it has been identified that these notions have not been considered. Instead, the studies have discussed several other aspects such as quality scopes, evaluation levels, quality characteristics, tools, stages, approaches and techniques (see Fig. 1). To this end, we proposed an iterative methodology for ontology quality evaluation by analyzing quality theories in software engineering and also based upon the experience gained through ontology development and evaluation. Moreover, we made an effort to map the existing quality theories with the relevant steps of the proposed methodology (see Sect. 3). This, in turn, provides a better understanding of how the quality theories are getting associated with the quality evaluation process. Accordingly, the paper has been structured as follows.

Section 2 discusses the quality aspects: *levels, characteristics, scopes, techniques, stages, tools* and *approaches* considered in the previous works. Section 3 presents the proposed methodology for ontology quality evaluation. This has been further exemplified in Sect. 4. Finally, Sect. 5 concludes the discussed methodology by highlighting the future work.

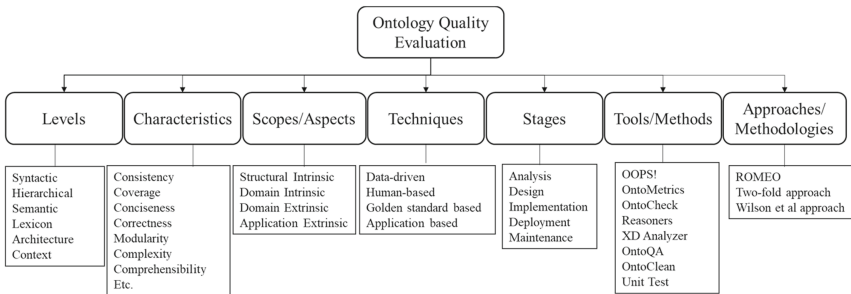**Fig. 1.** Overview of ontology quality evaluation.

2 Overview of the Ontology Quality

Ontology quality has been discussed under several aspects such as (see Fig. 1):

- Levels of ontologies: syntactic, semantic, hierarchy/taxonomy, architecture/structure, lexicon and context [7, 8]
- Characteristics (i.e., criteria, principles): consistency, conciseness, completeness, accuracy, adaptability, clarity, etc. [8–10]

- Scopes: conceptualization/structure, domain and technical/application [11, 12]
- Evaluation aspects: structural intrinsic, domain intrinsic, domain extrinsic and application extrinsic [5, 6]
- Techniques: data-driven, golden standard-based, human-based and application-based [7, 8, 11]
- Stages in the ontology life cycle: analysis, design (i.e., conceptualization, formalization), implementation, deployment and maintenance [5]
- Tools/methods [6, 13, 14]
- Approaches/methodologies [10, 15, 16]

The following sub-sections discuss the aforementioned aspects in detail.

2.1 Ontology Quality Evaluation Levels

In evaluating an ontology, different levels namely syntactic, hierarchical, semantic, lexicon, architecture and context have been taken into account rather than evaluating ontology as a whole component [7]. For instance, under the syntactic level, ontology is evaluated to confirm whether it complies with the specification and syntax defined in the representation language. Under the hierarchy level, the properties related to the taxonomic structure (i.e., subsumption relationship) of an ontology are evaluated. Thus, it is also known as the taxonomy level evaluation. OntoClean [17] is one of the well-known methodologies that supports assessing the taxonomic properties such as *essence*, *identity*, and *unity* of an ontology. Under the semantic level, it is checked whether the ontology content is coherent with the domain knowledge that has been used to model the ontology. Typically, precision and recall measures have been used to assess the characteristics (i.e., semantic consistency, domain coverage and conciseness) related to the domain coherency [10, 15, 18]. In addition to that, the ontology taxonomic measures such as maximum depth, maximum breadth and structural variance have been adopted to assess the semantic level of an ontology [19, 20]. Under the lexicon level of an ontology, the vocabulary used to identify the ontology components (i.e., concepts, relationships, attributes, and individuals) is evaluated. The pre-defined representation principles are evaluated under the architecture level and the practical usefulness in the operational environment is evaluated under the context level [7, 8].

2.2 Ontology Dimensions, Characteristics and Measures

Various characteristics/attributes and measures have been proposed for ontology evaluation. Initially, the author in [3, 9] proposed five characteristics such as clarity, coherence, expandability, minimum encoding bias and minimal ontological commitment for ontology evaluation. Thereafter, the author in [8] has proposed another set of characteristics by associating them with the ontology evaluation levels (see Table 1). For instance, the characteristic: correctness has been defined to evaluate the lexicon and syntactic levels of ontologies. Later, the researchers in [21] have proposed a set of quantitative attributes to assess the quality of an ontology. This set of attributes can also be utilized to select a suitable ontology from the ontology repositories for the intended purpose. Based on that, the

tools: OntoQA [21] and OntoMetrics [22] have been developed which support ontology developers to easily assess ontologies by avoiding the cost of manual evaluation.

Meanwhile, the semiotic metric suit has been proposed that consists of ten attributes grouped into four dimensions, namely: syntactic, semantic, pragmatic and social. Moreover, a set of measures to evaluate each attribute has been provided through that metric suit. Another comprehensive set of measures has also been introduced for ontology quality assessment in [23]. To this end, the attributes have been classified into three dimensions, namely: *structural*, *functional* and *usability-related*. Similarly, many researchers have proposed different characteristics, attributes and measures by grouping them into different dimensions (see Table 1). Therefore, we had to perform a comprehensive literature analysis to identify a set of characteristics that can be used for ontology quality evaluation [15, 24, 25]. Consequently, we proposed a quality model which consists of nineteen characteristics associated with different evaluation scopes [25]. Moreover, when analyzing the existing works, several discrepancies associated with the definitions provided for the characteristics, attributes and measures were observed. To this end, we made an effort to provide definitions for each characteristic identified through the literature that has been further discussed in [24, 25].

