Evaluating Domain Ontologies: Clarification, Classification, and Challenges

MELINDA MCDANIEL, Georgia Institute of Technology
VEDA C. STOREY, Georgia State University

The number of applications being developed that require access to knowledge about the real world has increased rapidly over the past two decades. Domain ontologies, which formalize the terms being used in a discipline, have become essential for research in areas such as Machine Learning, the Internet of Things, Robotics, and Natural Language Processing, because they enable separate systems to exchange information. The quality of these domain ontologies, however, must be ensured for meaningful communication. Assessing the quality of domain ontologies for their suitability to potential applications remains difficult, even though a variety of frameworks and metrics have been developed for doing so. This article reviews domain ontology assessment efforts to highlight the work that has been carried out and to clarify the important issues that remain. These assessment efforts are classified into five distinct evaluation approaches and the state of the art of each described. Challenges associated with domain ontology assessment are outlined and recommendations are made for future research and applications.

CCS Concepts: • **General and reference** → **Surveys and overviews**; *Evaluation*; • **Computing methodologies** → **Ontology engineering**; • **Software and its engineering** → **Formal language definitions**; **Semantics**; • **Information systems** → Web Ontology Language (OWL);

Additional Key Words and Phrases: Ontology, evaluation, metrics, domain ontology, applied ontology, ontology development, ontology application, task-ontology fit, assessment

ACM Reference format:
Melinda McDaniel and Veda C. Storey. 2019. Evaluating Domain Ontologies: Clarification, Classification, and Challenges. *ACM Comput. Surv.* 52, 4, Article 70 (September 2019), 44 pages.
https://doi.org/10.1145/3329124

1 INTRODUCTION

The explosive growth of information available on the World Wide Web has resulted in the need to capture and organize vast amounts of information about the real world and how it operates. Ontologies are intended to play a significant role in organizing this information by facilitating seamless information processing and interoperability among applications (Almeida 2009; Chalmeta et al. 2015; Fritzsche et al. 2017; Larsen et al. 2017; Obrst and Cassidy 2011; Roman et al. 2005). A domain ontology provides a formal representation of a specific domain and provides

This research was supported by the College of Computing, Georgia Institute of Technology, and the J. Mack Robinson College of Business, Georgia State University.

Authors' addresses: M. McDaniel, College of Computing, Georgia Institute of Technology, 801 Atlantic Drive, Room 135, Atlanta, GA 30332-0280; V. C. Storey, Computer Information Systems, J. Mack Robinson College of Business, Georgia State University, Atlanta Georgia 30302-4015.

0360-0300/2019/09-ART70 $15.00
https://doi.org/10.1145/3329124

contractual agreements about the meaning of terms within a discipline (Hepp et al. 2006). Thus, high-quality domain ontologies are essential for interoperability between software applications and for accurately modeling a domain of interest (Besheli 2018; Devi and Mittal 2016). Researchers in many areas have all recognized the need for ontologies to clearly define specialized vocabularies for these domains. These include Artificial Intelligence (Milne et al. 2013; Baclawski et al. 2017), Robotics (Sadik et al. 2017), the Internet of Things (Noguera-Arnaldos et al. 2017; Gaglio and Re 2014), Biomedical Informatics (Zhang et al. 2006; Cui et al. 2017), Natural Language Processing (Hirschberg et al. 2015; Brown et al. 2017; Guizzardi et al. 2002), Machine Perception (Dibley et al. 2012; Gemmeke et al. 2017), Machine Learning (Sinha et al. 2016), Database Management (Sugumaran and Storey 2005); and even Climate Science (Camporeale et al. 2015) and Citizen Science/Crowdsourcing (Goudos et al. 2006; Kiptoo et al. 2016; Lukyanenko and Parsons 2018).

The study of ontology deals with the nature of reality, that is, exploring the similarities, differences, and relationships between the types of things that exist in the real world (Guarino et al. 2009). The term "ontology" refers not only to the vocabulary itself but also to the concepts the vocabulary is intended to express (Guizzardi 2007). Domain ontologies, in particular, are content theories about the types of objects, properties of objects, and relationships between objects that are used in a particular domain of knowledge and provide terms for expressing a body of knowledge about the domain (Chandrasekaran et al. 1999). In spite of the recognition of the value of domain ontologies, it is difficult for knowledge engineers to find an appropriate one for a specific application. They are often forced to create an entirely new one, even though many high-quality domain ontologies might be available (Lange 2013). Identifying a suitable ontology for a given task, which we define as *task-ontology fit*, is nontrivial because ontologies are implemented using a variety of languages, methodologies, and platforms. Effective tools are thus needed to adequately address the ontology selection and evaluation problem (Paulheim et al. 2013). Furthermore, challenges arise when (1) automating the comparison of ontologies so the comparison can be carried out in real time (Obrst et al. 2007) and (2) evaluating an ontology for a specific task (Porzel and Malaka 2004). In the case of a knowledge engineer creating a new ontology, evaluating the quality of the resulting ontology is also essential to ensure the task will be performed accurately.

The objectives of this article are to (1) survey the state of the art in domain ontology evaluation, to highlight the most significant efforts and methods; (2) identify research challenges that limit ontology evaluation effectiveness; and (3) propose a framework for overcoming these challenges using an ontology assessment pipeline approach. The contributions are to (1) provide a scheme to classify significant research approaches to domain ontology selection and application; (2) clarify the issues related to each approach; and (3) elucidate the challenges faced. This work is intended to be helpful to those researchers seeking to obtain a deep understanding of domain ontologies and their assessment.

This article proceeds as follows. Section 2 clarifies the need for, and the issues related to, domain ontology evaluation, and defines the key terms needed to do so. Section 3 reviews research on ontology evaluation, classifying them into five evaluation approaches and detailing the strength and weaknesses of each. Section 4 presents a framework in which all five approaches can be combined to improve ontology assessment. Section 5 identifies significant challenges and proposes future research directions. Section 6 summarizes and concludes the article.

2 CLARIFICATION AND CLASSIFICATION

Both theoretical and applied research efforts recognize the need to develop and evaluate domain ontologies for use in many settings (Obrst et al. 2014). As a result, domain ontologies have continued to mature since Gruber (1993) proposed the definition of an ontology for practical use as "an explicit specification of a conceptualization," Dahlgren (1995) suggested a naïve approach to ontology development, and Berners-Lee et al. (2001) called for the development of ontologies as an integral part of the Semantic Web.

Ontology applications focus on creating models of the real world before selecting the representation systems or algorithms to be employed (Guarino and Musen 2005). These models need to be understood in such domains of inquiry as knowledge engineering, information systems modeling, artificial intelligence, formal and computational linguistics, information retrieval, library science, and knowledge management (Guarino and Musen 2005). Although ontologies are the fundamental data structure used to conceptualize knowledge, it is often possible to build many different ontologies to represent the same body of knowledge. Therefore, users (e.g., knowledge engineers) may not know whether a particular ontology will help them fill a need or solve a data or application problem. Communities might not be sure whether large ontologies formed from merging smaller ontologies will improve semantic operability for complex data and application needs (Obrst et al. 2007).

2.1 Ontology Evaluation Problems

There are two distinct ways to consider the ontology evaluation problem. The first, traditionally called the “glass box” or “component” evaluation, examines an ontology based on its individual characteristics. This type of evaluation should be conducted throughout the ontology life cycle to ensure it is of high quality (Hartmann et al. 2005). For a domain ontology, this evaluation assesses whether the ontology accurately, efficiently, and appropriately models the domain for which it is intended to be used. Detailed and correct criteria are needed to make this determination. The second type of ontology evaluation, commonly called “black box” or “task-based” evaluation, is employed when an ontology is tightly integrated into an application and serves to measure the ontology’s overall performance on a specific task (Hartmann et al. 2005). This type of evaluation could also be used when an ontology is being considered for reuse in a new task. Then, it is essential to be able to identify criteria for measuring whether an ontology is suitable for a given need (Brank et al. 2007).

Identifying the criteria for both types of ontology assessment is required in domain ontology evaluation. Many methods have been proposed, frameworks developed, and metrics applied, which are reviewed below. Before doing so, Table 1 summarizes the terms used in this article to aid in understanding the work that has been carried out.

Over time, the field of domain ontology engineering has matured, as have efforts to assess the quality of domain ontologies. There have been many development and assessment initiatives, making it difficult to analyze all of them. Figure 1 presents a timeline that highlights many of the most important ones.

As shown in Figure 1, there has been active research for more than 20 years in domain ontology quality assessment to promote interoperability between computer systems. One of the earliest approaches to ontology evaluation involved (1) identifying what quality attributes of ontologies need to be assessed and (2) developing metrics to assess them. Identifying errors in ontologies and removing them was the next logical step.

Other approaches to ensuring that ontologies are of high enough quality to be used for software system interoperability then emerged. An example is the ontology library approach in which ontologies are stored and maintained by curators who are responsible for their quality. Recent research has attempted to assess the task-ontology fit to determine whether an ontology is appropriate for an intended task (Pittet and Barthélémy 2015; Scheuermann and Leukel 2014). Intermixed with work on evaluating ontologies is research on ontology alignment and modularization of ontologies, whereby their individual components can be assessed separately. Some research efforts have tried to combine approaches but have generally taken different tactics

Table 1. Terms and Definitions Related to Domain Ontology Evaluation

Term	Citation	Definition
Adaptability	Vrandečić (2009)	Measures how well an ontology anticipates its future uses and whether it provides a secure foundation that is easily extended and flexible enough to react predictably to small internal changes
Alignment	Obrst et al. (2006)	Evaluates an ontology by comparing it to a reference ontology whose quality is known
Clarity	Gruber (1995)	Refers to whether an ontology effectively communicates the intended meaning of its defined terms and contains objective definitions that are independent of a particular context
Cohesion	H. Yao et al. (2005)	Refers to the degree to which the elements of a module belong together
Completeness	Gómez-Pérez (1996)	Refers to whether an ontology has sufficiency in its definitions to all possible domains
Conciseness	Gómez-Pérez (1996)	Refers to the absence of redundancies including redundancies that could be inferred from its definitions and axioms
Correctness	Gómez-Pérez (1996)	Refers to whether the concepts, instances, relationships, and properties modeled correlate with those in the world being modeled
Coupling	Orme et al. (2006)	Assesses how well the modules work together in systems of ontologies
Craftsmanship	Neuhaus et al. (2013)	Refers to whether the ontology is built carefully, including its syntactic correctness and consistent implementation
Deployability	Neuhaus et al. (2013)	Refers to whether the deployed ontology meets the requirements of the information system in which it will be used
Domain Ontology	Weber (2002)	Is a conceptualization specific to a particular domain
Domain Fit	Van Lamsweerde (2001)	Evaluation or improvement of an ontology in relationship to its performance on a specific set of tasks
Essence	Guarino and Welty (2002)	Refers to how essential the property is to an entity, and only includes properties that must hold for that entity
Expandability	Gómez-Pérez, (1996)	Refers to the ability of an ontology to be extended in order to describe specific application domains in a way that does not change its current definitions
Expressiveness	Hepp (2007)	Refers to an ontology's degree of detail
Extendibility	Gruber (1995)	Refers to whether a user is able to define new terms for special uses based on the existing vocabulary of an ontology, in a way that does not require the revision of the existing definitions
Fidelity	Neuhaus et al. (2013)	Refers to whether the axioms and the annotations of an ontology represent the intended domain correctly
Fitness	Neuhaus et al. (2013)	Refers to whether the ontology meets the requirements of its intended use
Intelligibility	Neuhaus et al. (2013)	Refers to the ability of all users to understand the intended interpretation of the ontology elements
Interoperability	Tolk and Muguira (2003)	Refers to the ability of two or more systems to communicate effectively both syntactically and semantically
Linked Data	Bizer et al. (2009)	Refers to a set of best practices for publishing and linking data on the web that allows for related data to be easily located
Linked Vocabularies	Vandenbussche et al. (2017)	Refers to an extension of Linked Data in which the data is expanded by providing definitions of the terms used within the datasets
Ontology	Gruber (1995)	Defined as an explicit specification of a conceptualization
Pragmatics	Stamper et al. (2000)	Defined as the relationships between signs and their consequences

(Continued)

Table 1. Continued

Term	Citation	Definition
Pruning	Maedche and Volz (2012)	A means of reducing the size of an ontology or module by removing elements outside of a specific application domain
Reusability	Duque-Ramos et al. (2011)	Refers to the degree to which an ontology, or a portion of an ontology, can be reused for a different purpose or to build other ontologies
Richness	Burton-Jones et al. (2005)	Refers to the proportion and type of features in the ontology language that have been used in a particular ontology
Rigidity	Guarino and Welty (2002)	Refers to a special form of essence in which a property is essential to all its instances
Semantics	Stamper et al. (2000)	Defined as the mapping between a sign and what it represents
Semantic Interoperability	Euzenat (2001)	Refers to the ability to correctly interpret the meaning of information imported from other languages or systems
Semantic Web	Berners-Lee et al. (2001)	Refers to an extension of the current web, in which the semantics of terms found in web pages will be explicitly defined using online ontologies
Semiotics	Sowa (2000)	Defined as the study of signs and used so that one entity can represent another entity to a particular agent
Sensitiveness	Gómez-Pérez, (1996)	Relates to how much a small change in a given definition can alter the existing structure of an ontology
Social Quality	Rittgen (2010)	Defined as the agreement between the interpretations of users and relates to consensus building
Syntactic	Stamper et al. (2000)	Defined as the relationship between signs including their formal logical arrangement
Task Fit	Porzel and Malaka (2004)	Refers to the evaluation of an ontology in relation to its performance on a specific set of tasks
Upper Ontology	Chandrasekaran et al. (1999)	Refers to an ontology that describes knowledge at a high level of generality

for solving the ontology assessment problem. Therefore, a classification scheme would be useful for understanding how these various approaches to ontology evaluation differ and whether there is congruence among them.

2.2 Purpose of Classification

Classification serves two fundamental purposes: (1) it reduces the amount of information that needs to be associated with each instance, and (2) it allows such information to be inferred by means of membership in a class (Parsons and Wand 2008). A well-designed scheme for classifying research contributions should accurately categorize the individual bodies of work that exist. It should also allow for explanations of the reasoning behind each work and provide generalizations between research carried out using the same approach. Most importantly, a good scheme should be useful for identifying gaps, pointing out congruencies, and predicting future research. We therefore classify the work that has been carried out, to develop a manageable number of classes that can aid in the analysis of the domain ontology evaluation.

2.3 Classification Procedures

A set of 172 research papers, published over the last two decades, that focus on ontology quality evaluation was collected. Each of the papers focuses specifically on a particular method for ontology quality assessment and includes validation of the method selected. The papers included work on determining how to measure the quality of an ontology either during development, for

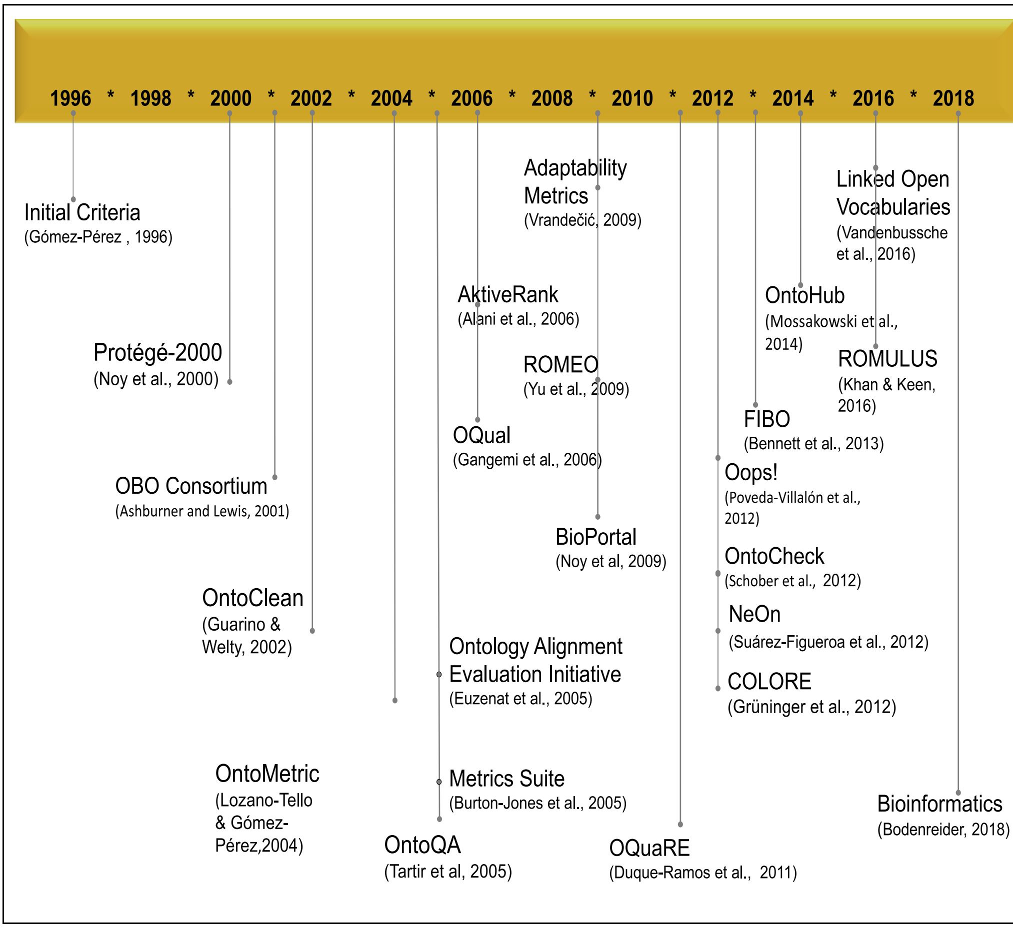

Fig. 1. Timeline of domain ontology evaluation initiatives.

reuse, or for library containment. There are many issues on which these efforts agree, as well as points of departure. The number of papers resulting from the search indicates the importance of the research area. The papers are summarized in Appendix A.