Table 1. Ontology quality characteristics.

Citation	Dimension	Characteristics/Attributes
[9]		Clarity, coherence, expandability, minimum encoding bias and minimal ontological commitment
[8]	Architecture	Soundness
	Syntax and lexicon	Correctness
	Content	Consistency, completeness, conciseness, expandability and sensitiveness
[21]	Schema	Relationship richness, attribute richness, inheritance richness
	Data	Class richness, average population, cohesion, importance, fulness, relationship richness, inheritance richness, connectivity and readability
[26]	Syntactic	Lawfulness, richness
	Semantic	Interpretability, consistency, clarity
	Pragmatic	Comprehensiveness, accuracy, relevance
	Social	Authority, history
[23]	Structural	Topological, logical, meta-logical
	Functional	Accuracy, precision, recall
	Usability-related	Recognition annotation, efficiency annotation, and interfacing annotation

(*continued*)

Table 1. (*continued*)

Citation	Dimension	Characteristics/Attributes
[27]	–	Structural, functional adequacy, reliability, operability, maintainability, compatibility, transferability, etc.[a]
[18]	–	Language conformance Completeness Conciseness Correctness (syntactic correctness, semantic correctness, representation) correctness, usefulness
[28]	Content	Correctness (internal consistency, external consistency, compatibility), completeness (syntactic completeness, semantic completeness)
	Presentation	Well-formedness, conciseness (non-redundancy), structural complexity (size, relation), modularity (cohesion, coupling)
	Usage	Applicability (definability, description complexity), adaptability (tailorability, composability, extendibility, transformability), efficiency (search efficiency, composition efficiency, invocation efficiency), comprehensibility

[a] http://miuras.inf.um.es/evaluation/oquare/Contenido.html.

2.3 Ontology Evaluation Scopes and Aspects

The researchers have discussed different scopes/aspects to be considered in quality evaluation. The authors in [11, 12] have presented the three main scopes, namely: conceptual scope, domain scope and technical scope. Conceptual scope evaluates whether the concepts associated with the taxonomy of ontologies are well represented. When comparing with the ontology levels, it can be understood that the quality of the taxonomic level is considered in this scope. The domain scope evaluates whether the ontology represents the domain knowledge required to accomplish the specified tasks in the considered use case. To this end, the authors in [12] have shown that the domain scope considers the evaluation of the lexicon and architectural levels. Moreover, the technical scope evaluates whether the ontology meets the specified application requirements which are required for ontology integration and application in practice [5, 7].

Similar to the evaluation scopes described just above, the ontology summit has also proposed a set of scopes (i.e., evaluation aspects) to be considered in evaluating the quality of ontologies [6]. However, they have defined four main scopes, namely: *structural intrinsic*, *domain intrinsic, domain extrinsic and application extrinsic*.

The domain extrinsic and application extrinsic consider the evaluation of domain requirements and application requirements of an ontology which are specifically needed for a particular application respectively. These evaluations are very much similar to the "black-box" or "task-based" testing in software engineering [29]. Thus, the quality assessment is performed without peering at the ontology structure and the content. This has also been defined as ontology validation in [5, 8]. In general, the quality under the

extrinsic aspect is determined by analyzing the correctness of the answers provided to competency questions with the involvement of users and domain experts [4].

The structural intrinsic and domain intrinsic scopes focus on the ontology structural and content quality evaluation respectively. In the structural intrinsic scope, the focus is given to the syntactic and structural requirements (i.e., syntactic, structural, architectural levels) which involve the specified conceptualization such as language compliance, conceptual complexity, and logical consistency. In the domain intrinsic aspect, quality is evaluated with reference to the domain knowledge which is used for the knowledge representation. At this stage (i.e., in the intrinsic aspect), mainly, the evaluation is performed by ontology engineers by considering the ontology as an isolated component separated from the system [5]. Somehow, ontology engineers need to obtain the assistance of domain experts when evaluating the domain intrinsic aspect as it requires some domain understanding. As we understood through the literature review, the semantic, vocabulary and architectural levels of ontologies are evaluated under the domain intrinsic aspect. Furthermore, the quality evaluation performed under the structural and domain intrinsic aspects are similar to the paradigm of "white-box" testing in software engineering [29]. Thus, the verification is being done under the intrinsic scope to ensure whether the ontology is built in the right way [8, 29].

2.4 Ontology Quality Evaluation Techniques

There are various techniques for carrying out a quality evaluation. The most commonly discussed techniques are human-based evaluation, data-driven evaluation, golden standard-based evaluation and application-based evaluation [7]. These techniques can be selected based upon the purpose, characteristics, scope, and/or levels to be evaluated of an ontology [24, 30]. For instance, the human-based technique is performing the quality assessment with the intervention of domain experts and/or users, and all the ontology levels can be assessed using this technique. In contrast to that, the data-driven evaluation assesses the ontology using a valid corpus and it is more suitable for ontology evaluation when it is difficult to acquire domain experts. However, this technique typically is used to evaluate the vocabulary and semantic levels of an ontology. When considering the golden-standard-based technique, it uses a standard ontology for the quality assessment and also it can be used in evaluating the levels: vocabulary, structure and semantics. However, the main difficulty of this technique is to find a standard ontology that has the quality at an acceptable level for the specified use case. The application-based technique is used to evaluate an ontology in an operational environment after it is integrated into the application. Thus, this technique can be adopted to evaluate the practical usefulness of an ontology and therefore, is suitable for assessing the context level of an ontology.

2.5 Ontology Evaluation Across the Ontology Life Cycle

Ontology quality could also depend on the success of activities that are followed to develop an ontology. For instance, if the requirement specification of an ontology is poorly performed at the requirement development phase, then, the resulting ontology will not succeed in providing knowledge to the specified tasks. To this end, the authors in

[5] have introduced a comprehensive set of criteria that can be used to evaluate the activities performed under each stage (i.e., *requirement development, ontological analysis, ontology design, system design, ontology development & reuse, system development & integration, deployment* and *maintenance*) of the ontology development.

In addition to that, the author in [31] has performed a comprehensive survey and analyzed quality criteria that can be used to evaluate ontologies at the design and implementation stages of the development. Consequently, it has been identified that a set of possible criteria such as *accuracy, adaptability, cognitive adequacy, completeness, conciseness, consistency, expressiveness* and *grounding* can be used for evaluating the design of an ontology. Moreover, the criteria: *computational efficiency, congruency, practical usefulness, precision* and *recall*, can be used to evaluate ontologies at the implementation stage. Nonetheless, as shown by the author in [31], the previous works have used only a few characteristics from these to evaluate the ontologies at each stage. Mainly, there is a piece of evidence in using the characteristic: *functional completeness (i.e., expressiveness)* and *practical usefulness* frequently in the previous works. To this end, the authors in [15, 31] have pointed out the necessity of introducing a complete methodology or approach for ontology evaluation in order to avoid the quality evaluation getting limited to a certain set of characteristics.

2.6 Ontology Quality Evaluation Tools

Appropriate tools can be utilized to make the ontology quality evaluation process easy, efficient and cost-effective. In the previous studies, many tools have been introduced and a few of them are summarized in Table 2. In analyzing the tools, it can be realized that a number of tools are available for evaluating the syntactic and structural properties. The tools: S-OntoEval, RepOSE, DoORS and OntoKeeper support to assess the characteristics in structural and domain intrinsic aspects. Nevertheless, it can be observed that only a few tools are available to use online such as RDF validator, OWL validator, OOPS!, OntoMetrincs, DoORS and RepOSE.

Table 2. Ontology tools and methods.

Tool/Method	Characteristics/Attributes	Aspects
RDF Validation Service [32]	Language compliance	Structural intrinsic
OWL Validator [33]	Language compliance	Structural intrinsic
OntoAnalyser [34]	Language compliance and internal consistency	Structural intrinsic
OntoKick [34]	Accuracy and functional completeness	Domain extrinsic

<div align="right">(continued)</div>

Table 2. (*continued*)

Tool/Method	Characteristics/Attributes	Aspects
OntoQA [21]	Complexity: relationship richness, attribute richness, inheritance richness, class richness, Average population, fullness Modularity: cohesion, importance, connectivity, instance relationship richness, etc	Structural intrinsic
S-OntoEval tool [35]	Complexity, modularity, internal consistency, external consistency, comprehensibility	Structural intrinsic and domain intrinsic
OntologyTest [36]	Accuracy and functional completeness	Domain extrinsic
OntoCheck [37]	Comprehensibility	Structural intrinsic attributes that are complementary to the domain intrinsic aspect have been automated
XD analyzer [38]	Coverage: isolated entities, missing types, missing domain or range in properties, missing inverse, Comprehensibility: instance missing labels and comments, unused imported ontologies	Structural intrinsic attributes that are complementary to the domain intrinsic aspect have been automated
RepOSE[a] [39]	Compliance and external consistency	Structural intrinsic and domain intrinsic
OOPS![b] [40]	Detect common pitfalls Compliance: e.g., P34, P35 and P38 Consistency: e.g., P05, P06, P07, P19 and P24 Coverage: e.g., P04, P10, P11, P12 and P13 Conciseness: e.g., P02, P03, P21 and P32 Comprehensibility: e.g., P08, P20 and P22 Availability: e.g., P36 and P37	Structural intrinsic, domain intrinsic/extrinsic and application extrinsic (i.e., only automated structural intrinsic metrics)

Table 2. (*continued*)

Tool/Method	Characteristics/Attributes	Aspects
OntoMetric[c] [41]	Complexity: basic metric, knowledge base metric Modularity: graph metrics, class metrics Comprehensibility: annotation metrics	Structural intrinsic and domain intrinsic (i.e., only automated structural intrinsic metrics)
OntoDebug [42]	Internal Consistency	Structural intrinsic
DoORS[d] [43]	Compliance, external consistency, conciseness, comprehensibility, accuracy, relevancy, and credibility	Structural intrinsic, domain intrinsic and domain extrinsic
OntoKeeper [44]	Compliance: lawfulness, richness, conciseness, comprehensibility, accuracy, relevancy, and credibility	Structural intrinsic, domain intrinsic, and domain extrinsic
Delta [45]	Complexity, modularity	Structural intrinsic

[a] RepOSE: https://www.ida.liu.se/~patla00/research/RepOSE/
[b] OOPS!: http://oops.linkeddata.es/
[c] OntoMetric: https://ontometrics.informatik.uni-rostock.de/ontologymetrics/
[d] DoORS: https://owlparser.herokuapp.com/