2.3.1 Agglomerative Hierarchical Clustering. To classify the research papers, we applied the agglomerative hierarchical clustering algorithm because it is an iterative classification method that groups objects based on their similarities and dissimilarities (Zhao and Karypis 2002). The similarities (or dissimilarities) can be selected to suit the data being classified. This clustering method is an appropriate choice because our goal was to find similarities between research studies. The algorithm determines the clusters by initially assigning each object to a separate cluster and then repeatedly merging pairs of clusters until an ending condition is reached. An important advantage of using an Agglomerative Hierarchical Clustering algorithm is that a binary clustering tree is produced that shows the progressive grouping of the data as it is iteratively clustered. From the resulting tree, it is possible to identify a reasonable number of classes into which to group the data.

2.3.2 Clustering Procedure. The choice of when to stop clustering, and what similarity or dissimilarity criterion should be applied, is subjective and based on the type of data being classified. To develop a classification system for ontology research, the type of data is textual, specifically, textual information in the form of published research papers. Therefore, the clustering criterion is based on the similarities in the keyword lists that indicate the topics covered by each research paper.

We obtained the papers through an extensive library search using the search terms *ontology assessment* and *ontology evaluation*. Papers that deal simply with ontology development or ontology application were filtered out, so only papers that focused on ontology quality assessment were retained. This initial procedure resulted in the 172 research articles that specifically focused on the evaluation of domain ontologies.

The agglomerative hierarchical clustering algorithm was applied to the set of research papers by using the keywords listed in each paper to identify its focus. (In the case where keyword lists were not included, they were manually extracted from the content of the paper.) Words in the list that were too general, such as *ontology* or *assessment*, which apply to all papers, were removed from consideration. Next the WordNet dictionary (Fellbaum 1998) was used to identify the synonym set for each of the keywords, expanding the lists to make sure that research containing "library," for example, matched with research assigned the keyword "repository." At each iteration of the clustering algorithm, research papers were grouped together that matched on a specific keyword. This process began with the first keyword continuing until all keywords had been included.

The stopping condition for the agglomerative hierarchical clustering algorithm was when the number of research articles became manageable. For our analysis, we repeatedly merged the clusters of papers based on the similarity of keywords pertaining to the research study. The keywords were ranked based on their level of importance to the study. For example, in Guarino and Welty's (2009) research "An Overview of OntoClean," the most common concepts are validation, evaluation, consistency, and pitfalls. The Poveda-Villalón et al. (2012) research "Validating Ontologies with Oops!" the most common concepts are ontology, pitfalls, ontology evaluation, and ontology validation. Therefore, these two ontologies were clustered together during the first round of agglomerative clustering. After the first round, over 40 clusters remained, which is an unmanageable number for a classification scheme. Therefore, for the second round, the keyword list of each research study was expanded to include synonyms of the original terms in the keyword list, using the synonym functionality of the WordNet lexical database for English (Fellbaum et al. 1998). For example, the term *pitfall* was expanded to include *error, mistake, drawback*, and *difficulty*. Then, the research by Schober et al. (2012), "OntoCheck: Verifying Ontology Naming Conventions and Metadata Completeness in Protégé 4," for example, could be added to the cluster because its keyword list included *error checking*. At the end of the second round of cluster analysis, there were 15 clusters, so the synonym list for each of the newly expanded keywords was also added to the new keyword list. After this round, five clusters emerged, a manageable number for a classification scheme. Algorithm 1 presents the agglomerative clustering algorithm.

From the cluster analysis, five distinct classes of research in ontology selection and evaluation emerged: (1) error checking and ontology cleaning; (2) creating domain-specific libraries in which to store and maintain the ontologies; (3) using metrics to quantify an ontology's quality; (4) modularizing ontologies to streamline the assessment task; and (5) determining an ontology's fitness for a specific task. Table 2 defines these classes.

Most ontology evaluation research focuses on a single method to determine quality. For example, OntoClean (Guarino and Welty 2002) and Oops (Poveda-Villalón et al. 2012) assess the errors or lack or errors in each. COLORE (Grüninger and Katsumi 2012), BioPortal (Noy et al. 2009), and OntoHub (Mossakowski et al. 2014, Codescu et al. 2017) focus on creating ontology libraries that allow for communities to assess and maintain the ontologies. OntoMetric (Lozano-Tello and Gómez-Pérez 2004), OntoQA (Tartir et al. 2010), and OQuaRE (Duque-Ramos et al. 2011) each consist of suites of metrics that can be used to obtain an overall evaluation of ontology quality.

Table 2. Classes of Ontology Evaluation Research

Class	Description
Domain/Task Fit	Identification of how an ontology can be developed or selected based on a specific use
Error Checking	Identification and removal of errors from ontologies
Libraries	Establishment of repositories for ontologies sharing a common domain or ontology language
Metrics	Development of metrics to assess the quality of ontologies based on specific attributes
Modularization	Subdivision of ontologies into smaller modules each with a separate purpose

ALGORITHM 1: Clustering Algorithm

```
for each paper
    assign the paper and its corresponding keyword list to an individual cluster
end
repeat
    for each cluster
      let n1 be the number of keywords in the cluster
        for each of the other clusters
            let n2 be the number of keywords in that cluster
          initialize match_rate = 0
            for each keyword, k1, in the first cluster:
                for each keyword, k2, in another cluster
                    if k1 == k2 then
                        match_rate = match_rate+ (n1 - k1's position in the keyword list + 1) +
                                                  (n2 - k2's position in the keyword list + 1)
    end
    for each cluster
      identify the cluster with the highest match_rate that is > 0
      merge the keyword list from that cluster with this cluster to form a new cluster
      for each keyword in the new keyword list
        append onto the list its synonym list from WordNet
    end
until number of clusters < threshold
```

The Financial Industry Business Ontology (FIBO) project (Bennett 2016; Fritzsche et al. 2017) and the Requirements Oriented Repository for Modular Ontologies (ROMULUS) (Khan & Keet 2016) both focus on the modularization of ontologies, splitting them into submodules that can be used separately. The NeOn project (Suárez-Figueroa et al. 2012) focuses on the goals required by the task for which an ontology is being developed and evaluates whether that ontology is meeting these goals.

Figure 2 identifies the five classes along with a selected group of representative studies. The studies included were selected because they are frequently cited and have influenced ontology evaluation efforts.

These five classes serve as the starting point for organizing the enormous amount of work being carried out in ontology evaluation, in an attempt to establish a common foundation upon which to build further research.

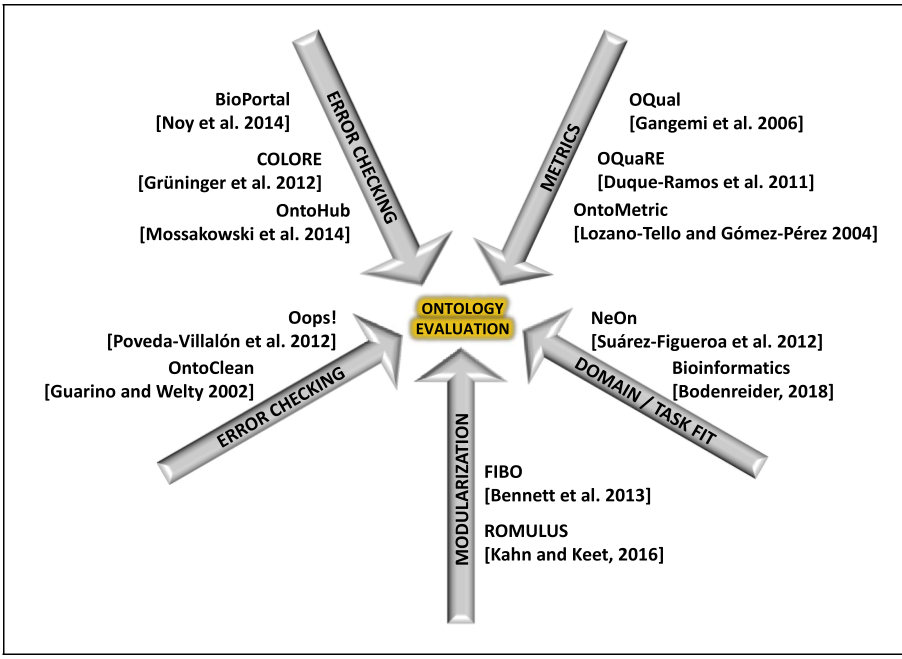

Fig. 2. Ontology evaluation research classes.

2.3.3 Expanded Search and Classification. The first search focused only on the literature where ontology evaluation was the main topic of the research. We then extended the search to those studies when evaluation was secondary to ensure that no topic had been omitted from our consideration. To do so, 1,252 research articles published within this decade (2010 through 2019) were extracted from the Web of Science Core Collection of six online databases (Science Citation Index Expanded (SCI-EXPANDED), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), Conference Proceedings Citation Index—Science (CPCI-S), Conference Proceedings Citation Index—Social Science & Humanities (CPCI-SSH), Book Citation Index—Science (BKCI-S), and Emerging Sources Citation Index (ESCI)). These articles are identified in an accompanying, online supplement.

The 1,252 articles were selected by searching for all scientific papers related to ontologies, which included an evaluation portion that was at least significant enough to appear in the abstract. The papers were then filtered by verifying that they were, indeed, articles pertaining to ontology evaluation. That is, the papers either (1) had ontology evaluation as the paper's subject or (2) reported on a specific ontology, ontology tool, or ontology method and included an actual evaluation of the tool, method, or ontology. A paper was not included if it simply mentioned the need for ontology evaluation. This resulted in the selection of 1,208 articles. (The references for the search articles are included in a separate, supplementary document.) Only a small portion of the retrieved abstract was actually related to our research topic, that is, the evaluation and validation of ontologies. Most of the abstract pertained to the paper's main purpose: a new ontology for a specific task or a new method for creating, matching, or searching ontologies. Therefore, the abstracts were preprocessed to identify only the portion related to ontology evaluation.

A modified version of the TextRank algorithm, an unsupervised machine-learning technique developed by Mihalcea and Tarau (2004) for the automated summarization of texts, was used to create a keyword list for each of the articles based on the related portion of the abstract, the title of the paper, and the author-supplied keyword list. The TextRank algorithm identifies patterns in a complete set of text to identify what makes each individual paper unique. For example, if, in almost all of the corpus (all 1,252 papers) the term *measure* appeared, this term would not be included in the keyword list for one particular article. In other words, the keyword list is the list of words for each article that sets that article apart from the rest of the corpus. In this way, the articles were preprocessed to produce a keyword list from the abstract, title, and author-provided keyword list retrieved from the Web of Science. The TextRank algorithm was applied using the summarize capabilities of the Gensim Python module developed by Řehůřek and Sojka (2011). Both the TextRank algorithm and the Gensim module have been shown effective in research for summarizing Yelp reviews (Taddy 2015), many other types of web documents (Roul and Sahay 2014; Maslova and Potapov 2017; Mukku et al. 2016), and even text in other languages such as found in the work of Al-Anzi and AbuZeina (2017).

The next step was to cluster the keyword lists using another machine-learning technique that groups keyword lists based on their similarity to each other. This Natural Language Processing technique called Latent Semantic Analysis (LSA) takes advantage of the WordNet library of synonyms and matches documents together, even if the words were not exact. Using the corpus of all 1,252 keyword lists, the LSA revealed that some of the lists have more in common with each other than with other lists and produces a document term matrix (DTM) that can be analyzed for patterns in the text. Another LSA technique, singular value decomposition (SVD), reduces the text data into a manageable number of dimensions for the analysis to cluster the terms in the matrix. The Gensim Python module (Řehůřek and Sojka 2011) was used to perform the SVD and has been shown to be effective for text classification in the work of Campr and Jezek (2015), Adams and Bedrick (2014), Huang et al. (2014), and Wang et al. (2017). By setting a similarity threshold, documents are identified based on the similarities of their terms. If the threshold is set too low, the conditions to determine that two documents match are low, resulting in many groups; a higher threshold results in much fewer groups. Because some researchers used more than one evaluation method, a paper could be clustered into more than one group. After experimenting with different thresholds, a 30% threshold was selected, which resulted in four clusters.

To identify the main evaluation method used by each of the clusters, we manually examined both the merged keyword lists and the clustered abstracts. This led to the four classes shown below, listed in the order of how many papers were grouped into each class, from greatest to least:

(1) Using models to prove that the ontology adequately matches its intended domain
(2) Checking the ontology for errors
(3) Using metrics to assess the quality of the ontology
(4) Using the ontology for a specific task to show its adequacy and usefulness

Although not matching our original set of classes exactly, each of the classes that emerged from the second clustering had previously been identified in our first clustering algorithm. This indicates that we did not overlook any significant topics related to ontology evaluation research, our primary goal in conducting this expanded literature search.

Figure 3 shows the expanded literature search process, inspired by the approaches of Sidorova et al. (2008) and Larsen et al. (2008) and by the recommendations of Evangelopoulos et al. (2012). An example is detailed in Appendix B.

3 CLASSES OF ONTOLOGY EVALUATION RESEARCH

The five ontology assessment classes identified by our original clustering algorithm each has a specific focus for the evaluation of ontologies. Research that focuses on removing errors from ontologies is categorized into the *Error Checking* class. Research that focuses on establishing

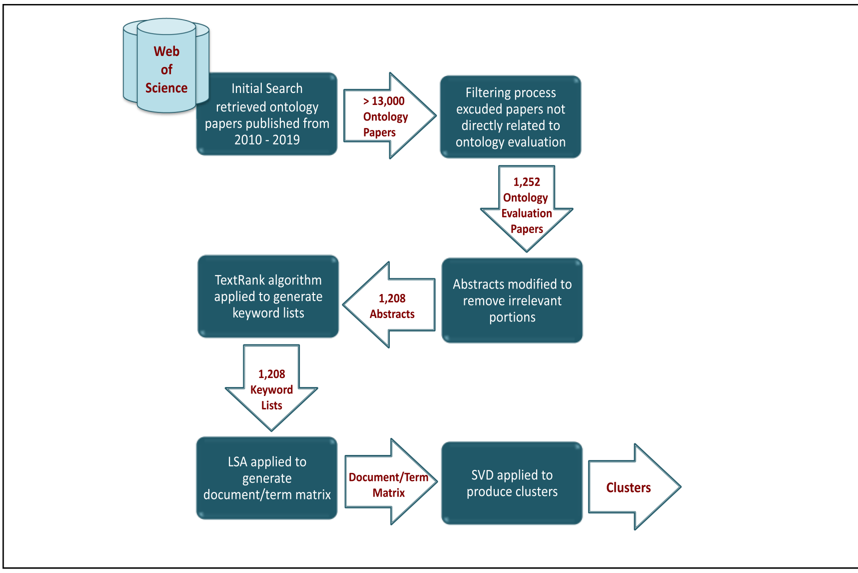

Fig. 3. Overview of expanded literature clustering process.

repositories to store and maintain the quality of ontologies is categorized into the *Libraries* class. Research on developing metrics to quantify certain attributes of ontology quality appears in the *Metrics* class. The research efforts that subdivide ontologies into smaller modules for easier assessment and application are in the *Modularization* class. Finally, research that attempts to assess an ontology's fitness for a particular task, or within a specific context, falls into the *Domain/Task Fit* class. Each class is described below, identifying its advantages and disadvantages and providing representative examples.

3.1 Domain/Task Fit

Domain/Task fit is the evaluation or improvement of a given ontology in relationship to its performance on a specific set of tasks. Software requirements engineering includes research on specifying the goal of a software product first before developing the software, in an effort to ensure that the goals are being met and are the intended ones (Van Lamsweerde 2001). This approach to software creation can also be applied to ontology creation and selection. Knowing for which task an ontology will be used can help select or build an ontology specific to that goal. For this method to work, however, ontology engineers must have a precisely defined, and realistically achievable, goal. Another approach is to evaluate a pre-existing ontology's fitness relative to a specific goal. This effectiveness can be quantified only if there is a measurable way to assess the performance of the ontology for a given task (Porzel and Malaka 2004). Porzel and Malaka (2004), for example, created a methodology to evaluate ontology performance on given tasks and then augment them to better fit the task requirements. By performing a language translation task, heavily dependent upon the ontology for accurate results, they were able to identify errors such as superfluous, missing, or ambiguous concepts and then correct the ontology to reduce these errors. The choice of ontology for a strongly language-dependent application is crucial to achieving correct results.

Brewster et al. (2004) developed data-driven techniques for ontology assessment. Their method assesses how well an ontology fits a given corpus by examining the internal structure of the ontology. By understanding how closely the terms of the corpus are clustered in the ontology, an assessment can be made of the ontology's fitness. The NeOn methodology for ontology creation also takes into consideration specific goals when an ontology is being developed, as well as the input, output, and constraints of a task (Suárez-Figueroa et al. 2012). The NeOn methodology includes procedures for ontology selection, reuse, and re-engineering with each process that is part of the NeOn framework, including an assessment to ensure that the goals are being met.

Most tasks for which ontologies are needed are tightly bound to a specific domain. For example, biomedical ontologies provide resources for clinical decision support systems and data integration systems for medical research and health analytics, with the quality of these resources having a direct impact on health care and biomedical research (Bodenreider 2018). It is challenging to separate the task for which an ontology is needed from the domain's coverage area. When evaluating the quality of an ontology, researchers assess the ontology's performance on a specific task or tasks and on how accurate and applicable the ontology is for the task and domain. Therefore, both ontology task fitness and ontology domain fitness are categorized in a single class.

Advantages: If an ontology is to be used for a given task, as is the case for most domain ontologies, its quality is directly related to how effectively it performs an intended task (Kim and Storey 2012). An evaluation method that includes a measurement of the ontology's performance on a specific task would provide an excellent indication of quality (Porzel and Malaka 2004). Although this type of measurement is difficult to achieve, it is essential for a complete evaluation, if the evaluation is being carried out for the selection of an ontology for reuse or repurposing (Choi et al. 2006).

Challenges: For ontology selection to be performed by a computer system, which is a goal of the Semantic Web and a necessity when dealing with large ontologies (Whetzel et al. 2011), a program needs to be able to calculate a metric that is an aggregate number, representing an overall quality evaluation. The problem of assigning a value to the fitness metric (relevance) is difficult because a metric for fitness is not easily obtained. For most ontology quality assessment metrics suites, ontology concepts, including both classes and properties, are compared to the terminology used in the domain (Strasunskas and Tomassen 2008). This type of concept matching is an oversimplification of the complex nature of matching a domain to a particular ontology. The relationship between an ontology and a conceptualization is dependent upon (1) the agent that conceives the conceptualization and (2) the means by which it is encoded. Therefore, at best, a fitness measurement can only be an approximation (Gangemi et al. 2006).

3.2 Error Checking

Research to identify errors and "clean" them (Guarino and Welty 2002; Gómez-Pérez 2001) includes error types ranging from simple syntax errors to complicated semantic and structural problems. Syntax is built upon rules for the construction of expressions; identifying that a particular rule has been broken is usually quite straightforward. Semantics, however, deals with meanings, so contradictory meanings and faulty interpretations of meaning must also be identified. Table 3 provides examples of the type of error that might occur in an ontology, based on error-checking examples in the literature.

Figure 4 illustrates an example from a construction domain ontology (El-Diraby et al. 2005). A French door cannot be both an opening window and a glass door because a glass door is a subclass of door, which is disjoint from window.

Early work on ontology cleaning evaluated ontologies to ensure they met specific, basic requirements for validity. Gómez-Pérez (1996), for example, proposed a framework that identifies redundancy errors, semantic errors, and incompleteness. OntoClean (Guarino and Welty 2002) was developed to assess ontologies using the formal notions of essence and rigor. The OntoClean

Table 3. Examples of Possible Errors in Domain Ontologies

Description of Error	Type of Error
Using different naming criteria in the ontology	Syntax error
Missing inverse relationships	Syntax error
Using a recursive definition	Syntax error
Merging different concepts in the same class	Semantic error
Missing equivalent properties	Semantic error
Missing disjointness relationships	Semantic error
Creating synonyms as new classes	Semantic error
Swapping intersection and union	Semantic error

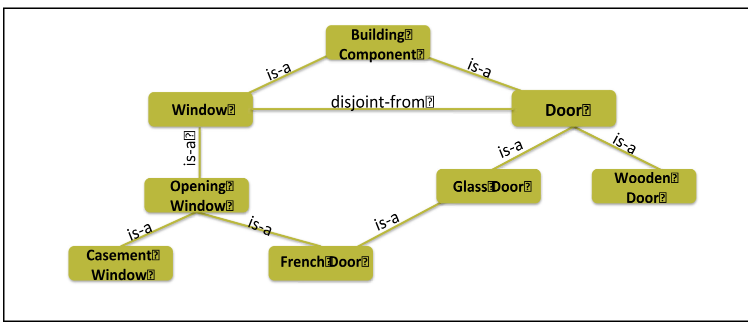

Fig. 4. Consistency error in building construction ontology (adapted from El-Diraby et al. (2005)).

framework consists of two steps. First, concepts in an ontology are tagged according to the meta-properties of rigidity, unity, dependency, and identity. Next, the tagged concepts are checked for errors using predefined constraints dependent on the assigned tags. Aeon (Völker et al. 2008) is an attempt to automate the well-known OntoClean methodology (Guarino and Welty 2002) in order to reduce costs and to improve interoperability between software systems.

Oops! (Poveda-Villalón et al. 2014) is a simple-to-use, web-based tool that provides automatic checking for common errors, called *pitfalls*, such as naming conflicts or consistency problems in ontologies that have been uploaded by users. The error list can be easily expanded to include other types of errors, and the user interface is designed to be user friendly. Oops! does not attempt to correct the errors found in an ontology; rather, a comprehensive list is provided to the user that includes a description and severity level for each error.

Advantages: The error-checking approach to ontology evaluation has the potential to be automated. This could be extremely advantageous for very large ontologies such as those for the biomedical domain, some of which contain hundreds of thousands of classes (Noy et al. 2009). Although not all types of errors or potential errors are easily located, any removal of common errors or structural problems can be effective for improving the usefulness of an ontology.

Challenges: Error-checking methods, although providing useful information, do not provide a thorough enough evaluation of an ontology to solve the challenges related to selecting an ontology to fit a particular domain or task. This method would need to be combined with other approaches to provide an appropriate selection.

3.3 Ontology Libraries

Because it is less expensive for data providers to reuse existing, well-established ontologies than to create new ones, ontology libraries have been developed (d'Aquin and Noy 2012). Some of these

Table 4. Examples of Ontology Libraries

Ontology Library	Objective
BioPortal	A searchable repository for biomedical ontologies that includes tools for ontology evaluation and recommendation (Noy et al. 2009)
COLORE	A repository of ontologies that supports the design, evaluation, and application of ontologies through first-order logic (Grüninger 2009)
Linked Open Vocabularies	A high-quality catalog of reusable vocabularies for the description of data on the web (Vandenbussche et al. 2015)
OntoHub	An open ontology repository for Distributed Ontology Language-conforming ontologies (Mossakowski et al. 2014)
ROMULUS	A requirements-oriented repository for modular ontologies (Khan and Keet 2016)

libraries store and maintain ontologies related to a specific domain, such as the BioPortal ontology for biomedical ontologies (Noy et al. 2009). Others are multipurpose libraries allowing not only domain ontologies from many different domains to reside there but also high-level ontologies and other types of vocabularies or schemas. Table 4 shows representative examples of ontology libraries and their stated objectives.

Reuse of existing ontologies improves semantic interoperability because, when knowledge engineers use the same ontology, integration between applications is easier (d'Aquin and Noy 2012). As the number of new ontologies increases, more libraries will be needed, requiring different versions of evaluation systems for comparison (Grüninger et al. 2012). Some of these libraries are built using automated systems, such as OntoSelect (Buitelaar et al. 2004), which monitor the World Wide Web for newly published ontologies that match a specified format and add them to a library. Quality assessment is even more essential in these automatically created ontology libraries to ensure they are within acceptable quality levels (Buitelaar et al. 2004).

The Open Biomedical Ontologies (OBO) consortium (Ashburner and Lewis 2001) was established to identify the best practices for the development of bioscience ontologies. Guidelines were needed to deal with the vast amount of available data associated with the biosciences and the rapidly expanding number of ontologies being developed to store that data. This work has progressed to the point where it also includes the OBO foundry, which serves as a repository for biomedical ontologies designed using established guidelines (Ashburner and Lewis 2001; Smith et al. 2007).

An extension to ontology libraries is linked vocabularies, which usually include not only ontologies but also metadata, vocabularies, and dictionaries. The Linked Open Vocabulary (LOV) project provides a vocabulary collection that is maintained by curators who are responsible for ensuring the quality of the vocabulary. The latest version of the LOV system includes an automated portion, with human intervention ensuring that any vocabulary included in this library can be trusted (Vandenbussche et al. 2015).

Advantages: Assessing ontology quality within a given community has the advantage of providing specific domain knowledge that the community members possess (Hepp et al. 2006). Although general-purpose ontology repositories exist, such as OntoHub (Mossakowski et al. 2014) and COLORE (Grüninger and Katsumi 2014), most ontology libraries are domain specific. One of the largest is BioPortal (Noy et al. 2009), an open repository of biomedical ontologies. It provides access to existing ontologies and has the capability for a user to add new ontologies and notes; contribute mappings between terms; and review ontologies based on criteria such as usability,

Table 5. Basic Ontology Quality Attributes (Gómez-Pérez 1996)

Attribute	Definition
Consistency	Whether it is possible to obtain contradictory conclusions from valid input data
Completeness	Whether all that is supposed to be in the ontology is explicitly set out in it or can be inferred using other definitions and axioms
Conciseness	Whether all the information gathered in the ontology is useful and precise
Expandability	The ease of adding new definitions and new information to an existing definition without altering the set of well-defined properties that are already guaranteed
Sensitiveness	How small changes in a given definition alter the set of well-defined properties that are already guaranteed

coverage of the domain, accuracy, and level of available documentation. BioPortal also includes a recommender system that provides users with a list of ontologies that match a specific domain in order to assess their quality based on domain coverage, community acceptance, detail of knowledge, and amount of specialization (Martínez-Romero et al. 2017).

Challenges: Although some libraries standardize the web ontology language and the file format used, most libraries allow uploading of ontologies to their repository in any readable ontology language (Gonçalves et al. 2017). This lack of consistency creates challenges for carrying out ontology evaluation within the libraries. There are also redundancies between the libraries that could be avoided if the existing ontology library systems were able to network with each other, sharing both content and ontology application tools (d'Aquin and Noy 2012).

3.4 Metric Based

Ontology evaluation is best carried out by software, rather than by humans (Hendler and Berners-Lee 2010), especially for applications such as those for the Semantic Web (Obrst et al. 2014). For this to happen, however, an objective, rather than subjective, ontology assessment is required to translate a quality level based on specific attributes into numerical scores. This is indeed challenging though, because there is no consensus on which attributes of an ontology correlate to a high level of quality for diverse domains and applications (Obrst 2007; Thalheim 2013).

The systematic evaluation of ontologies requires accurate, well-defined, and easy-to-apply metrics. The metrics can be used to evaluate the quality of an ontology or to compare ontologies when there are multiple candidates that fit a set of requirements (Tartir et al. 2005). Applying metrics, rather than assessing an ontology as simply effective or ineffective, can evaluate specific aspects of it (Tartir et al. 2005). Furthermore, for a given domain, some attributes of an ontology might be more significant than others. As a result, suites of metrics have been developed to provide the user with the ability to weigh each aspect differently. In response, many ontology assessment metrics have been developed. They range from very specific metrics focusing on only one aspect of ontology quality to complex suites of metrics that attempt to provide an overall rating for an ontology, based on all aspects of its quality.

Gómez-Pérez (1996) identified consistency, completeness, conciseness, expandability, and sensitiveness as the most important qualities an ontology must possess. These five, as defined in Table 5, are important for verifying the definitions and axioms that are explicitly included in an ontology, as well as those that can be inferred. Over the past two decades, these five basic attributes have continued to appear in many evaluation frameworks and tools (e.g., Burton-Jones et al. 2005; Gangemi et al. 2002; Lozano-Tello and Gómez-Pérez 2004; Jones et al. 1998; Kang et al. 2012; McDaniel et al.

2018; Völker et al. 2008; Vrandečić 2009; Zhang et al. 2010). This work was an early recognition of attributes that could be used to assess ontology quality.

Representative examples of quality assessment methods that use metric-based approaches are discussed below. These examples were selected based on their proven contribution to the field. They have expanded upon Gomez-Perez's (1996) original five basic attributes and shown the necessity for additional metrics to provide a true valuation of an ontology.

They thus contribute toward the set of assessment categories recognized as essential by the 2013 Ontology Summit on ontology evaluation (Neuhaus et al. 2013), which focused on a life cycle of ontology development and use.

Several studies provide a broad assessment of an ontology's quality from multiple perspectives. Lozano-Tello and Gómez-Pérez (2004) developed a hierarchical framework, OntoMetric, which has 160 characteristics (e.g., Ease of Integration, Quality of Manuals, Number of Different Domains, etc.), organized along five dimensions to evaluate the quality of an ontology and its suitability to a user's requirements (Lozano-Tello and Gómez-Pérez 2004). These five dimensions capture (1) the content of the ontology, (2) the language used, (3) the method of development, (4) the building tools and (5) the associated costs. OntoMetric recognized (two decades before the ontology summit on ontology evaluation) that, to determine the quality of an ontology, one must understand that an ontology is not only a model of the language for human and computer representation but also a portion of the deployed software that incorporates a larger system (Neuhaus et al. 2013).

Burton-Jones et al. (2005) propose a metric suite that is built upon a semiotic framework for sign quality assessment (Stamper et al. 2000) that evaluates whether symbols are good or bad, clear or unclear. Since ontologies use symbols to describe terms and the relationships among them, semiotic theory provided an appropriate theoretical basis from which to derive the metric suite.

OntoQA (Tartir et al. 2010) separates its metrics into two classes: schema metrics and knowledge base metrics. Schema metrics measure the success of a schema in modeling a real-world domain by evaluating its structure. Knowledge base metrics assess whether a populated ontology is a rich and accurate representation of the real world by evaluating its content. Together, the resulting set of metrics can assist a user in deciding whether a specific ontology is suitable for an individual's needs.

OQuaRE (Duque-Ramos et al. 2011) provides both a model and a set of metrics to assess ontologies based on established standards for software quality evaluation. Acknowledging that an ontology can be considered a software artifact, the characteristics of reliability, operability, maintainability, compatibility, transferability, and functional adequacy from software standards are reused and adapted to evaluate ontological quality. Each characteristic is broken down into multiple subcharacteristics with a metric provided for assessing its value. For example, the characteristic of *maintainability* is subdivided into *modularity, reusability,* and *analyzability*, each of which has a set of metrics to assess its value for an ontology.

Advantages: Metric-based techniques quantify the quality of a particular attribute of an ontology (Tartir et al. 2005). A numerical quality score can help a user make the best selection if multiple ontologies are available. Having an evaluation mechanism during the design process of a new ontology is also advantageous because designers could make changes before the ontology reaches its final form. Probably the biggest advantage of this approach is that, if the calculation can be automated, then machines can calculate and compare the ontology quality assessment scores (McDaniel et al. 2018; Obrst et al. 2007).

Challenges: Because of the attractive nature of quantifiable metrics, much work has been attempted using this approach to ontology assessment. An important aspect of the ontology development process is being able to "prove" that an ontology is of high quality and appropriate for the domain for which it is used, even though there is no generally accepted procedure for doing

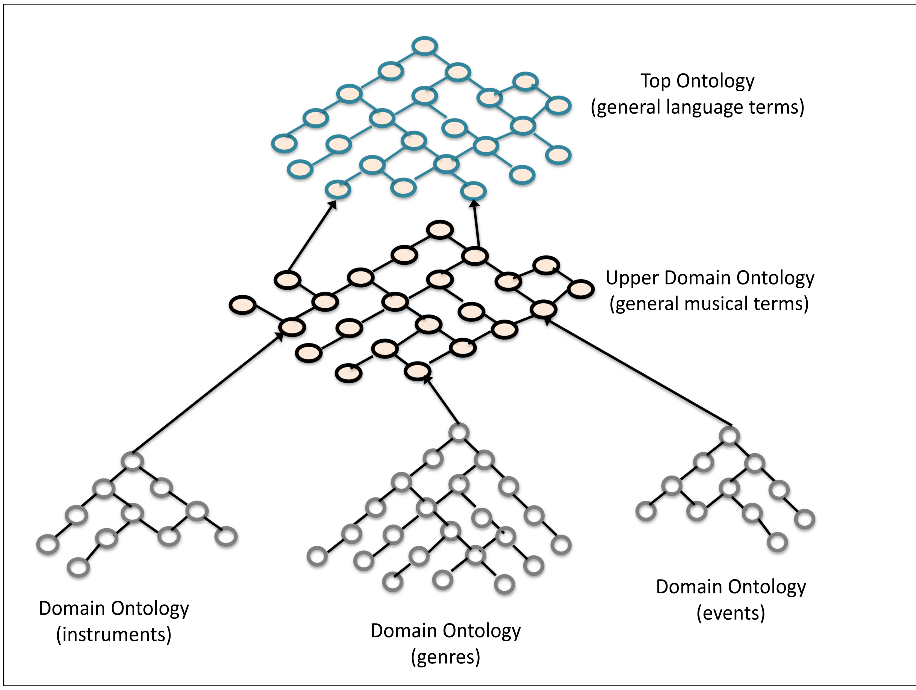

Fig. 5. Modularization for a music ontology (Han et al. 2010).

so (Neuhaus et al. 2013). Proving that the metrics used are, indeed, valid and appropriate requires repeated, empirical studies.

3.5 Modularization

The idea of modularization of ontologies is derived from software engineering and refers to a way of developing structured software so that it is easy to understand, maintain, and reuse (d'Aquin et al. 2007). Software that is divided into smaller pieces, and thus modular, is easier to understand and apply, especially if more than one person is involved in the software's development and use. Modularizing existing large ontologies, or developing new ontologies in a modular nature, allows for distribution of effort, greater control over visibility, and increased scalability because each module can be developed and controlled independently. A modularly designed ontology usually consists of several layers of ontology files frequently including a top-level ontology for general concepts shared by all modules; one or more midlevel domain ontologies including terms used more in some of the modules; and several domain-specific modules related to specific domains. Figure 5 illustrates how a modular ontology of music could be developed using a top ontology, an ontology for the entire music domain, and separate domain-specific modules for terms related to specific concepts related to music.