2.7 Ontology Quality Approaches and Methodologies

It is vital to have approaches and methodologies that systematically describe theories to be used, processes to be followed and the steps to be carried out in doing some work. With respect to the ontology quality evaluation, a few contributions can be found, such as the ROMEO methodology [10], the Two-Fold quality assurance Approach [16] and the Wilson et al. approach [15]. ROMEO methodology provides a set of guidelines to identify the intrinsic ontology quality characteristics from the defined ontology quality requirements of a system. Mainly, it has been constructed by employing the Goal-Question-Metrics (GQM) paradigm in information systems [46]. The Two-Fold quality approach has been introduced to monitor and assess the quality of an evolving knowledge base. It consists of two main phases, namely: coarse grain analysis and fine-grain analysis. During the coarse grain analysis, a quantitative analysis is performed to detect the high-level changes and common quality issues in a knowledge base. On the other hand, fine-grain analysis is performed to detect the detailed changes in a knowledge base and to detect all possible quality issues. The Wilson et al. approach mainly presents the steps to be carried out in determining quality characteristics from the intended needs of an ontology and presents how the identified characteristics are evaluated. Similar to the ROMEO methodology, this approach is also based on the GQM paradigm. However, the Wilson et al. approach supports to separately identify both intrinsic and extrinsic quality characteristics considering the ontology evaluation aspects (i.e., scopes).

3   Ontology Quality Evaluation Methodology

Although quality evaluation has been considered as a phase in most of the ontology methodologies, it is an iterative process that consists of several activities such as; *identification of intended needs*, *elicitation of quality requirements from the identified needs, prioritizing the quality requirements*, *specifying quality characteristics* (i.e., intrinsic and extrinsic aspect) and *performing quality assessment* across the ontology development [5, 47]. Moreover, these activities should be carried out in parallel with the development of the ontology starting from the stage of requirements analysis. However, there is no proper methodology that provides a set of steps that need to be followed under the quality evaluation of ontologies. To this end, we propose a methodology for the ontology quality evaluation by examining software quality theories described in SQuaRE (Systems and software Quality Requirements and Evaluation) [48] and also based on the experience gained through our previous studies [15, 24]. Accordingly, the proposed methodology consists of four main steps, namely: *quality requirement specification, plan & design, execution* and *user acceptance test* (see Fig. 2). These steps should be followed in parallel with the steps in the ontology development as illustrated in Fig. 3. The outer circle of Fig. 3 presents the steps in the ontology development life cycle as defined in [5]. The inner-circle presents the steps to be followed through our methodology which are further described in the below subsections.
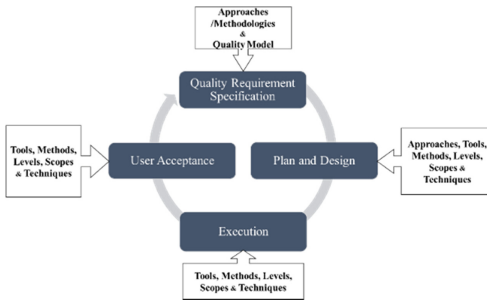

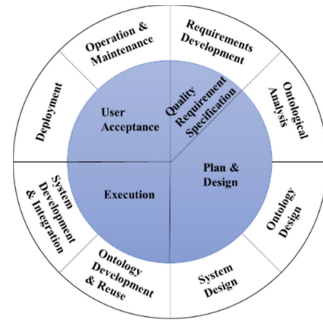**Fig. 2.** Ontology quality evaluation methodology.

Fig. 3. The quality evaluation stages associated with the ontology life cycle.

3.1   Quality Requirement Specification

Under this stage, mainly, the quality requirements to be achieved through an ontology should be identified. It is noteworthy that quality requirements have been also defined as *quality characteristics* [49]. Thus, hereinafter, we use the term *quality characteristics* to refer to the term *quality requirements.* To identify the quality characteristics, initially, it is required to recognize,

- Who are the intended users?
- What is the intended purpose of using an ontology/ontology-driven system?

- What is the context of use?
- What are the intended needs[1] in the considered context (i.e., use case)? What are the competency questions?
- What is the scope of the ontology? What set of Competency Questions (CQs)/user needs will be covered through the ontology?
- What are the resources available for the ontology quality evaluation (i.e., text corpora, documents, domain experts, other related ontologies, users, tools, budget, time)?

Generally, ontology engineers identify the aforementioned factors at the requirement analysis phase in ontology development [5]. Thus, they can be considered when specifying the quality characteristics during this step. To this end, an appropriate approach/methodology can be followed. For instance, the Wilson et al. approach [15] can be used which explains how the characteristics for evaluating each aspect (i.e., intrinsic and extrinsic) are derived from the intended needs. In addition to this approach [15], the ROMEO methodology can also be used. However, it only supports deriving the characteristics associated with the intrinsic aspect of an ontology [10]. Moreover, quality models are required for the mentioned approaches that support determining the characteristics. This is due to the fact that quality models[2] present a possible set of characteristics that are applicable for an artifact evaluation. In software engineering, ISO/IEC 25010 defines a quality model that supports specifying the characteristics required for the considered software product [48]. Moreover, it describes a set of measures that can be used to assess quality characteristics. However, for ontology engineering, there are no such agreed quality models that can be used for the quality requirement specification. Instead, a number of characteristics/metrics proposed in the previous works can be seen such as characteristics/metrics proposed in [3, 8, 23, 26, 43]. Nevertheless, it is difficult for researchers and developers to analyze all these existing works and to identify a proper set of quality characteristics. To this end, we constructed an ontology quality model after performing a comprehensive literature analysis that presents nineteen main characteristics [15, 25]. Meantime, it describes a set of measures that can be used to assess characteristics. Therefore, this model can be utilized when following any of the aforementioned approaches to specify quality characteristics.