The advantages of modular software are especially significant for ontologies (Grau et al. 2006). As ontologies increase in size, it becomes more important for portions of the ontology to be verified and reused individually to meet specific requirements and to improve scalability. The ability to reuse parts of an ontology is only possible if the portions are truly completely separate modules that can be extracted without loss of meaning (Khan and Keet 2015). Domain ontologies should therefore be created so there are extractable parts that can be reused outside of the original context of the complete ontology. The use of individual modules requires each module to be of high

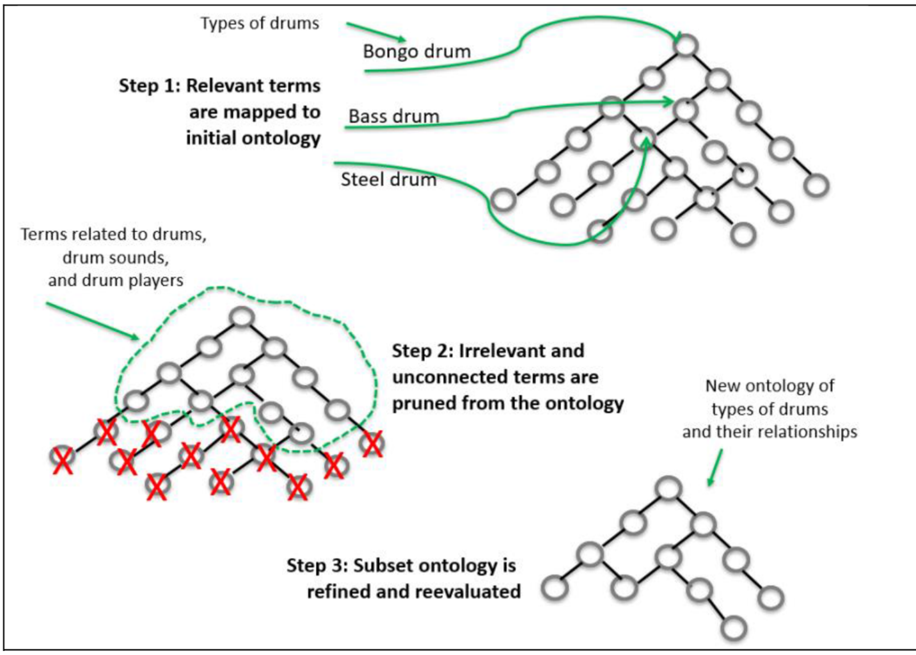

Fig. 6. Pruning irrelevant terms from a music ontology.

quality, fit the intended task, and be used independently from the rest of the original ontology. Therefore, there must be a way to evaluate the semantic, syntactic, and pragmatic quality of individual modules, as well the overall modularization of an ontology (Kutz and Hois 2012).

If an ontology was not originally developed modularly, it is still possible to separate it into separate modules using a process called *pruning*. The goal of pruning, which reduces the size of an ontology or module by removing elements outside of a specific application domain (Maedche and Volz 2012), is to create a balance between the completeness and preciseness of an ontology. Creating a totally complete model of a domain may lead to an ontology that is overly large, unwieldy, and hard to manage. On the other hand, a model of a domain that is too narrowly focused could lead to an ontology with limited expressiveness. The goal is to create a single ontology, or a set of ontology modules, that provide a rich conceptualization of the target domain, while excluding any parts that are outside of its specific focus (Maedche and Volz 2012). A system to measure how closely this balance is attained should be a part of any overall ontology assessment. Figure 6 illustrates how pruning could be used to remove unnecessary concepts from the music ontology, COMUS (Han et al. 2010), while maintaining the connections between the remaining nodes.

Being able to assess the quality of individual modules would greatly aid in ontology reusability because specialized modules could be combined together to form a complete ontology that accurately models a new domain. The Financial Industry Business Ontology (FIBO), for example, is an extensive ontology created from a large number of smaller ontologies, each of which models a specific financial area (Fritzsche et al. 2017).

The FIBO initiative includes two levels of ontologies: (1) core domains, which represent industry terms and relationships at the level of conceptual models not dependent upon application restraints, and (2) modules, which are operational ontologies developed for specific use cases and therefore dependent upon the restraints of specific applications (Neuhaus et al. 2013). Currently

FIBO consists of 11 core domains and 49 modules as well as over 400 individual ontology files (Atkins 2017).

Advantages: Creating ontologies from more specialized, focused, and self-contained modules greatly improves ontology evolution and reuse. The combination and extension of smaller modules and patterns can result in the formation of larger ontologies containing only the most relevant information. Less human intervention is necessary if the individual modules are drawn from ontologies in which the usefulness and quality had already been proven.

Challenges: The very act of extracting modules must not detract from the syntactic, semantic, and pragmatic quality of the individual modules. Further assessment is needed to determine whether this is indeed the case.

3.6 Combination Examples

To improve the accuracy of evaluation results and to overcome the challenges associated with each, some researchers combine two or more approaches in creative ways. Efforts have been made to create metrics that measure the quality of the individual modules when a modular approach is employed. Much research has focused on attributes of cohesion, coupling, and complexity because of their contribution to assessing whether an ontology is easily adapted for a new purpose (Orme et al. 2006; Yang et al. 2006; Yao et al. 2005; Ouyang et al. 2011). Yao et al. (2005), for example, establish cohesion as a fundamental characteristic of an ontology and proposed a set of metrics to measure cohesion for modular ontologies. Oh et al. (2011) add additional ontology modularity metrics to find a way to quantify the relationship between entities in an ontology. Ma et al. (2010) propose four additional ontology cohesion metrics based on ontological semantics, and Orme et al. (2006) define coupling metrics based on commonly accepted software engineering measurements, which assess how well the modules work together in systems of ontologies.

The ROMEO methodology combines the use of metrics with task fit to match the task for which the ontology will be needed to specify metrics that evaluate the ontology's suitability to that application. The methodology consists of three steps: (1) ask the user questions about the task, (2) map the task to specific metrics, and (3) assess the ontology's quality based on the specific metrics identified as relevant (Yu et al. 2009).

One effective means of combining research approaches is by adding error-checking and metric computation capabilities to an ontology editor, thus providing quality assessment to any ontology that is edited. Protégé is the most widely used environment for ontology development and modification (Khondoker and Mueller 2010). This tool provides a hierarchical structure of an ontology's contents as well as valuable information about its classes and axioms (Gennari et al. 2003). As an open-source project, plug-ins are available to expand its capabilities. One of these, OntoCheck (Schober et al. 2012), adds verification of naming conventions and metadata completeness to provide quality evaluation.

3.7 Evaluation throughout the Life Cycle

Gómez-Pérez (1996) was the first to acknowledge that an ontology by its nature is incomplete, because it is impossible to capture everything known about the real world in a finite structure. She therefore called for the verification of complete, consistent, and concise definitions at all stages of the ontological development process. At least one of the ontology assessment approaches should be applied at each step of the process to ensure that a minimum level of quality is maintained.

The Life Cycle Evaluation Approach, resulting from the 2013 Ontology Summit (Neuhaus et al. 2013), proposes that ontologies should be evaluated throughout the life cycle of their development and use. An extensive literature review (Ontology Evaluation Across the Ontology Lifecycle 2013), as part of the summit, identified a lack of consistency in methods for evaluating ontologies,

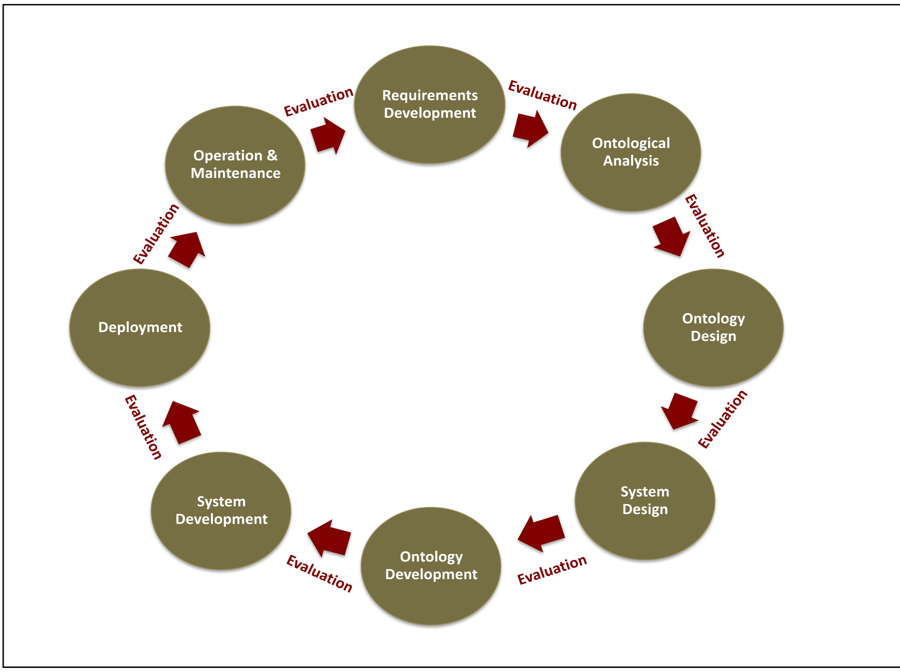

Fig. 7. The Life Cycle model of ontology evaluation (adapted from Neuhaus et al. (2013)).

resulting in many ontologies being developed without applying proper evaluation techniques or tools. Ontologies are described as being "human-intelligible and machine-interpretable representations of some portions and aspects of a domain" (Neuhaus et al. 2013, p. 180). To be both human intelligible and machine interpretable, however, an ontology must be recognized as (1) a domain model for human consumption, (2) a domain model for machine consumption, and (3) deployed software that is part of a larger system (Neuhaus et al. 2013). Five high-level characteristics must be evaluated throughout all phases of ontology development and use: intelligibility, fidelity, craftsmanship, fitness, and deployability. Phases identified as part of the life cycle of an ontology are ontological analysis, ontology design, system design, ontology development and reuse, system development and integration, deployment, and operation and maintenance. Competency questions should be answered at each phase of the ontology life cycle to improve the overall quality of ontologies being deployed. Figure 7 illustrates how this cycle should occur. Combining domain ontology assessment throughout the life cycle with more exploration of the combinations of approaches would most likely progress domain ontology assessment.

3.8 Summary of Evaluation Classes

Table 6 presents the five classes of ontology assessment, summarizing the advantages and challenges of each.

As Table 6 indicates, using the five classes seems to be promising as a way to organize the work in this area. Examining the table shows that a vast amount of research has focused on finding ways to evaluate ontologies based on their quality and suitability for a particular task, and that progress is being made on doing so. However, problems related to ontology evaluation remain. For example, the problem of selecting a domain ontology for a specific task requires a more thorough evaluation of the ontology than many of the research efforts have been able to achieve. This includes both

Table 6. Summary of Five Classes for Ontology Evaluation

Class	Examples	Advantages	Challenges
Error Checking	Gherasim et al. (2012) Guarino and Welty (2002) Keet et al. (2013) Poveda-Villalón et al. (2014) Schober et al. (2012)	- Possible to automate removal of many types of errors - Error removal easy to apply if requirements are stringent	- Often difficult to identify each error's level of urgency -Even after error removal, no surety that the ontology is high quality or is suitable for a particular task
Libraries	D'Aquin and Noy (2012) Schwartz (2014) Whetzel et al. (2011) Ziemba et al. (2015)	- Provides specific domain knowledge the community members possess - Provides additional functionality to the ontologies such as mappings, ontology reviews, and documentation - Recommender systems and matching services provided by some domain-specific libraries are increasing likelihood of finding appropriate ontology	- Many libraries allow uploading of ontologies to their repository in any readable ontology language, causing a lack of consistency within the library - Few general-purpose ontology libraries so not all ontologies have a place to be stored
Metric Based	Albarrak and Sibley (2012) Alm et al. (2013) Amirhosseini and Salim (2011) Bansal and Chawla (2015) Batet and Sánchez (2014) Bera et al. (2014) Choukri (2014) Duque-Ramos et al. (2014) de Villiers et al. (2017) Hicks (2017) McDaniel et al. (2018) Zhu et al. (2017)	- Metric-based techniques quantify the quality of a particular attribute - Many of the calculations can be automated so computers can calculate and compare ontology quality assessment scores	- Proving that the metrics applied are valid and appropriate requires repeated, empirical studies - Proving that the attributes assessed are correct for an ontology's intended use
Modularity	Ensan and Du (2013) Grüninger et al. (2012) Grüninger et al. (2012) Kahn and Keet (2016) Kutz and Hois (2012)	- Individual modules can be more specialized - Modules can be used to form larger ontologies containing only relevant concepts - Modules can be created from modules in which the quality has already been determined	- Very act of extracting modules from larger ontologies may inadvertently reduce their quality
Domain/Task Fit	Al-Khalifa et al. (2012) Ashraf et al. (2018) Bandeira et al. (2016) Hlomani (2014) Hobbs et al. (2014) Hoehndorf et al. (2012) Ismail et al. (2014) Krogstie (2012) Liu et al. (2017) Ouyang et al. (2012) Pittet and Barthélémy (2015) Rico et al. (2014) Romano and McDonald (2011)	- Allows the ontology to be evaluated on the specific qualities that are needed for a given domain or task	- Fitness measurements are not easily obtained - Concept matching is oversimplification of the complex nature of matching a domain to a particular ontology

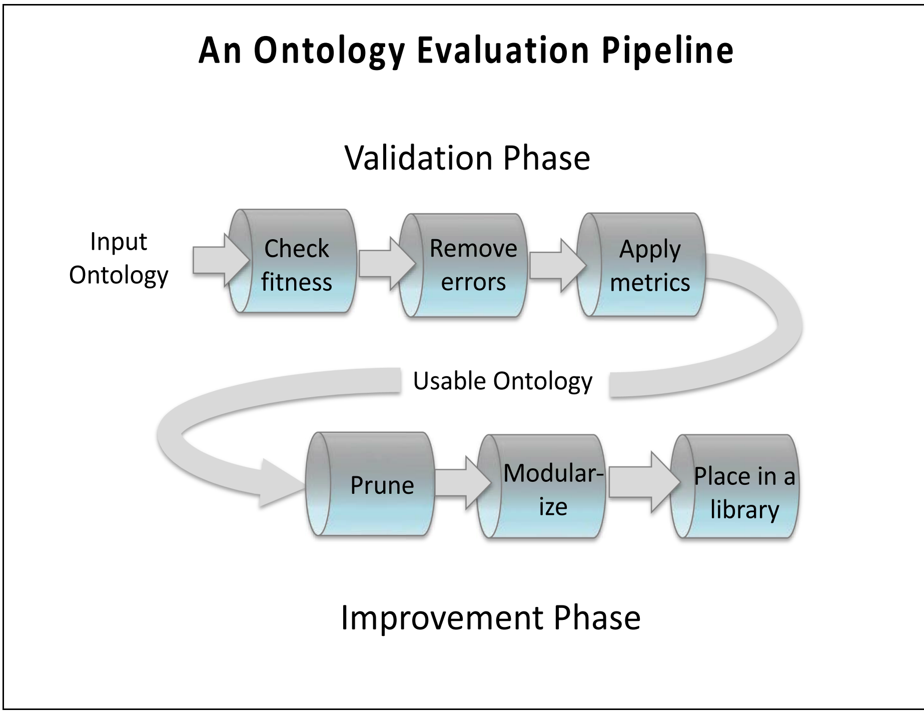

Fig. 8. An ontology evaluation pipeline: Validation and improvement.

an overall evaluation and a specific evaluation dependent on task and domain requirements. An attempt to address this problem by combining research categories is outlined in the following section.

4 POTENTIAL SOLUTION TO ONTOLOGY SELECTION PROBLEM

Selecting the best ontology to suit a specific need requires that the ontology be free from errors, modular in nature, and stored in an ontology repository where it can easily be found. It also requires that the ontology scores high on specific assessment attributes, aligns well with the required domain, and fits the task for which it is needed. However, this requires existing research efforts to be combined.

The combination of work to be carried out, we propose, can be represented by the sequential Ontology Evaluation Pipeline shown in Figure 8. Here, the ontology is the input, and the resulting output is an appropriate ontology for a given purpose. This framework, which is intended to serve as a guideline for researchers and practitioners, envisions ontology selection in two distinct phases. The first is to ascertain that an ontology is of high enough quality to be used and is the correct ontology for a specific task or domain. The second phase further improves the ontology by pruning it, modularizing it, and placing it in an appropriate ontology library where it can be found and repurposed.

4.1 Ontology Validation Phase

Ontologies can be problematic for a variety of reasons. They might have been automatically created by software and are incomplete, or perhaps created by novice ontology engineers without following a standard process. Yet, they may still contain valuable information. One of the first steps in ontology evaluation is checking its fitness for a given task and a specific domain. Another important step in evaluation is to apply ontology cleaning practices to remove errors as, for example, found in the work of Poveda-Villalón et al. (2014) and Schober et al. (2012). Finally, metrics should be applied to assess how high the quality of the metric is on certain levels applicable to ontology quality.

- **Check Domain and Task Fitness:** Attempts to ascertain fitness for a specific task include matching ontologies to goals (Bandeira et al. 2017) or to listing requirements made by stakeholders (Tatarintseva et al. 2013). Using a well-chosen method for fitness matching would help the user to select the most appropriate ontology to reuse from those available. Not only should an ontology match task requirements but also it should fit a particular area or domain that is neither too broad nor too narrow (Suárez-Figueroa et al. 2012). There needs to be a way to identify which ontology is most aligned with a specific domain. One possible way to do so is by comparing each of the ontologies to a corpus, such as that found in the work of Achichi et al. (2016).
- **Remove Errors:** Cleaning is required to make the ontology usable. Cleaning processes include removal of syntactical errors (Poveda-Villalón et al. 2014), checking for inconsistencies (Fahad and Qadir 2008), and reducing the number of redundancies (Rico et al. 2014).
- **Apply Metrics:** Once the selection has been narrowed down to a smaller list of ontologies, each of which closely matches the domain and task, quality metrics should be applied to identify which of the ontologies is the highest quality based on user-selected criteria. For this purpose, a ranking system such as that found in the work of Baliyan and Kumar (2016), Duque-Ramos et al. (2011), or McDaniel et al. (2018) could be applied.

The order in which these tasks are performed can vary depending on the input ontology. For example, if there is a problem reading the ontology, then the error checking and cleaning phase will need to be completed before the task and domain fitness can be assessed.