The followings should be the outputs of this stage.

- Specification of ontology quality requirements (i.e., quality characteristics)
- Specification of quality measures

3.2 Plan and Design

After specifying the quality characteristics, evaluators (i.e., ontology engineers, and curators) can identify the related ontology levels or scopes (see Fig. 1) in which the characteristics are associated. Moreover, they can further identify the measures to be

---

[1] Intended needs: business requirements that the intended users (i.e., persons, parties or organizations) expect from the ontology/ ontology driven information system.

[2] Quality model consists of a set of characteristics and the relationships between them that provide the basis for specifying quality requirements and evaluating quality [49].

used to assess each characteristic under each level with the support of a quality model. Also, they can define the decision criteria[3] for the selected measures. Furthermore, the evaluators can select the tools to be utilized, methods and techniques to be applied in order to assess the measures derived for the characteristics. For instance, assume that *functional accuracy* has been identified as a quality characteristic for an ontology in a particular context. Then, it can be derived through the approaches [10, 15, 18] and the quality model [15, 25], that *external consistency, internal consistency* and *syntactic accuracy* are a set of quality characteristics that are to be achieved through the ontology for *functional accuracy*. Meantime, the levels and scopes in which the characteristics are associated can be identified with the support of approaches. Table 3 represents the levels and scopes related to the characteristics that we have identified. Moreover, it includes the related measures, tools/methods, and techniques that can be used to assess the characteristics. When selecting techniques and tools/methods, it is required to analyze the resources that are available for the evaluation. For instance, *external consistency* is "the degree to which an ontology (i.e., ontology definitions) is coherent with the specified domain knowledge" [15, 25]. To this end, a frame of reference is required to check the domain coherency. A frame of reference could be a text corpus, domain experts and a standard ontology. If we have a valid text corpus to check the domain coherency, then the data-driven evaluation techniques can be utilized to assess the *external consistency*. On the other hand, if the domain expert intervention is readily available to the evaluators, then the human-based techniques can be easily carried out to evaluate the characteristic. Similarly, for all the relevant quality characteristics, the factors: *scope, tool, method* and *techniques* should be identified and shall be documented under this stage.

Table 3. The quality aspects related to the selected user requirement: functional accuracy.

Characteristic	Scope	Level	Measures	Tool/Methods	Techniques
External consistency	Domain intrinsic	Semantic	Precision	Manually	Data driven, Human based
Internal consistency	Structural intrinsic	Semantic and structural	Logical contradictions	Reasoners	An evaluator can use the tools to observe the logical contradictions
Syntactic accuracy	Structural intrinsic	Syntactic	Lawfulness	OOPS!, protégé	An evaluator can use the tools to observe the syntactic violence

---

[3] Decision criteria may be "*numerical threshold that used to determine the level of confidence in a given results. These will often be set with respect to quality requirements and corresponding evaluation criteria*" [48].

The following should be the outputs of this stage.

- Specification of the selected quality measures, tools, methods, and techniques to be used in the quality evaluation.
- Specification of decision criteria for ontology quality measures.
- Specification of resources available for the evaluation.
- Specification of detailed quality evaluation plans including time and budget [48].

3.3   Execution

During the execution stage, the selected quality measures of the characteristics shall be applied to the developed ontology or ontology being developed to check the required quality characteristics are achieved. In other terms, both *white-box testing* (i.e., domain intrinsic and structural intrinsic) and *black-box testing* (i.e., domain extrinsic and application extrinsic) shall be performed under this stage. Then, the results of the measures shall be reported and reviewed using the decision criteria. Consequently, any specific deficiencies can be identified with regard to the quality requirements. Thereafter, those deficiencies can be informed to the ontology developers to take the required actions to address them. Moreover, any limitations, constraints and exclusions in an evaluation can be reported including their impact on the use [48].

The following should be the outputs of this stage.

- Results of the ontology quality measures associated with the characteristics.
- Report of the limitations, constraints and exclusions in an evaluation.

3.4   User Acceptance

Under this stage, an ontology shall be evaluated in order to ensure whether the ontology meets the intended needs. In the case of an ontology-driven system, the ontology shall be assessed with the intervention of end-users of the system. To this end, the criteria and techniques defined for the acceptance testing in system and software engineering [48] can also be adopted due to the ontology will not appear as an individual component to the end-users at this level. In the case where the ontology is deployed as a standalone reference ontology, the appropriate application-based techniques can be used to observe whether the intended needs in an operational environment are achieved through the ontology [5, 7, 18].

Through this stage, the intended needs that are not covered through the ontology can be detected. These needs may be (i.) a set of new requirements for the ontology or (ii.) the requirements that have been identified during the requirement analysis, but, have not been addressed through the development. In these cases, the missing requirements shall be reported to the development team for further action. Consequently, the quality requirement specification can be refined and the methodology can be repeated.

The following should be the outputs of this stage.

- Results of the user acceptance test, i.e., new user needs, user feedback, comments and suggestions.