4.2 Ontology Improvement Phase

After the validation phase, the user is provided with a set of quality ontologies that are usable, but it might be possible to further improve their use.

- **Pruning and Modularization:** Large ontologies that attempt to cover specific terminology for multiple domains are difficult to reuse. An application might need only a small portion of an ontology for a specific purpose. A feasible approach could be to divide an ontology into irreducible modules using a procedure such as the framework developed by Grüninger et al. (2012).
- **Library Placement:** The final task in this phase is to create or find an appropriate, well-maintained ontology repository in which to place the ontology so it is easily accessible (and sharable) for future use. Many ontology libraries and linked datasets exist that not only are appropriately curated but also have searching capabilities (Silva et al. 2013; Katsumi and Grüninger 2017; Martínez-Romero et al. 2017; Vandenbussche et al. 2017).

Thus, the ontology improvement phase of the assessment pipeline should prune and modularize an ontology, if possible, before placing it in a library to improve the ability for the ontology to be reused.

4.3 Application of the Ontology Evaluation Pipeline

The two phases of the ontology evaluation process (ontology validation and ontology enhancement), as well as the ontology library that stores the ontologies, form a sequential pipeline.

The pipeline is one way to ensure the quality of an ontology and to improve the availability of ontologies suitable for reuse. Using a portion of the pipeline or combining two or more of the approaches in a similar framework could aid in solving other problems related to ontology evaluation. Although one approach to ontology evaluation may not be sufficient for a complete assessment, combining two or more methods could improve the accuracy of the results. The framework is thus intended to provide a way to consider elements from the five ontology evaluation classes from a problem-solving perspective.

5 CHALLENGES AND FUTURE WORK

A systematic method for ontology evaluation has long been needed (Cheatham and Pesquita. 2017; Degbelo 2017; Obrst et al. 2007) because ontology quality varies in terms of coverage, consistency, and intelligibility. Such a method should account for the fact that ontologies may be represented in different ways, and the tasks for which the ontologies are created may differ greatly. Yet, despite these differences, a high standard of quality must emerge based on tested theories, standard units of measure, and well-defined engineering practices. However, high-level assessment remains difficult because no single approach to ontology evaluation is applicable to all ontologies. There are many other challenges, the most notable of which are discussed below.

5.1 Interoperability between Computer Systems

An ontology is an enabling methodology for knowledge exchange (Gruber, 1993, 1995), with many research initiatives attempting to enrich available information with machine processing of semantic information (Guarino and Musen 2015). Semantics, in general, is a challenging problem for computer science (Bera et al. 2014; Ma et al. 2014; Sowa 2010; Purao and Storey 2005; Storey and Thalheim 2017). It deals with formal logic and truth, specifically, the truth of the representations of signs (Sowa 2015). The World Wide Web is so overburdened with formatting tags that there is no room left for representing semantics for the thousands, and possibly even millions, of competing vocabularies that exist (Sowa 2006). Adopting ontologies as a potential or partial solution to capturing semantic information requires that such ontologies be of high quality, hence the need for ontology quality assessment.

Interoperability, the ability of systems to communicate with each other, both syntactically and semantically, remains a critical need for communication among programs (Fritzsche et al. 2017). Because of the heterogeneous nature of software systems, in which data may be stored in different formats, knowledge sharing and communication are difficult to achieve (Kataria et al. 2008). A high-quality ontology can help with information sharing, and reusability and intercommunication between heterogeneous software systems, by providing a common vocabulary (Yao et al. 2009). Health care systems, for example, may have many different departments, locations, and types of computer systems, all of which need the ability to share patient information (Riano et al. 2012). Semantic interoperability is needed to efficiently provide information and services to an end-user, at any time and at any location (Yao et al. 2009). This requires both understanding and operationalization of semantics.

For interconnected biological repositories, with their vast amount of scientific data, system interoperability is critical but difficult to attain (Hoehndorf et al. 2012). The information stored in these databases comes from a variety of reference sources, communities, and organizations, making both interoperability and reusability challenging. The ability to easily and inexpensively evaluate the quality of these sources, however, could help address interoperability and reusability problems (Stvilia 2007; Lee and Stvilia 2017). The need to do so must to be recognized by other types of engineers when developing domain-specific systems such as components of the Internet of Things (Perera et al. 2014; Underwood et al. 2015), signal processing (Gemmeke et al. 2017), and bioinformatics (Zhou et al. 2017). Ontology engineers should also continue to share and update experiences with ontology management and assessment.

5.2 Semantic Complexity

As access to information and computer resources continues to expand, computer systems are expected to be able to seamlessly communicate. Well-designed and error-free ontologies are key to semantic integration and interoperability between these systems (Fritzsche et al. 2017). Developing ontologies, however, remains difficult because ontology development deals with capturing and representing stocks of knowledge about the real world. These stocks of knowledge, defined as accumulated knowledge assets (DeCarolis and Deeds, 1999), may be represented in a variety of forms. The entire field of knowledge representation recognizes such difficulties (Bimson and Hull 2016; Guarino and Guizzardi 2016; Sowa 2014, 2018).

Other challenges arise when tools are developed and used to automatically populate ontologies from a variety of sources, because such automatically generated ontologies may contain inconsistencies and redundancies (Brank et al. 2007; Park et al. 2007). Furthermore, many ontologies are used successfully, even though they may lack consistency or coverage, which should be included in the ultimate evaluation of an ontology but is difficult to assess (Obrst et al. 2007).

5.3 Upper Ontologies

Ontological evaluation can consider whether an ontology has been mapped to a large foundational (Guizzardi et al. 2008), or upper level, ontology (e.g., CYC (Lenat 2005; Matuszek et al. 2006), SUMO (Niles and Pease 2001)) to define its broad, general terms and axioms (Willner et al. 2015). If so, the upper-level ontology and the mappings between it and the newly developed ontology must also be evaluated (Obrst et al. 2007; Guarino 2017). In spite of this, the use of upper-level ontologies to assist in the development of new ontologies can be extremely beneficial because an upper-level ontology, such as Dolce (Gangemi et al. 2002), has already defined the basic categories and relationships and decisions have been made on how to represent reality (Neuhaus et al. 2013).

5.4 Availability of Ontologies

Practical concerns center around knowledge management. Ontology engineers and scientists must be able to locate existing ontologies instead of developing them from scratch (d'Aquin and Noy 2012) and understand the importance of evaluation (Neuhaus et al. 2013). Well-curated ontology libraries are needed that contain not only ontologies but also information about evaluations that have been performed on the ontologies in the libraries. In this way, ontology metadata can help guide the selection of an ontology for reuse from the supply of available ontologies (Obrst et al. 2014).

5.5 Progression of Ontology Field

Applied ontology, as interdisciplinary research on ontological analysis and conceptual modeling, may be moving toward a new science (Guarino and Musen 2015). From a high-level perspective, there are challenges to building complex information systems that will benefit our society with the evaluation of ontologies as part of a "larger endeavor to systematize the construction of information systems" (Obrst et al. 2007, p. 18).

Becoming a science, however, requires recognition that a genuine new body of knowledge is emerging with its own set of distinctive problems, which require the development of tools and methodologies to solve them (Shapere 1984). As with any science, effective procedures must be developed for measurement to become credible. A science has well-defined problems, resources to deal with problems (theories, models, and artifacts, evaluated for their effectiveness), and a recognizable (identifiable) domain. How well the domain is delineated depends on the development of the science (Kitcher 1995). If we consider the domain to be that of ontology development, formal theories, methodologies, and experiences are progressing (e.g., Ma et al. (2014); Bera et al. (2014); Jabar et al. (2014); Guizzardi et al. (2015); Musen (2015)). These efforts address theoretical challenges and contribute to shifts in research approaches to ones that might have previously been neglected. Our review of domain ontology evaluation metrics reveals areas in which more effort is required, before a standard, concrete, set of assessment metrics emerges.

6 CONCLUSION

Ontologies are increasingly important as intelligent applications continue to be developed and used, making the need for accurate communication between applications essential. Although significant work has been carried out on creating ontologies, it is important to be able to assess the quality of these ontologies in a systematic way. Doing so can help developers to select an ontology from among available choices or to create their own for domain and task-ontology fit. Automated tools, which create ontologies from existing systems, especially need a mechanism to assess whether the ontologies being created are correct, meaningful, and useful.

This article has reviewed work on ontology evaluation and clarified the issues involved in an attempt to identify challenges that still need to be overcome as ontologies continue to be integrated into information system applications that depend on domain knowledge. Five categories were derived to classify existing ontology evaluation approaches, based on a set of approximately 170 research articles that focus specifically on evaluating domain ontologies. The classification was applied to clarify the work that has been carried out as well as to identify where additional work is needed. A pipeline approach to evaluation was then proposed. Challenges and future research were discussed that deal with issues related to interoperability, progression of the Semantic Web, and other applications.

APPENDIX A: RESEARCH STUDIES USED FOR THE DEVELOPMENT OF THE ONTOLOGY RESEARCH CLASSIFICATION SCHEME

Key: **D**: Domain/Task Fit; **E**: Error Checking; **M**: Metrics; **O**: Modularization; **L**: Libraries	
D	Hend S. Al-Khalifa, Maha M. Al-Yahya, Alia Bahanshal, and Iman Al-Odah. 2012. On the evaluation of linguistic ontological models: An application on the SemQ ontology. In *7th International Conference on Digital Information Management (ICDIM'12)*. 341–345. Keywords: semantics, linguistics, feature extraction, user context, user content, requirements
D	Mauricio Barcellos Almeida. 2009. A proposal to evaluate ontology content. 2009. *Applied Ontology* 4, 3 (2009), 245–265. Keywords: ontologies, modeling, user-centered evaluation, task-fitness, content, context
D	Jamshaid Ashraf, Omar K. Hussain, Farookh Khadeer Hussain, and Elizabeth J. Chang. 2018. Evaluation of U ontology. In *Measuring and Analysing the Use of Ontologies*. Springer, Cham, 243–268. Keywords: task-specific evaluation, application specific evaluation, context-dependent evaluation, gold standard evaluation, alignment
D	Judson Bandeira, Ig Ibert Bittencourt, Patricia Espinheira, and Seiji Isotani. 2016. FOCA: A Methodology for Ontology Evaluation. *arXiv preprint arXiv:1612.03353*. Keywords: ontology evaluation, goal, question, methodology
D	Janez Brank, Dunja Madenic, and Marko Groblenik. 2006. Gold standard based ontology evaluation using instance assignment. In *Proceedings of the 4th Workshop on Evaluating Ontologies for the Web (EON'06)*. Keywords: ontology instances, evaluation, task, matching, gold standard ontologies
D	Robert Porzel and Rainer Malaka. 2005. A task-based framework for ontology learning, population and evaluation. *Ontology Learning from Text: Methods, Evaluation and Applications*, Vol. 123. 107–122. Keywords: framework, ontology learning, ontology evaluation, text-based learning, assessment

D	Núria Casellas. 2009. Ontology evaluation through usability measures. In *On the Move to Meaningful Internet Systems OTM 2009 Workshops*. Springer, Berlin/Heidelberg, 594–603. Keywords: ontology evaluation, usability, user study
D	Erik L. Clarke, Salvatore Loguercio, Benjamin M. Good, and Andrew I. Su. 2013. A task-based approach for gene ontology evaluation. *Journal of Biomedical Semantics* 4, 1 (2013). Keywords: task fit, ontology matching, ontology evaluation, gene ontologies
D	Michael Cochez and Vagan Y. Terziyan. 2012. Quality of an ontology as a dynamic optimisation problem. In *ICTERI*. 249–256. Keywords: Semantic Web, Ontology features, Ontology quality, Contextual optimization, user context
D	de Almeida Falbo, Monalessa Perini Barcellos, Julio Cesar Nardi, and Giancarlo Guizzardi. 2013. Organizing ontology design patterns as ontology pattern languages. In *Extended Semantic Web Conference*. Springer, Berlin, Heidelberg, 61–75. Keywords: ontology design patterns, pattern matching, ontology pattern language, ontology evaluation
D	Muhammad Fahad, Muhammad Abdul Qadir, and Syed Adnan Hussain Shah. 2008. Evaluation of ontologies and DL reasoners. In *International Conference on Intelligent Information Processing*. Springer, Boston, MA, 17–27. Keywords: ontology evaluation, DL Reasoners, information, knowledge, logic
D	Miriam Fernández, Chwhynny Overbeeke, Marta Sabou, and Enrico Motta. 2009. What makes a good ontology? A case-study in fine-grained knowledge reuse. In *Asian Semantic Web Conference*. Springer, Berlin, Heidelberg, 61–75.
D	Miriam Fernández, Iván Cantador, and Pablo Castells. 2006. CORE: A tool for collaborative ontology reuse and evaluation. In *4th International Conference on Evaluation of Ontologies for the Web (EON'06)*. Edinburgh, UK. Keywords: Ontology reuse, ontology quality, collaboration, task-based
D	Aldo Gangemi, Carola Catenacci, Massimiliano Ciaramita, and Jos Lehmann. 2006. Qood grid: A metaontology-based framework for ontology evaluation and selection. In *Proceedings of the EON 2006 Workshop*. Keywords: ontology evaluation, ontology selection, task fitness, parameterized framework
D	Aldo Gangemi, Carola Catenacci, Massimiliano Ciaramita, and Jos Lehmann. 2005. *Ontology Evaluation and Validation: An Integrated Formal Model for the Quality Diagnostic task*. Technical report. Laboratory of Applied Ontologies, CNR, Rome, Italy. http://www.loa-cnr.it/Files/OntoEval4OntoDev_Final.pdf. Keywords: ontology validation, ontology assessment, framework, model, evaluation
D	Gintare Grigonyte. 2010. *Building and Evaluating Domain Ontologies: NLP Contributions*. Logos Verlag, Berlin, GmbH. Keywords: ontology matching, ontology evaluation, ontology quality
D	Giancarlo Guizzardi, Ricardo de Almeida Falbo, and Renata S. S. Guizzardi. 2008. Grounding software domain ontologies in the unified foundational ontology (UFO): The case of the ODE software process ontology. In *CIbSE*. 127–140. Keywords: upper ontology, matching, semantic model
D	Jon Atle Gulla, Darijus Strasunskas, and Stein L. Tomassen. 2006. Semantic Interoperability in Multi-Disciplinary Domain. Applications in Petroleum Industry. *Contexts and Ontologies: Theory, Practice and Applications* (2006), 30. Keywords: ontology standards, ontology content, user context
D	Peter Haase and York Sure. 2005. Incremental Ontology Evolution-Evaluation. Institute AIFB, University of Karlsruhe. Sekt deliverable d3, 1. Keywords: Semantic Web, Ontology quality, quality determination, user context
D	Hlomani, Hlomani. 2014. *Multidimensional Data-driven Ontology Evaluation*. PhD Dissertation. Keywords: ontology, evaluation, assessment, data, task, ontology selection
D	Mike Hobbs, Cristina Luca, Arooj Fatima, and Mark Warnes. 2014. Ontological analysis for dynamic data model exploration. *Electronic Journal of Applied Statistical Analysis: Decision Support Systems and Services Evaluation* 5, 1 (2014), 42–56. Keywords: Ontology, model, classifier, concept mapping
D	Robert Hoehndorf, Michel Dumontier, and Georgios V. Gkoutos. 2012. Evaluation of research in biomedical ontologies. *Briefings in Bioinformatics* 14, 6 (2012), 696–712. Keywords: biomedical ontology, ontology evaluation, task-based evaluation, evaluation criteria
D	Khairul Nurmazianna Ismail and Zainab Abu Bakar. 2015. DuriO concept evaluation. In *IEEE Conference on e-Learning, e-Management and e-Services (IC3e'15)*. 116–119. Keywords: ontology enrichment, evaluation, artificial intelligence, ontology application
D	Xu Jianliang and Ma Xiaowei. 2009. A web-based ontology evaluation system. In *International Conference on Advanced Language Processing and Web Information Technology (ALPIT'08)*. IEEE. Keywords: ontology evaluation, context driven, evaluation tool, expert evaluation, framework, user driven