4   Application

Quality Rquirement Specification: to exemplify the methodology that we have proposed, we consider a use case in the agriculture domain explained in [15, 50]. In Sri Lanka, agriculture is one of the main industries, of which, the farmer is the main stakeholder who struggles to access the right information at the right time to make the right decisions. To address this issue, the requirement of producing an agricultural ontology and the intended needs to be achieved through that ontology have been identified in [50]. With respect to our methodology, we initially identified the answers to the questions which are highlighted in Sect. 3.1 and have been summarized in Table 4. A few of the identified intended needs were selected for the explanation as given below.

a)   Users need the necessary and sufficient contextual information.
b)   Users need trustworthy information.
c)   Users need information in an understandable way.
d)   Users need up-to-date information.

These needs are a set of inputs for the quality requirement specification in the proposed quality methodology. We illustrate how the quality requirement (i.e., characteristics) can be specified considering only the first requirement (i.e., *a*) in order to maintain the simplicity of the explanation. However, Table 4 summarized the quality characteristics which are associated with the other needs specified from **b** to **d**.

To identify the quality characteristics which are expressed in the user need: *"a"*, we employed the Wilson et al. approach [15]. Accordingly, we formulated the questions by considering the need mentioned below.

Q1: Does the ontology provide contextual information in a specified context of use?

Q2: Does the ontology provide necessary and sufficient information in a specified context of use?

With the support of the ontology quality model in [15, 25], we identified that the mentioned questions describe the characteristics: *relevancy* and *completeness* of ontology information respectively. By further subdividing these questions, the ontology characteristics to be achieved at the domain intrinsic and structural intrinsic levels were identified. For example, the characteristics: *conciseness* and *compliance* have been identified as characteristics to be satisfied at the intrinsic level (i.e., domain, structural) to achieve *relevancy* (Q1). To achieve *completeness* (Q2), the required internal ontology characteristics have been identified as *coverage* and *compliance*. Similarly, the ontology characteristics associated with the other user needs can also be identified. Due to the page limitation, we have not described in this paper the steps in detail which are to be followed in deriving the quality characteristics and further detail is available in [15]. Accordingly, Table 4 presents all the related quality characteristics derived from the mentioned needs.

Table 4. Summarization of the quality requirements.

Main users	Farmers, Agriculture Instructors, Experts
Purpose	To support the decision-making process of farmers
Context of use	Pest and disease management in Brinjal
Resources	Agronomist, domain experts, terminology documents, list of packages of practices, tools: Protégé, OOPS! (online), OntoMetrics (online), etc
User needs	*Quality requirement (see Q1–Q6 in* [15]*)*
a	Q1. Relevancy: conciseness and compliance Q2. Completeness: coverage and compliance
b	Q3. Accuracy: external consistency and internal consistency Q4. Credibility
c	Q5. Understandability: comprehensibility and compliance
d	Q6. Timeliness
Scope and CQs	Scope and CQs are defined in [15, 50]

Plan and Design: During this stage, it is required to plan and design how the specified quality characteristics shall be evaluated. Accordingly, we identified the measures that can be used to assess the specified quality characteristics (i.e., relevancy, completeness, conciseness, coverage and compliance), the tools, methods, and techniques based upon the available resources (see Table 5). In selecting the measures for the characteristic evaluation, we utilized the quality model that we have constructed in [15, 25]. To evaluate the structural intrinsic characteristics, the available tools were explored and selected such as OOPS! [40], OntoMetrics [22], Protégé [51] (see Table 5). Moreover, we defined the data-driven techniques to measure the characteristics: *external consistency, conciseness, comprehensibility* and *coverage* as we have the documents provided by the domain experts. To assess the domain extrinsic scope, the human-based techniques were defined as the domain expert and user assistance can be obtained for the considered context. To this end, the unit test discussed in [52, 53] can also be performed by maintaining test cases. Thus, in that case, all the required test cases should be documented and then, they can be used during the test execution. Similarly, the appropriate evaluation measures, tools, methods, techniques and scope are required to decide during this stage in order to use in the quality execution phase.

Table 5. Quality characteristics and the related evaluation aspects. *Note: *E* denotes extrinsic characteristics, *I* denotes intrinsic characteristics, *O* denotes a modeled ontology, and *F* denotes a frame of reference.

Characteristics	Measures	Techniques	Methods/Tools
Relevancy (E)	The number of competency questions received relevant answers	Human-based	Protégé,

(*continued*)

Table 5. (*continued*)

Characteristics	Measures	Techniques	Methods/Tools
Completeness (E)	The number of competency questions received sufficient answers	Human-based	Protégé,
Accuracy (E)	Number of competency questions received correct answers	Human-based	Protégé,
Credibility (E)		Human-based	Manual
Understandability (E)		Human-based	Manual
Timeliness (E)	Average update rate		Manual
Conciseness (I)	Precision (O; F)	Data-driven	Manual
Coverage (I)	Recall (O; F)	Data-driven	Manual
External consistency (I)	Precision (O; F)	Data-driven Human-based	Manual
Internal consistency (I)	Number of logical contradictions	Human-based	Reasoners
Compliance (I)	Measures of lawfulness, richness, class richness, naming conventions, etc. [15]	Human-based	Protégé, OOPS! and OntoMetrics

Execution: As planned and designed in the previous stage, the quality evaluation can be performed in this stage. According to the example, during the intrinsic level, the structural intrinsic characteristics were evaluated using the selected tools such as OOPS! [40], OntoMetrics [22], Protégé [51] and reasoners [54]. The characteristics in the domain intrinsic scope were assessed manually using data-driven techniques. Finally, the characteristics in the domain extrinsic scope were evaluated using Protégé by running the SPARQL queries defined for the CQs. In this case, the answers produced to the CQs were validated using the documents provided by the domain experts and obtaining their support. After assessing the identified measures using the selected methods and techniques, the results of this evaluation were reported. The evaluation results related to this example, i.e., the use case in agriculture, can be found in [15].

User Acceptance Test: In the considered use case, an ontology is used as a component of a decision support system. Therefore, the user acceptance test can be performed by giving the ontology-driven system to the end-users, i.e., farmers and agriculture inspectors. To this end, the quality in use criteria defined in ISO/IEC 25010 can be used to assess the effectiveness, efficiency and user satisfaction of the ontology-driven system [49]. Due to the main system being under development, the acceptance test in a

real environment has not yet been performed for the use case. In the future, we expect to present a thorough result analysis including this step.

5   Conclusion

The usefulness of a methodology for ontology quality evaluation has been identified through theoretical and empirical reviews. As an initial step, we constructed a methodology by analyzing the theories in software engineering, i.e., SQuaRE [48, 49] and experience gained through the development and evaluation of ontologies. Consequently, the developed methodology consists of four main steps, namely: *quality requirement specification, plan & design, execution* and *user acceptance test* (see Fig. 2). These steps can be performed iteratively and parallelly with the ontology development life cycle. The applicability of this methodology for real applications has been exemplified by discussing a use case in agriculture. Moreover, the methodology has been introduced for undergraduate students who are currently doing research in ontology engineering. Most of the students provided positive feedback by highlighting that the methodology is useful for them to understand the proper set of characteristics to be assessed and how the quality concepts discussed in the literature are associated with it. To this end, we further expect to observe the results of the experiments using the proposed methodology in many use cases and to enhance it with a more comprehensive set of guidelines.

References

1. Gal, A.: Ontology engineering. In: Liu, L., Özsu, M.T. (eds.) Encyclopedia of Database Systems, pp. 1972–1973. Springer, Boston (2009). https://doi.org/10.1007/978-1-4614-8265-9

2. Gómez-Pérez, A., Fernández-López, M., Corcho, O.: Theoretical foundations of ontologies. In: Gómez-Pérez, A., Fernández-López, M., Corcho, O. (eds.) Ontological Engineering: With Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web, pp. 1–45. Springer, London (2004). https://doi.org/10.1007/1-85233-840-7_1

3. Uschold, M., Gruninger, M.: Ontologies: principles, methods and applications. Knowl. Eng. Rev. **11**, 93–136 (1996)

4. Gruninger, M., Fox, M.S.: Methodology for design and evaluation of ontologies. In: Workshop on Basic Ontological Issues in Knowlege Sharing (1995)

5. Neuhaus, F., et al.: Towards ontology evaluation across the life cycle: the ontology summit 2013. Appl. Ontol. **8**, 179–194 (2013)

6. Ontology Summit 2013. http://ontolog.cim3.net/OntologySummit/2013/. Accessed 26 June 2021

7. Brank, J., Grobelnik, M., Mladeni, D.: A survey of ontology evaluation techniques. In: Proceedings of the Conference on Data Mining and Data Warehouses, pp. 166–170 (2005)

8. Gómez-Pérez, A.: Towards a framework to verify knowledge sharing technology. Expert Syst. Appl. **11**, 519–529 (1996)

9. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. Int. J. Hum. Comput. Stud. **43**, 907–928 (1993)
10. Yu, J., Thom, J.A., Tam, A.: Requirements-oriented methodology for evaluating ontologies. Inf. Syst. **34**(8), 766–791 (2009). https://doi.org/10.1016/j.is.2009.04.002. ISSN 0306-4379
11. 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
12. Lantow, B.: OntoMetrics: putting metrics into use for ontology evaluation. In: KEOD, pp. 186–191 (2016)
13. Corcho, O., Fernández-López, M., Gómez-Pérez, A.: Methodologies, tools and languages for building ontologies. Where is their meeting point? Data Knowl. Eng. **46**, 41–64 (2003)
14. Aruna, T., Saranya, K., Bhandari, C.: A survey on ontology evaluation Tools. In: 2011 International Conference on Process Automation, Control and Computing, pp. 1–5 (2011)
15. Wilson, S.I., Goonetillake, J.S., Ginige, A., Walisadeera, A.I.: Towards a usable ontology: the identification of quality characteristics for an ontology-driven decision support system. IEEE Access **10**, 12889–12912 (2022)
16. Rashid, M., Torchiano, M., Rizzo, G., Mihindukulasooriya, N., Corcho, O.: A quality assessment approach for evolving knowledge bases. Semant. Web **10**, 349–383 (2019)
17. Guarino, N., Welty, C.A.: An overview of OntoClean. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies, pp. 151–171. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24750-0_8
18. 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)
19. Sánchez, D., Batet, M., Martínez, S., Domingo-Ferrer, J.: Semantic variance: an intuitive measure for ontology accuracy evaluation. Eng. Appl. Artif. Intell. **39**, 89–99 (2015)
20. Fernández, M., Overbeeke, C., Sabou, M., Motta, E.: What makes a good ontology? A case-study in fine-grained knowledge reuse. In: Gómez-Pérez, A., Yu, Y., Ding, Y. (eds.) ASWC 2009. LNCS, vol. 5926, pp. 61–75. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-10871-6_5
21. Tartir, S., Arpinar, I.B., Moore, M., Sheth, A.P., Aleman-Meza, B.: OntoQA: metric-based ontology quality analysis. In: The IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources (2005)
22. Lantow, B.: OntoMetrics: application of on-line ontology metric calculation. In: BIR Workshops, pp. 1–12 (2016)
23. 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
24. Wilson, R.S.I., Goonetillake, J.S., Indika, W.A., Ginige, A.: Analysis of ontology quality dimensions, criteria and metrics. In: Gervasi, O., et al. (eds.) ICCSA 2021. LNCS, vol. 12951, pp. 320–337. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86970-0_23
25. Wilson, R.S.I.: A Conceptual Model for Ontology Quality Assessment. http://semantic-web-journal.net/system/files/swj3003.pdf. Accessed 04 Apr 2022
26. 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)
27. 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)
28. 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)