D	Prodromos Kolyvakis, Alexandros Kalousis, Barry Smith, and Dimitris Kiritsis. 2008. Biomedical ontology alignment: an approach based on representation learning. *Journal of Biomedical Semantics* 9, 1 (2018), 21. Keywords: Semantic similarity, Sentence embeddingg, Word embeddings, Denoising autoencoder, Outlier detection, domain, task, suitability
D	John Krogstie. 2012. Model-based development and evolution of information systems: A quality approach. Springer Science & Business Media. Keywords: ontology quality, modeling, user centered evaluation, context-centered
D	Holger Lewen, Kaustubh Supekar, Natalya F. Noy, and Mark A. Musen. 2006. Topic-specific trust and open rating systems: An approach for ontology evaluation. In *Workshop on Evaluation of Ontologies for the Web*. Keywords: Ontology reuse, ontology quality, task-based criteria, application-centered
D	Ning Li, Enrico Motta, and Mathieu d'Aquin. Ontology summarization: an analysis and an evaluation. 2010. Keywords: ontology, evaluation, summarization, modeling, pruning
D	Danfeng Liu, Antonis Bikakis, and Andreas Vlachidis. 2017. Evaluation of semantic web ontologies for modelling art collections. In *Advances in Databases and Information Systems*. Springer, Cham. Keywords: task-based, domain, art collections, ontology, quality
D	Gilbert Maiga and D. Williams. 2008. A systems approach to user evaluation of biomedical ontologies. Keywords: user testing, ontology evaluation, biomedical, bio-ontologies
D	Imen Bouaziz Mezghanni and Faiez Gargouri. 2017. A gold standard-based approach for arabic ontology evaluation. In *European Conference on Knowledge Management*. Academic Conferences International Limited, 1153–1161. Keywords: legal ontology, Arabic legal domain, gold standard
D	Nandana Mihindukulasooriya, María Poveda-Villalón, Raúl García-Castro, and Asunción Gómez-Pérez. 2016. Collaborative ontology evolution and data quality-an empirical analysis. In *International Experiences and Directions Workshop on OWL*. Springer, Cham, 95–114. Keywords: ontology development, ontology assessment, user centered, data quality
D	L. Ouyang, B. Zou, and J. Lin. 2012. BDHL: A framework of index system for evaluating ontology content. *International Review on Computers and Software* 7, 2 (2012). Keywords: ontology, framework, content, user context
D	Ted Pedersen, Serguei V. S. Pakhomov, Siddharth Patwardhan, and Christopher G. Chute. 2007. Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics* 40, 3 (2007), 288–299. Keywords: ontology matching, semantic similarity, semantic relatedness
D	Helena Sofia Pinto and João P. Martins. 2000. Reusing ontologies. In *AAAI 2000 Spring Symposium on Bringing Knowledge to Business Processes*. AAAI Press, 77–84. Keywords: ontology application, knowledge evaluation, knowledge modeling, business processes
D	Perrine Pittet and Jérôme Barthélémy. 2015. Exploiting users' feedbacks: Towards a task-based evaluation of application ontologies throughout their lifecycle. In *International Conference on Knowledge Engineering and Ontology Development*, Vol. 2. Keywords: Application Ontology, Task-based Ontology Evaluation, Ontology Revision, Semantic Annotation, Ontology Lifecycle, Crowdsourcing
D	Natalia Ponomareva. 2008. Towards an Optimal Ontology Construction. Doctoral Dissertation. Polytechnic University of Valencia Valencia, Spain. Keywords: ontology construction, ontology development, ontology content, user specifications
D	Mariela Rico, María Laura Caliusco, Omar Chiotti, and María Rosa Galli. 2014. OntoQualitas: A framework for ontology quality assessment in information interchanges between heterogeneous systems. *Computers in Industry* 65, 9 (2014), 1291–1300. Keywords: ontology evaluation, information interchange, ontology requirements
D	Rosetta Romano and Craig McDonald. 2011. Assessing the Quality of Ontology. In *Proceedings of MCIS 2011* (2011), 1–11. Keywords: Semantic Web, Ontology quality, quality determination, user context
D	Delia Rusu, Blaž Fortuna, and Dunja Mladenić. 2014. Measuring concept similarity in ontologies using weighted concept paths. *Applied Ontology* 9, 1 (2014), 65-95. Keywords: ontology matching, concept matching, semantic similarity, ontologies
D	Marta Sabouand Miriam Fernández. 2012. Ontology (network) evaluation. In *Ontology Engineering in a Networked World* (2012), 193–212. Keywords: ontology evaluation, reference ontology, network
D	Marta Sabou, Vanessa Lopez, Enrico Motta, and Victoria Uren. 2006. *Ontology Selection: Ontology Evaluation on the Real Semantic Web*. Keywords: ontology selection, ontology evaluation, semantic web, requirements oriented selection
D	David Sánchez, Montserrat Batet, David Isern, and Aida Valls. Ontology-based semantic similarity: A new feature-based approach. *Expert Systems with Applications* 39, 9 (2012), 7718–7728. Keywords: ontology matching, ontology similarity, semantic relatedness, WordNet

D	Syed Adnan Hussain Shah, Adnan Khalid, and Muhammad Abdul Qadir. 2008. OntoFetcher: An approach for query generation to gather ontologies and ranking them by ensuring user's context. In *4th International Conference on Emerging Technologies (ICET'08)*. 247–252. Keywords: ontology searching, ontology ranking, user criteria, user context
D	Peter Spyns. 2010. Assessing iterations of an automated ontology evaluation procedure. *On the Move to Meaningful Internet Systems (OTM'10)*. 1145–1159. Keywords: ontology evaluation, user context, ontology ranking
D	Mark Stevenson. 2002. Combining disambiguation techniques to enrich an ontology. In *Proceedings of the Fifteenth European Conference on Artificial Intelligence (ECAI'02), Workshop on Machine Learning and Natural Language Processing for Ontology Engineering*. Keywords: ontology enrichment, evaluation, artificial intelligence, ontology application
D	Darijus Strasunskas and Stein L. Tomassen. 2008. Empirical insights on a value of ontology quality in ontology-driven web search. In *OTM Confederated International Conferences On the Move to Meaningful Internet Systems*. Springer, 1319–1337. Keywords: ontology quality evaluation, task-driven evaluation
D	Darijus Strasunskas and Stein L. Tomassen. 2008. On significance of ontology quality in ontology-driven web search. *World Summit on Knowledge Society*. 469–478. Keywords: ontology evaluation, task fitness, context driven, domain specific
D	Darijus Strasunskas and Stein L. Tomassen. 2007. Quality aspects in ontology-based information retrieval. In *IRMA International Conference*. 1048–1105. Keywords: ontology evaluation, information quality, user context, domain content
D	Darijus Strasunskas and Stein L. Tomassen. The role of ontology in enhancing semantic searches: the EvOQS framework and its initial validation. *International Journal of Knowledge and Learning* 4, 4 (2008), 398–414. Keywords: Ontology application, ontology quality, quality framework, information retrieval, semantic search, task-driven context
D	Darijus Strasunskasand and Stein Tomassen. 2007. Web search tailored ontology evaluation framework. In *Advances in Web and Network Technologies, and Information Management*. 372–383. Keywords: ontology, ontology evaluation, task fitness, search engine
D	Vijayan Sugumaran and Veda C. Storey. 2002. Ontologies for conceptual modeling: their creation, use, and management. *Data & Knowledge Engineering* 42, 3 (2002), 251–271. Keywords: ontology evaluation, life cycle, domain, ontology task, ontology fitness
D	Kaustubh Supekar. 2005. A peer-review approach for ontology evaluation. In *8th International Protege Conference*. 77–79. Keywords: ontology evaluation, ontology selection, user-centered, requirements-oriented
D	Jonathan Yu, James A. Thom, and Audrey Tam. 2009. Requirements-oriented methodology for evaluating ontologies. *Information Systems* 34, 8 (2009), 766–791. Keywords: ontology evaluation, requirements, user evaluation, ontology fitness
D	Elias Zavitsanos, George Paliouras, and George A. Vouros. 2011. Gold standard evaluation of ontology learning methods through ontology transformation and alignment. *IEEE Transactions on Knowledge and Data Engineering* 23, 11 (2011), 1635–1648. Keywords: ontology matching, ontology learning, ontology alignment, gold standard ontology
E	Ismailcem Budak Arpinar, Karthikeyan Giriloganathan, and Boanerges Aleman-Meza. 2006. Ontology quality by detection of conflicts in metadata. In *Proceedings of the 4th International EON Workshop*. Keywords: Semantic Web, data conflict, Rule, Ontology Quality, Ontology Evaluation.
E	Joachim Baumeister and Dietmar Seipel. 2005. Smelly owls-design anomalies in ontologies. In *FLAIRS Conference*, Vol. 215. 220. Keywords: ontology cleaning, errors, anomalies, ontology evaluation
E	Olivier Bodenreider. 2016. Identifying missing hierarchical relations in SNOMED CT from logical definitions based on the lexical features of concept names. ICBO/BioCreative. Keywords: description logics; SNOMED CT, quality assurance, lexical features
E	Muhammad Fahad and Muhammad Abdul Qadir. 2008. A framework for ontology evaluation. *ICCS Supplement* 354 (2008), 149–158. Keywords: Ontological Errors, Ontology Verification, Ontology Design Anomalies, Ontology Mapping and Merging, Semantic Web
E	Muhammad Fahad, Muhammad Abdul Qadir, and Muhammad Wajahaat Noshairwan. 2008. Ontological errors-inconsistency, incompleteness and redundancy. In *ICEIS* (3-2). 253–285. Keywords: Ontology reuse, ontology quality, quality criteria, metrics
E	Muhammad Fahad, Muhammad Abdul Qadir, and Muhammad Wajahaat Noshairwan. 2007. Semantic inconsistency errors in ontology. In *2007 IEEE International Conference on Granular Computing (GRC'07)*. 283–283. Keywords: errors, ontology cleaning, consistency, inconsistency, knowledge base, semantics

E	Toader Gherasim, Giuseppe Berio, Mounira Harzallah, and Pascale Kuntz. 2012. Problems impacting the quality of automatically built ontologies. *Knowledge Engineering and Software Engineering (KESE8'12)*. 22. Keywords: ontology extraction, ontology evaluation, error checking, pitfalls, anomalies in ontologies
E	Nicola Guarino and Christopher Welty. 2002. Evaluating ontological decisions with Onto Clean. *Communications of the ACM* 45, 2 (2002), 61–65. Keywords: ontology cleaning, ontology evaluation, error checking, error removal
E	C. Maria Keet, Mari Carmen Sunarez-Figueroa, and Maria Poveda-Villalon. 2013. The current landscape of pitfalls in ontologies. In *Proceedings of the 5th International Conference on Knowledge Engineering and Ontology Development*. Keywords: ontology errors, pitfalls, ontology cleaning, evaluation
E	Sandra Lovrencic and Mirko Cubrilo. 2008. Ontology evaluation-comprising verification and validation. In *CECIIS-2008*. Keywords: ontology checking, validation, error checking
E	Wajahat Noshairwan, Muhammad Abdul Qadir, and Muhammad Fahad. 2007. Sufficient knowledge omission error and redundant disjoint relation in ontology. In *Advances in Intelligent Web Mastering*. Springer, Berlin, Heidelberg, 260–265. Keywords: world wide web, ontology relations, errors, redundancy, evaluation
E	Alessandro Oltramari, Aldo Gangemi, Nicola Guarino, and Claudio Masolo. 2002. Restructuring WordNet's top-level: The OntoClean approach. In *Workshop Proceedings of OntoLex*, Vol. 2. 17–26. Keywords: ontology restructuring, error removal, ontology cleaning, ontoclean
E	Anthony M. Orme, Haining Yao, and Letha H. Etzkorn. 2007. Indicating ontology data quality, stability, and completeness throughout ontology evolution. *Journal of Software: Evolution and Process* 19, 1 (2007), 49–75. Keywords: data quality, ontology assessment, completeness, consistency, ontology evolution
E	Maria Poveda-Villalón, Asunción Gómez-Pérez, and Mari Carmen Suárez-Figueroa. 2014. Oops! (ontology pitfall scanner!): An on-line tool for ontology evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS)* 10, 2 (2014), 7–34. Keywords: pitfalls, bad practices, ontology evaluation, ontology engineering
E	Maria Poveda-Villalón, Mari Carmen Suárez-Figueroa, and Asunción Gómez-Pérez. 2012. Did you validate your ontology? OOPS! In *Extended Semantic Web Conference*. Springer, Berlin, Heidelberg, 402–407. Keywords: pitfalls, bad practices, ontology evaluation, ontology engineering
E	Muhammad Abdul Qadir and Wajahat Noshairwan. 2007. Warnings for Disjoint Knowledge Omission in Ontologies. In *Second International Conference on Internet and Web Applications and Services (ICIW'07)*. IEEE, 45–45. Keywords: ontology reuse, errors, coverage, semantics, assessment
E	Muhammad Abdul Qadir, Muhammad Fahad, and Syed Adnan Hussain Shah. 2007. Incompleteness errors in ontology. In *IEEE International Conference on Granular Computing (GRC'07)*. IEEE, 279–279. Keywords: ontology errors, ontology cleaning, incompleteness, comprehensiveness
E	Daniel Schober, Ilinca Tudose, Vojtech Svatek, and Martin Boeker. 2012. OntoCheck: verifying ontology naming conventions and metadata completeness in Protégé 4. *Journal of Biomedical Semantics* 3, 2 (2012), S4. Keywords: ontology cleaning, error checking, completeness
E	Christopher Welty and William Andersen. 2005. Towards ontoclean 2.0: A framework for rigidity. *Applied Ontology* 1, 1 (2005), 107–116. Keywords: ontology evaluation, rigidity, framework, ontology restructuring, ontology cleaning
L	Kenneth Baclawski and Todd Schneider. 2009. The open ontology repository initiative: Requirements and research challenges. In *Proceedings of Workshop on Collaborative Construction, Management and Linking of Structured Knowledge at the International Semantic Web Conference (ISWC'09)*. 18. Keywords: ontology, repository, metadata, federated repositories, interoperability
L	Mathieu d'Aquin and Natalya F. Noy. 2012. Where to publish and find ontologies? A survey of ontology libraries. *Web Semantics: Science, Services and Agents on the World Wide Web 11*. 96–111. Keywords: ontology repository, ontology evaluation, content, gatekeeping, ontology matching
L	Immanuel Normann and Oliver Kutz. 2010. Ontology reuse and exploration via interactive graph manipulation. In *Proceedings of the International Semantic Web Conference (ISWC) Workshop on Ontology Repositories for the Web (SERES'10)*. Keywords: Ontology repositories, modularity, matching, information visualization
L	David G. Schwartz. 2014. Enhancing knowledge marketplaces through the theory of knowledge measurement. *Handbook of Strategic e-Business Management*. 735–748. Keywords: Knowledge repositories, Knowledge management, Information quality, Metrics Ontology Measurement, Semantics
L	Barry Smith. 2008. Ontology (science). In *Formal Ontology in Information Systems (FOIS'08)*. 21–35. Keywords: scientific method, expert peer review, ontology engineering, biomedical informatics, Gene Ontology, OBO Foundry, ontology library
L	Ontology Summit. 2008. Toward an open ontology repository: http://ontolog.cim3.net/cgibin/wiki.pl. Ontology Summit 2008. Keywords: ontology, ontology reuse, ontology repository

L	Patricia L. Whetzel, Natalya F. Noy, Nigam H. Shah, Paul R. Alexander, Csongor Nyulas, Tania Tudorache, and Mark A. Musen. 2011. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Research* 39, suppl_2 (2011), W541–W545. Keywords: ontology quality, ontology repository, biomedical ontology
L	Paweł Ziemba, Jarosław Jankowski, Jarosław Wątróbski, and Jarosław Becker. 2015. Knowledge management in website quality evaluation domain. In *Computational Collective Intelligence*. Springer, Cham, 75–85. Keywords: repository, ontology, evaluation, comparison, assessment, ontology evaluation, knowledge conceptualization, website quality
M	Harith Alani and Christopher Brewster. 2006. Metrics for ranking ontologies. In *Evaluating Ontologies for the Web Workshop (EON'06), 15th International World Wide Web Conference, 23-26 May 2006, Edinburgh, Scotland*. Keywords: ontology ranking, assessment, metrics, semantics, structure, syntax
M	Harith Alani, Christopher Brewster, and Nigel Shadbolt. 2006. Ranking ontologies with AKTiveRank. In *Proceedings of the 5th International Semantic Web Conference (ISWC'06), Athens, GA, USA, November 5-9, 2006*. 1–15. Keywords: ontology evaluation, structure analysis, structural metrics
M	Khalid Albarrak and Edgar Sibley. 2012. Measuring expressivity between ontology models. In *Proceedings of the 11th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED'12)*. Keywords: Knowledge representation, ontology evaluation, ontology quality measurement, ontology metrics
M	Rebekka Alm, Sven Kiehl, Birger Lantow, and Kurt Sandkuhl. 2013. Applicability of quality metrics for ontologies on ontology design patterns. In *Proceedings of the 5th International Conference on Knowledge Engineering and Ontology Development (KEOD'13), Vilamoura, Portugal*. 48–57. Keywords: ontology metrics, knowledge engineering, design patterns, attributes
M	Maziar Amirhosseini and Juhana Salim. 2011. Ontoabsolute as a ontology evaluation methodology in analysis of the structural domains in upper, middle and lower level ontologies. In *International Conference on Semantic Technology and Information Retrieval (STAIR'11)*. 26–33. Keywords: ontology evaluation, methodology, structural domains analysis, simplicity, semantic relations, metric criteria
M	Ritika Bansal and Sonal Chawla. 2015. Evaluation metrics for computer science domain specific ontology in semantic web based IRSCSD system. *International Journal of Computer (IJC)* 19, 1 (2015), 129–139. Keywords: Information retrieval; Ontology quality; Ontology evaluation; Evaluation metrics; Ontology verification; Ontology validation; Sparql queries
M	Montserrat Batet and David Sánchez. 2014. A semantic approach for ontology evaluation. In *IEEE 26th International Conference on Tools with Artificial Intelligence (ICTAI'14)*. 138–145. Keywords: Knowledge representation; Ontology evaluation; semantics; dispersion, metrics
M	Palash Bera, Andrew Burton-Jones, and Yair Wand. 2014. Research note—how semantics and pragmatics interact in understanding conceptual models. *Information Systems Research* 25, 2 (2014), 401–419. Keywords: ontology, ontology quality metric, semiotic theory, semantics, pragmatics
M	Andrew Burton-Jones, Veda C. Storey, Vijayan Sugumaran, and Punit Ahluwalia. 2005. A semiotic metrics suite for assessing the quality of ontologies. *Data & Knowledge Engineering* 55, 1 (2005), 84–102. Keywords: ontology, ontology quality metric, semiotic theory, ontology auditor
M	Andrew Burton-Jones, Veda C. Storey, Vijayan Sugumaran, and Punit Ahluwalia. 2003. Assessing the Effectiveness of the DAML Ontologies for the Semantic Web. In *Natural Language Processing and Information Systems, 8th International Conference on Applications of Natural Language to Information Systems (Spreewald)*. 56–69. Keywords: ontology, effectiveness, assessment, evaluation, semantics, pragmatics, metrics
M	Shyam R. Chidamber and Chris F. Kemerer. 1994. A metrics suite for object oriented design. *IEEE Transactions on Software Engineering* 20, 6 (1994), 476–493. Keywords: object oriented design, metrics suite, software development, ontologies
M	Djellali Choukri. 2014. A new distributed expert system to ontology evaluation. In *Proceedings of the 5th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN'14)*. Procedia Computer Science 37, 48–55. Keywords: Semantic Web, Ontology, logic, consistency, subsumption, instantiation, DIG, metrics
M	Valerie Cross and Anindita Pal. 2008. An ontology analysis tool. *International Journal of General Systems* 37, 1 (2008), 17–44. Keywords: ontology evaluation, content analysis, structure analysis, information content, metrics
M	Johan Pieter de Villiers, Richard W. Focke, Gregor Pavlin, Anne-Laure Jousselme, Valentina Dragos, Kathryn Blackmond Laskey, P. C. Costa, and Erik Blasch. 2017. Evaluation metrics for the practical application of URREF ontology: An illustration on data criteria. In *20th International Conference on Information Fusion (Fusion'17)*. 1–8. Keywords: ontology evaluation, data criteria, metrics, evaluation criteria, use cases
M	Astrid Duque-Ramos, Jesualdo Tomás Fernández-Breis, Robert Stevens, and Nathalie Aussenac-Gilles. 2011. OQuaRE: A SQuaRE-based approach for evaluating the quality of ontologies. *Journal of Research and Practice in Information Technology* 43, 2 (2011), 159–176. Keywords: Semantic Web, Ontology Quality, Software Quality, ISO 25000, metrics

M	Erki Eessaar. 2008. Towards a semiotic quality framework of software measures. In *Proceedings of the 10th International Conference on Enterprise Information Systems (ICEIS'08), June 12-16, 2008, Barcelona, Spain.* 41–48. Keywords: Metrics, Measures, Semiotics, Quality, Metamodel, Database design
M	Letha H. Etzkorn. 2006. Semantic metrics, conceptual metrics, and ontology metrics: an analysis of software quality using IR-based systems, potential applications and collaborations. In *Proceedings of the International Conference on Software Maintenance.* Keywords: semantics, conceptual metrics, ontology metrics, quality evaluation
M	Jesualdo Tomás Fernández-Breis, Mikel Egaña Aranguren, and Robert Stevens. 2009. A quality evaluation framework for bio-ontologies. In *Proceedings of the International Conference on Biomedical Ontology (ICBO'09), July 24-26, 2009, Buffalo, NY.* 127–130. Keywords: framework, bio-ontology, biomedical knowledge management, quality, metrics
M	Asunción Gómez-Pérez. 1994. Some Ideas and Examples to Evaluate Ontologies. Knowledge Systems Laboratory, Stanford University. Keywords: ontology assessment, metrics, evaluation, attributes, ontology criteria
M	Asunción Gomez-Perez and Adolfo Lozano-Tello. 2005. Applying the ontometric method to measure the suitability of ontologies. In *Business Systems Analysis with Ontologies.* 249–269. Keywords: ontology evaluation, ontology selection, systems analysis, systems requirements
M	Asunción Gómez-Pérez. 1999. Evaluation of Taxonomic Knowledge in Ontologies and Knowledge Bases. Keywords: ontology, knowledge base, taxonomy, evaluation, metrics, assessment, criteria
M	Asunción Gómez-Pérez. 1996. Towards a framework to verify knowledge sharing technology. *Expert Systems with Applications* 11, 4 (1996), 519–529. Keywords: ontology evaluation, evaluation criteria, framework, metrics
M	Michael Grüninger and Mark S. Fox. 1995. Methodology for the design and evaluation of ontologies. In *International Joint Conference on Artificial Intelligence (IJCAI'95), Workshop on Basic Ontological Issues in Knowledge Sharing, August 20-25, 1995 Montreal, Quebec, Canada.* Keywords: framework, evaluation, ontology design, ontology management, metrics
M	Nicola Guarino. 2004. Toward a formal evaluation of ontology quality. *IEEE Intelligent Systems* 19, 4 (2004), 78–79. Keywords: model, semantics, ontology assessment, metrics, attributes
M	Amanda Hicks. 2017. Metrics and methods for comparative ontology evaluation. *Ciência da Informação* 46, 1 (2017). Keywords: ontologies, ontology evaluation, metrics, methods, comparative metrics
M	Hlomani Hlomani, Mitchell G. Gillespie, Daniel Kotowski, and Deborah A. Stacey. 2011. Utilizing a compositional knowledge framework for ontology evaluation: A case study on BioSTORM. In *Conference on Knowledge Engineering and Ontology Development (KEOD'11), October 26-29, 2011, Paris, France.* 167–175. Keywords: knowledge engineerings, ontology, ontology evaluation, ontology development, framework, metrics
M	Marzanah A. Jabar, M. A. Khalefa, Mohd Syazwan Abdullah, and Rusli Abdullah. 2014. Meta-analysis of ontology software development process. *International Review of Computers and Software* 9, 1 (2014), 29–37. Keywords: ontology development, ontology evaluation, metrics
M	Clement Jonquet, Nigam H. Shah, and Mark A. Musen. 2009. Prototyping a biomedical ontology recommender service. *Bio-Ontologies: Knowledge in Biology, ISMB/ECCB SIG* (2009). Keywords: ontology recommendation, ontology assessment, ontology quality, metrics
M	Karim Kamoun and Sadok Ben Yahia. 2012. Information content similarity measure to assess stability during ontology enrichment. *International Review on Computers and Software* 7, 3. Keywords: ontology development, ontology evaluation, metrics
M	Karim Kamoun and Sadok Ben Yahia. 2012. A novel global measure approach based on ontology spectrum to evaluate ontology enrichment. *Complexity* 39, 17. Keywords: Ontology evaluation, semantic similarity measure, ontology enrichment, ontology quality, ontology stability, quality metrics
M	Dimitris N. Kanellopoulos. 2009. ODELO: an ontology-driven model for the evaluation of learning ontologies. *International Journal of Learning Technology* 4, 1–2 (2009), 73–99. Keywords: ontology evaluation, learning ontologies, modeling, pragmatics, metrics
M	Dionysios D. Kehagias, Ioannis Papadimitriou, Joana Hois, Dimitrios Tzovaras, and John Bateman. 2008. A methodological approach for ontology evaluation and refinement. In *the 2nd International Conference of ASK-IT. June 2008, Nuremberg, Germany.* Keywords: Ontologies, evaluation criteria, refinement process, metrics
M	John Krogstie, Odd Ivar Lindland, and Guttorm Sindre. 1995. Defining quality aspects for conceptual models. *Information System Concepts.* 216–231. Keywords: Quality, information system models, semiotics, metrics, assessment
M	Liubo Ouyang, Beiji Zou, Miaoxing Qu, and Chengming Zhang. 2011. A method of ontology evaluation based on coverage, cohesion and coupling. In *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'11)*, Vol. 4. 2451–2455. Keywords: coverage, cohesion, coupling, ontology metrics

M	Adolfo Lozano-Tello and Asunción Gómez-Pérez. 2004. Ontometric: A method to choose the appropriate ontology. *Journal of Database Management (JDM)* 15, 2 (2004), 1–18. Keywords: ontology evaluation, ontology task, ontology matching, ontology selection, metrics, criteria
M	Yinglong Ma, Haijiang Wu, Xinyu Ma, Beihong Jin, Tao Huang, and Jun Wei. 2011. Stable cohesion metrics for evolving ontologies. *Journal of Software: Evolution and Process* 23, 5 (2011), 343–359. Keywords: ontology evaluation, ontology metrics, cohesion
M	Yinglong Ma, Ling Liu, Ke Lu, Beihong Jin, and Xiangjie Liu. 2014. A graph derivation based approach for measuring and comparing structural semantics of ontologies. *IEEE Transactions on Knowledge and Data Engineering* 26, 5 (2014), 1039–1052. Keywords: ontologies, semantics, OWL, unified modeling language, graphical models, transforms, evaluation metrics
M	Liam Magee. 2010. A framework for assessing commensurability of semantic web ontologies. *Electronic Journal of Knowledge Management* 8, 1 (2010), 91–102. Keywords: ontology matching, ontology evaluation, ontology quality
M	Diana Maynard, Wim Peters, and Yaoyong Li. 2006. Metrics for evaluation of ontology-based information extraction. In *International World Wide Web Conference*. 1–8. Keywords: ontology, ontological classification, evaluation, assessment, metrics
M	Melinda H. McDaniel. 2017. *An Automated System for the Assessment and Ranking of Domain Ontologies*. Computer Science Dissertation. Georgia State University (scholarworks.gsu.edu). Keywords: Domain ontology, interoperability, metrics, ontology assessment, ontology evaluation
M	Melinda McDaniel, Veda C. Storey, and Vijayan Sugumaran. 2018. Assessing the quality of domain ontologies: Metrics and an automated ranking system. *Data & Knowledge Engineering* 115 (2018), 32–47. Keywords: Domain ontology, interoperability, metrics, ontology assessment, ontology evaluation, ranking, ontology, semiotics, semiotic layers, domain ontology ranking system
M	Lalit Mohan, Gollapudi V. R. J. Sai Prasad, Sridhar Chimalakonda, Y. Raghu Reddy, and Venkatesh Choppella. 2017. *A Lightweight Approach for Evaluating Sufficiency of Ontologies*. Software Engineering Research Center, IIIT Hyderabad, Hyderabad, India. Keywords: ontology evaluation; sufficiency; natural language processing
M	Huang Ning and D. I. A. O. Shihan. 2006. Structure-based ontology evaluation. In *International Conference on e-Business Engineering (ICEBE'06)*. 132–137. Keywords: ontology evaluation, structure, syntax, criteria, assessment, metrics, attributes
M	Sunju Oh, Heon Y. Yeom, and Joongho Ahn. 2011. Cohesion and coupling metrics for ontology modules. *Information Technology and Management* 12, 2 (2011), 81–96. Keywords: ontology metrics, ontology evaluation, cohesion, coupling
M	Anthony Mark Orme, Haining Yao, and Letha H. Etzkorn. 2009. Complexity metrics for ontology based information. *International Journal of Technology Management* 47, 1–3 (2009), 161–173. Keywords: ontologies, ontology-based systems, semantic web, metrics, complexity metrics, bioinformatics, genomics
M	Jinsoo Park, Wonchin Cho, and Sangkyu Rho. 2007. Evaluation framework for automatic ontology extraction tools: an experiment. In *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops*. Springer, Berlin/Heidelberg, 511–521. Keywords: ontology tools, ontology evaluation, metrics
M	Chintan Patel, Kaustubh Supekar, Yugyung Lee, and E. K. Park. 2003. OntoKhoj: a semantic web portal for ontology searching, ranking and classification. In *Proceedings of the 5th ACM International Workshop on Web Information and Data Management*. 58–61. Keywords: ontology selection, ontology ranking, metrics, classification
M	H. Sofia Pinto, Steffen Staab, and Christoph Tempich. 2004. DILIGENT: Towards a fine-grained methodology for distributed, loosely-controlled and evolving engineering of ontologies. In *Proceedings of the 16th European Conference on Artificial Intelligence*. IOS Press, 393–397. Keywords: ontology, knowledge base, taxonomy, evaluation, metrics, assessment, criteria
M	F. Porzel and R. Malaka. 2004. A task-based approach for ontology evaluation. In *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI'04) Workshop on Ontology Learning and Population, Valencia, Spain*. Keywords: ontology, effectiveness, assessment, evaluation, semantics, pragmatics, metrics
M	Lila Rao and Kweku-Muata Osei-Bryson. 2007. Towards defining dimensions of knowledge systems quality. *Expert Systems with Applications* 33, 2 (2007), 368–378. Keywords: quality dimensions, ontology, knowledge assets, criteria, metrics
M	Marcos Martínez Romero, José M. Vázquez-Naya, Cristian R. Munteanu, Javier Pereira, and Alejandro Pazos. 2010. An approach for the automatic recommendation of ontologies using collaborative knowledge. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. 74–81. Keywords: ontology recommendation, context, coverage, metrics, social context, semantics

M	Marco Rospocher, Sara Tonelli, Luciano Serafini, and Emanuele Pianta. 2012. Corpus-based terminological evaluation of ontologies. *Applied Ontology* 7, 4 (2012), 429–448. Keywords: Corpus based ontology evaluation, terminological ontology evaluation, key-concept extraction, metrics, ontology building environment
M	Jiabin Ruan and Yubin Yang. 2010. Assess content comprehensiveness of ontologies. In *Second International Conference on Computer Modeling And Simulation (ICCMS'10)*, Vol. 1. 536–539. Keywords: corpus, ontology, comprehensiveness, evaluate, metric
M	Pum-Mo Ryu and Key-Sun Choi. 2004. Determining the specificity of terms based on information theoretic measures. In *Proceedings of CompuTerm 2004: 3rd International Workshop on Computational Terminology*. 87–90. Keywords: ontology evaluation, metrics, specificity, relevance, assessment, information modeling
M	Ying Shen, Daoyuan Chen, Buzhou Tang, Min Yang, and Kai Lei. 2018. EAPB: entropy-aware path-based metric for ontology quality. *Journal of Biomedical Semantics* 9, Article 20 (2018), 9–20. Keywords: Ontology evaluation,Ontology modeling, Entropy-based metrc, Knowledge representation, Big data and semantics
M	Muhammed Shoaib and, M., K. Kalsoom, S. Majid, and Fiaz Majeed. 2011. Software metrics for an efficient design of ontologies. *Pakistan Journal of Science* 63, 1. Keywords: ontology assessment, ontology metrics, ontology
M	Miguel-Angelo Sicilia, M. A., Daniel Rodríguez, Elena García-Barriocanal, and Salvador Sánchez-Alonso. 2012. *Empirical findings on ontology metrics. Expert Systems with Applications* 39, 8 (2012), 6706–6711. Keywords: ontology, metrics, Swoogle, OntoRank
M	Peter Spyns. 2005. *EvaLexon: Assessing Triples Mined from Texts.* Technical Report 09. STAR Lab, Brussels, Belgium. 1–16. Keywords: text mining, ontology evaluation, metrics, accuracy, recall, precision
M	Steffen Staab, Asunción Gómez-Pérez, W. Daelemana, M.-L. Reinberger, and Natalya F. Noy. 2004. Why evaluate ontology technologies? Because it works! *IEEE Intelligent Systems* 19, 4 (2004), 74–81. Keywords: ontology learning, ontology application, ontology evaluation, framework, metrics
M	Besiki Stvilia, Les Gasser, Michael B. Twidale, and Linda C. Smith. 2007. A framework for information quality assessment. *Journal of the Association for Information Science and Technology* 58, 12 (2007), 1720–1733. Keywords: ontology evaluation, evaluation criteria, framework, metrics
M	Besiki Stvilia. 2007. A model for ontology quality evaluation. *First Monday* 12, 12. Keywords: Ontologies, quality measurement, activity theory, biodiversity, metrics
M	He Tan, Anders Adlemo, Vladimir Tarasov, and Mats E. Johansson. 2017. Evaluation of an application ontology. In *Proceedings of the Joint Ontology Workshops Episode 3: The Tyrolean Autumn of Ontology Bozen-Bolzano*, Vol. 2050. Keywords: application ontologies, ontology evaluation, ontology quality features, ontology verbalization, usability, correctness, applicability
M	Samir Tartir and I. Budak Arpinar. 2007. Ontology evaluation and ranking using OntoQA. In *International Conference on Semantic Computing (ICSC'07).* 185–192. Keywords: ontology evaluation, ontology validation, metrics, criteria
M	Samir Tartir, I. Budak Arpinar, and Amit P. Sheth. 2010. Ontological evaluation and validation. In *Theory and Applications of Ontology: Computer Applications.* 115–130. Keywords: ontology evaluation, ontology validation, metrics, criteria
M	Samir Tartir, I. Budak Arpinar, Michael Moore, Amit P. Sheth, and Boanerges Aleman-Meza. 2005. *OntoQA: Metric-Based Ontology Quality Analysis.* Keywords: ontology evaluation, ontology validation, metrics, criteria
M	Samir Tartir. 2009. *Ontology-driven Question Answering and Ontology Quality Evaluation.* PhD Dissertation. University of Georgia, Athens, GA, USA. Keywords: ontology evaluation, metrics, automation, tool
M	Paola Velardi, Roberto Navigli, Alessandro Cuchiarelli, and Francesca Neri. 2005. Evaluation of OntoLearn, a methodology for automatic learning of domain ontologies. In *Ontology Learning and Population.* Keywords: ontology learning, ontology application, ontology evaluation, framework, metrics
M	Johanna Völker, Denny Vrandečić, York Sure, and Andreas Hotho. 2008. AEON–An approach to the automatic evaluation of ontologies. *Applied Ontology* 3, 1–2 (2008), 41–62. Keywrds: OntoClean, ontology evaluation, metrics, automation
M	Denny Vrandečić and York Sure. 2007. How to design better ontology metrics. In *The Semantic Web: Research and Applications.* 311–325. Keywords: ontology evaluation, normalization, stable metrics
M	Pornpit Wongthongtham and Behrang Zadjabbari. 2009. Signifying ontology complexity for knowledge sharing. In *International Conference for Internet Technology and Secured Transactions (ICITST'09).* 1–5. Keywords: ontology, complexity, metrics, assessment