29. McDaniel, M., Storey, V.C.: Evaluating domain ontologies: clarification, classification, and challenges. ACM Comput. Surv. **52**, 1–44 (2019)
30. 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)
31. 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)
32. W3C RDF Validation Service. https://www.w3.org/RDF/Validator/. Accessed 29 Dec 2021
33. Horridge, M.: OWL 2 Validator. http://mowl-power.cs.man.ac.uk:8080/validator/. Accessed 29 Dec 2021
34. Sure, Y., Erdmann, M., Angele, J., Staab, S., Studer, R., Wenke, D.: OntoEdit: collaborative ontology development for the semantic web. In: Horrocks, I., Hendler, J. (eds.) ISWC 2002. LNCS, vol. 2342, pp. 221–235. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-48005-6_18
35. Dividino, R., Romanelli, M., Sonntag, D.: Semiotic-based ontology evaluation tool S-OntoEval. In: Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008) (2008)
36. García-Ramos, S., Otero, A., Fernández-López, M.: OntologyTest: a tool to evaluate ontologies through tests defined by the user. In: Omatu, S., et al. (eds.) IWANN 2009. LNCS, vol. 5518, pp. 91–98. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02481-8_13
37. Schober, D., Tudose, I., Svatek, V., et al.: OntoCheck: verifying ontology naming conventions and metadata completeness in Protégé 4. J. Biomed. Semant. **3**, S4 (2012). https://doi.org/10.1186/2041-1480-3-S2-S4
38. Daga, E.: XDTools - NeOn Wiki. http://neon-toolkit.org/wiki/XDTools.html. Accessed 29 Dec 2021
39. Lambrix, P., Liu, Q.: Debugging the missing is-a structure within taxonomies networked by partial reference alignments. Data Knowl. Eng. **86**, 179–205 (2013)
40. Poveda-Villalón, M., Gómez-Pérez, A., Suárez-Figueroa, M.C.: Oops! (ontology pitfall scanner!): an on-line tool for ontology evaluation. Int. J. Semant. Web. Inf. Syst. IJSWIS. **10**, 7–34 (2014)
41. Lozano-Tello, A., Gomez-Perez, A.: ONTOMETRIC: a method to choose the appropriate ontology. J. Database Manag. JDM **15**, 1–18 (2004)
42. Schekotihin, K., Rodler, P., Schmid, W.: OntoDebug: interactive ontology debugging plug-in for Protégé. In: Ferrarotti, F., Woltran, S. (eds.) FoIKS 2018. LNCS, vol. 10833, pp. 340–359. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-90050-6_19
43. 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)
44. Amith, M., et al.: OntoKeeper: semiotic-driven ontology evaluation tool for biomedical ontologists. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1614–1617 (2018)
45. Kondylakis, H., et al.: Delta: a modular ontology evaluation system. Information **12**, 301 (2021)
46. van Solingen, R., Basili, V., Caldiera, G., Rombach, H.D.: Goal question metric (GQM) approach. In: Encyclopedia of Software Engineering. American Cancer Society (2002)
47. Carriero, V.A., Gangemi, A., Mancinelli, M.L., Nuzzolese, A.G., Presutti, V., Veninata, C.: Pattern-based design applied to cultural heritage knowledge graphs. Semantic Web. **12**, 313–357 (2021)

48. ISO/IEC 25040:2011(en), Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Evaluation process. https://www.iso.org/obp/ui/#iso:std:iso-iec:25040:ed-1:v1:en. Accessed 03 Jan 2022
49. ISO/IEC 25010:2011(en), Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. https://www.iso.org/obp/ui/#iso:std:iso-iec:25010:ed-1:v1:en. Accessed 26 June 2021
50. Walisadeera, A.I., Ginige, A., Wikramanayake, G.N.: User centered ontology for Sri Lankan farmers. Ecol. Inform. **26**, 140–150 (2015). https://doi.org/10.1016/j.ecoinf.2014.07.008
51. Protégé. https://protege.stanford.edu/. Accessed 14 Nov 2021
52. Vrandečić, D., Gangemi, A.: Unit tests for ontologies. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM 2006. LNCS, vol. 4278, pp. 1012–1020. Springer, Heidelberg (2006). https://doi.org/10.1007/11915072_2
53. Fernández-Izquierdo, A., García-Castro, R.: How to validate ontologies with themis. In: Hitzler, P., et al. (eds.) ESWC 2019. LNCS, vol. 11762, pp. 52–57. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32327-1_11
54. List of Reasoners | OWL research at the University of Manchester. http://owl.cs.manchester.ac.uk/tools/list-of-reasoners/. Accessed 14 Nov 2021