M	Haining Yao, Anthony Mark Orme, and Letha Etzkorn. 2005. Cohesion metrics for ontology design and application. *Journal of Computer Science* 1, 1 (2005), 107–113. Keywords: metrics, cohesion metrics, ontology, ontology based system, semantic web, modularity
M	Jonathan Yu, James A. Thom, and Audrey Tam. 2007. Ontology evaluation using wikipedia categories for browsing. In *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management.* 223–232. Keywords: Ontology evaluation, Browsing, User Studies, Wikipedia, Quality Criteria, Metrics
M	Hongyu Zhang, Yuan-Fang Li, and Hee Beng Kuan Tan. 2010. Measuring design complexity of semantic web ontologies. *Journal of Systems and Software* 83, 5 (2010), 803–814. Keywords: ontology, ontology quality metrics, complexity
M	Y. A. N. G. Zhe, Dalu Zhang, and Y. E. Chuan. 2006. Evaluation metrics for ontology complexity and evolution analysis. In *Proceedings of the IEEE International Conference on e-Business Engineering (ICEBE'06).* IEEE, 162–170. Keywords: ontology evaluation, ontology metrics, ontology complexity
M	Hong Zhu, Dongmei Liu, Ian Bayley, Arantza Aldea, Yunfei Yang, and Ying Chen. 2017. Quality model and metrics of ontology for semantic descriptions of web services. *Tsinghua Science and Technology* 22, 3 (2017), 254–272. Keywords: semantic web services; ontology; quality model; ontology evaluation; ontology metrics
M	Hongwei Zhu and Harris Wu. 2014. Assessing the quality of large-scale data standards: A case of XBRL GAAP Taxonomy. *Decision Support Systems* 59 (2014), 351–360. Keywords: information quality, data quality, data standards, quality assessment, metrics
M	Rok Žontar and Marjan Heričko. 2012. Adoption of object-oriented software metrics for ontology evaluation. In *Proceedings of the Fifth Balkan Conference in Informatics.* 298–301. Keywords: ontology evaluation, object-oriented methodology, software metrics
O	Bernardo Cuenca Grau, Ian Horrocks, Yevgeny Kazakov, and Ulrike Sattler. 2008. Modular reuse of ontologies: Theory and practice. *Journal of Artificial Intelligence Research* 31 (2008), 273–18. Keywords: ontology reuse, modularity, refinement, pruning, assessment, ontology evaluation
O	Faezeh Ensan and Weichang Du. 2013. A semantic metrics suite for evaluating modular ontologies. *Information Systems* 38, 5 (2013), 745–770. Keywords: modular ontologies, ontology evaluation, measurements, semantic metrics, cohesion, coupling
O	Michael Grüninger, Torsten Hahmann, Ali Hashemi, Darren Ong, and Atalay Ozgovde. 2012. Modular first-order ontologies via repositories. *Applied Ontology* 7, 2 (2012), 169–209. Keywords: ontology repository, modularity, first-order logic, non-conservative extension, conservative extension, reducibility, similarity
O	Michael Grüninger. 2013. *Ontology Evaluation Workflow in COLORE.* http://ontolog.cim3.net/file/work/OntologySummit2013/2013-02-14_OntologySummit2013_OntologyEvaluation-SoftwareEnvironment/wip/Gruninger_colore-summit_20130214a.pdf. Keywords: ontology repository, ontology evaluation, ontology integration, modularity
O	Zubeida Casmod Khan and C. Maria Keet. 2016. Dependencies between modularity metrics towards improved modules. In *Proceedings of the Knowledge Engineering and Knowledge Management: 20th International Conference (EKAW'16), November 19-23, 2016, Bologna, Italy.* 400–415. Keywords: ontology quality, modularity, assessment, metrics, modularity metrics
O	Oliver Kutz and Joana Hois. 2012. Modularity in ontologies. *Applied Ontology* 7, 2 (2012), 109–112. Keywords: ontology quality, modularity, semantics
O	Farhad Mostowfi and Farshad Fotouhi. 2006. Improving quality of ontology: An ontology transformation approach. In *Proceedings of the 22nd International Conference on Data Engineering Workshops (ICDEW'06).* 61– 61. Keywords: framework, ontology quality, ontology refining
O	Mina Ardakani Movaghati and Ahmad Abdollahzadeh Barforoush. 2016. Modular-based measuring semantic quality of ontology. In *Proceedings of the 6th International Conference on Computer and Knowledge Engineering (ICCKE'16).* Keywords: ontology; ontology measurement; semantic quality assessment; semantic similarity metric; modular-based measurement

APPENDIX B: ILLUSTRATIVE EXAMPLE OF EXPANDED PAPER CLASSIFICATION

This appendix includes an example of a paper that is retrieved from the Web of Science, consolidated into a list of keywords, analyzed using Latent Semantic Analysis techniques, and clustered into an appropriate group. This paper was one of the papers not predominately on ontology evaluation, but one of the larger set of 1,252 papers identified as being about ontologies and including at least a small portion on how an ontology is evaluated.

Step 1 - Document retrieval from the Web of Science

The following paper was retrieved from the Web of Science as a result of our search parameters because its title contained the word *ontology* and the phrase *the system was evaluated* appeared in the abstract.

Title:

Ontology-Based Information Extraction for Subject-Focused Automatic Essay Evaluation (Ajetunmobi and Daramola 2017)

Abstract:

Automatic essay evaluation (AEE) systems are designed to assist a teacher in the task of classroom assessment in order to alleviate the demands of manual subject evaluation. However, although numerous AEE systems are available, most of these systems do not use elaborate domain knowledge for evaluation, which limits their ability to give informative feedback to students and also their ability to constructively grade a student based on a particular domain of study. This paper is aimed at improving on the achievements of previous studies by providing a subject-focused evaluation system that considers the domain knowledge while scoring and provides informative feedback to its user. The study employs a combination of techniques such as system design and modelling using Unified Modelling Language (UML), information extraction, ontology development, data management, and semantic matching in order to develop a prototype subject-focused AEE system. **The developed system was evaluated to determine its level of performance and usability. The result of the usability evaluation showed that the system has an overall mean rating of 4.17 out of maximum of 5, which indicates "good usability." In terms of performance, the assessment done by the system was also found to have sufficiently high correlation with those done by domain experts, in addition to providing appropriate feedback to the user.**

Step 2 - Create relevant keyword list

Only the last portion of the abstract (boldfaced) related to how the ontology was evaluated for quality. Therefore, the keywords generated by Latent Semantic Analysis using only this portion of the abstract generated the relevant keyword.

Relevant Keywords: developed, evaluated, determine, level, performance, usability, result, evaluation, showed, overall, mean, rating, maximum, indicates good usability

Step 3: Classification

Latent Semantic Analysis works by first taking a large set of documents (called the corpus) and analyzing patterns of words in the text. Part of this analysis is ranking each word in the corpus on its influence in a particular document as opposed to its overall use throughout the corpus. For example, if a word such as *system* appears many times in the corpus, it is not considered to be an influential word in a particular paper. From the words in the keyword list that were ranked most influential (high-loading words) in setting the Ajetunmobi and Daramola (2017) paper apart from the other papers in the corpus are the following terms:

Highest-loading terms in the sample document:

mean, usability, rating, maximum good

Another technique of Latent Semantic Analysis is clustering documents in the corpus based on their common high-loading words. After a clustering was performed for our corpus of documents, four groups emerged as having a large number of documents that matched a particular set of terms. We examined each of these sets of terms, along with the documents that matched them most closely. Each cluster of papers and its corresponding set of terms was assigned a meaningful name. One the clusters, containing documents related closely to the following set of terms, was shown to be a cluster related to using metrics for ontology quality ranking. The Ajetunmobi and Daramola (2017) paper ranked the highest among all the papers as a match for this classification.

Highest-loading terms in the "using metrics" cluster of documents:
performance, rating, mean, usability, evaluate, determine, level, result overall mean, maximum, indicate good ranking, benchmark analysis

ACKNOWLEDGEMENTS

The authors wish to thank Editor-in-Chief Sartaj Sahni, Associate Editor Wenji Mao, and the anonymous reviewers for their work in processing our article and for their insightful comments and suggestions. The authors also thank their colleagues for their assistance: Balasubramaniam Ramesh, Rajshekhar Sunderraman, Kai Larsen, Roman Lukyanenko, Jeffrey Parsons, Leo Obrst, Maria Poveda-Villalón, John Sowa, and Vijayan Sugumaran.

REFERENCES

Manel Achichi, Michelle Cheatham, Zlatan Dragisic, Jérôme Euzenat, Daniel Faria, Alfio Ferrara, and Ernesto Jiménez-Ruiz. 2016. Results of the 2016 ontology alignment evaluation initiative. In *CEUR Workshop Proceedings*, 1766. 73–129.

Stephanie Abimbola Ajetunmobi and Olawande Daramola. 2017. Ontology-based information extraction for subject-focussed automatic essay evaluation. In *2017 International Conference on Computing Networking and Informatics (ICCNI'17)*. 1–6.

Joel Adams and Steven Bedrick. 2014. Automatic classification of Pubmed abstracts with latent semantic indexing: Working notes. In *CLEF (Working Notes)*. 1275–1282.

Fawaz S. Al-Anzi and Dia AbuZeina. 2017. Toward an enhanced Arabic text classification using cosine similarity and latent semantic indexing. *Journal of King Saud University-Computer and Information Sciences* 29, 2 (2017), 189–195.

Khalid Albarrak and Edgar Sibley. 2012. Measuring expressivity between ontology models. In *Proceedings of the 11th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED'12)*.

Hend S. Al-Khalifa, Maha M. Al-Yahya, Alia Bahanshal, and Iman Al-Odah. 2012. On the evaluation of linguistic ontological models: An application on the SEMQ ontology. In *2012 7th International Conference on Digital Information Management (ICDIM'12)*. 341–345

Kiehl Alm, B. Lantow, and K. Sandkuhl. 2013. Applicability of quality metrics for ontologies on ontology design patterns. In *Proceedings of the 5th International Conference on Knowledge Engineering and Ontology Development (KEOD'13)*. 48–57.

Maziar Amirhosseini and Juhana Salim. 2011. Ontoabsolute as an ontology evaluation methodology in analysis of the structural domains in upper, middle and lower level ontologies. In *2011 International Conference on Semantic Technology and Information Retrieval (STAIR'11)*. 26–33.

Jamshaid Ashraf, Omar K. Hussain, Farookh Khadeer Hussain, and Elizabeth J. Chang. 2018. Evaluation of u ontology. In *Measuring and Analysing the Use of Ontologies*. Springer, Cham, 243–268.

Mauricio Barcellos Almeida. 2009. A proposal to evaluate ontology content. *Applied Ontology* 4, 3 (2009), 245–265.

Ritika Bansal and Sonal Chawla. 2015. Evaluation metrics for computer science domain specific ontology in semantic web based IRSCSD system. *International Journal of Computer (IJC)* 19, 1 (2015), 129–139.

Michael Atkins. 2017. FIBO ontologies by domain and module. Retrieved from https://edmcouncil.org.

Kenneth Baclawski, Mike Bennett, Gary Berg-Cross, Donna Fritzsche, and Andrea Westerninen. 2017. Ontology summit 2017 communiqué - ai, learning, reasoning and ontologies. *Applied Ontology* 13, 1 (2017), 3–18.

Niyati Baliyan and Sandeep Kumar. 2016. A behavioral metrics suite for modular ontologies. In *Proceedings of the 2nd International Conference on Information and Communication Technology for Competitive Strategies*. ACM, 133.

Judson Bandeira, Ig Ibert Bittencourt, Patricia Espinheira, and Seiji Isotani. FOCA: A methodology for ontology evaluation. *Arxiv Preprint Arxiv*:1612.03353.

Montserrat Batet and David Sánchez. 2014. A semantic approach for ontology evaluation. In *IEEE 26th International Conference on Tools with Artificial Intelligence (ICTAI'14)*. IEEE, 138–145.

Mike Bennett. 2013. The financial industry business ontology: Best practice for big data. 2013. *Journal of Banking Regulation* 14, 3–4 (2013), 255–268.

Pooyan Ramezani Besheli. 2018. The pattern of patterns: What is a pattern in conceptual modeling? In *12th International Workshop on Value Modeling and Business Ontologies (VMBO'18)*. 99–106.

Kent D. Bimson and Richard D. Hull. 2016. Unnatural language processing: Characterizing the challenges in translating natural language semantics into ontology semantics. In *Semantic Web*. 119–135.

Christian Bizer, Tom Heath, and Tim Berners-Lee. 2009. Linked data-the story so far. In *Semantic Services, Interoperability and Web Applications: Emerging Concepts*. 205–227.

Palash Bera, Andrew Burton-Jones, and Yair Wand. 2014. Research note-how semantics and pragmatics interact in understanding conceptual models. *Information Systems Research* 25, 2 (2014), 401–419.

Tim Berners-Lee, James Hendler, and Ora Lassila. 2001. The semantic web. *Scientific American*. 284, 5 (2001), 28–37.

Olivier Bodenreider. 2004. The unified medical language system (UMLS): Integrating biomedical terminology. *Nucleic Acids Research* 32, Suppl 1 (2004), D267–D270

Oliver Bodenreider. 2018. Evaluating the Quality and Interoperability of Biomedical Terminologies. Lister Hill Nation Center for Biomedical Communications. https://lhncbc.nlm.nih.gov/system/files/pub9754.pdf.

Janez Brank, Marko Grobelnik, and Dunja Mladenić. 2007. Automatic evaluation of ontologies. In *Natural Language Processing and Text Mining*. 193–219.

John G. Breslin, Stefan Decker, Andreas Harth, and Uldis Bojars. 2006. SIOC: An approach to connect web-based communities. *International Journal of Web Based Communities* 2, 2 (2006), 133–142.

Christopher Brewster, Harith Alani, Srinandan Dasmahapatra, and Yorick Wilks. 2004. Data driven ontology evaluation. In *International Conference on Language Resources and Evaluation*. 24–30.

Susan Brown, Claire Bonial, Leo Obrst, and Martha Palmer. 2017. The rich event ontology. In *Proceedings of the Events and Stories in the News Workshop*. 87–97.

Paul Buitelaar, Thomas Eigner, and Thierry Declerck. 2004. OntoSelect: A dynamic ontology library with support for ontology selection. In *Proceedings of the Demo Session at the International Semantic Web Conference (ISWC'04)*.

Andrew Burton-Jones, Veda C. Storey, Vijayan Sugumaran, and Punit Ahluwalia. 2005. A semiotic metrics suite for assessing the quality of ontologies. *Data and Knowledge Engineering* 55, 1 (2005), 84–102.

Cecilia Camporeale, Antonio De Nicola, and Maria Luisa Villani. 2015. Semantics-based services for a low carbon society: An application on emissions trading system data and scenarios management. *Environmental Modelling & Software* 64 (2015), 124–142.

Michal Campr and Karel Ježek. 2015. Comparing semantic models for evaluating automatic document summarization. In *International Conference on Text, Speech, and Dialogue*. Springer, Cham, 252–260.

Ricardo Chalmeta and Verónica Pazos. 2015. A step-by-step methodology for enterprise interoperability projects. *Enterprise Information Systems*. 9, 4 (2015), 436–464.

Balakrishnan Chandrasekaran, John R. Josephson, and V. Richard Benjamins. 1999. What are ontologies, and why do we need them? *IEEE Intelligent Systems* 1 (1999), 20–26.

Michelle Cheatham and Catia Pesquita. 2017. Semantic data integration. In *Handbook of Big Data Technologies*. Springer, 263–305.

Jianpeng Cheng and Mirella Lapata. 2106. Neural summarization by extracting sentences and words. *Arxiv Preprint ArXiv:1603.07252*.

Namyoun Choi, Il-Yeol Song, and Hyoil Han. 2006. A survey on ontology mapping. *ACM SIGMOD Record* 35, 3 (2006), 34–41.

Djellali Choukri. 2014. A new distributed expert system to ontology evaluation. *Procedia Computer Science* 37 (2014), 48–55.

Mihai Codescu, Eugen Kuksa, Oliver Kutz, Till Mossakowski, and Fabian Neuhaus. 2017. Ontohub: A semantic repository engine for heterogeneous ontologies. *Applied Ontology* 12, 3–4 (2017), 275–298.

Licong Cui, Wei Zhu, Shiqiang Tao, James T. Case, Olivier Bodenreider, and Cuo-Qiang Zhang. 2017. Mining non-lattice subgraphs for detecting missing hierarchical relations and concepts in SNOMED CT. *Journal of the American Medical Informatics Association* 24, 4 (2017), 788–798.

Kathleen Dahlgren. 1995. A linguistic ontology. *International Journal of Human-Computer Studies* 43, 5 (1995), 809–818.

J. P. de Villiers, R. W. Focke, G. Pavlin, A. L. Jousselme, V. Dragos, K. B. Laskey, P. C. Costa, and E. Blasch. 2017. Evaluation metrics for the practical application of URREF ontology: An illustration on data criteria. In *2017 20th International Conference on Information Fusion (Fusion'17)*. 1–8.

Auriol Degbelo. 2017. A snapshot of ontology evaluation criteria and strategies. In *Proceedings of the 13th International Conference on Semantic Systems*. ACM, 1–8.

M. Syamala Devi and Himani Mittal. 2016. Machine learning techniques with ontology for subjective answer evaluation. *Arxiv Preprint Arxiv*:1605.02442.

Michael Dibley, Haijiang Li, Yacine Rezgui, and John Miles. 2012. An ontology framework for intelligent sensor-based building monitoring. *Automation in Construction* 28 (2012), 1–14.

Astrid Duque-Ramos, Jesualdo Tomas Fernández-Breis, Robert David Stevens, and Nathalie Aussenac-Gilles. 2011. OQuaRE: A SQuaRE-based approach for evaluating the quality of ontologies. *Journal of Research and Practice in Information Technology* 43, 2 (2011), 159.

Astrid Duque-Ramos, Martin Boeker, Ludger Jansen, Stefan Schulz, Miguela Iniesta, and Jesualdo Tomás Fernández-Breis. 2014. Evaluating the good ontology design guideline (GoodOD) with the ontology quality requirements and evaluation method and metrics (OQuaRE). *PloS One* 9, 8 (2014), e104463.

Mathieu d'Aquin, Anne Schlicht, Heiner Stuckenschmidt, and Marta Sabou. 2007, September. Ontology modularization for knowledge selection: Experiments and evaluations. In *International Conference on Database and Expert Systems Applications*. Springer, 874–883.

Mathieu d'Aquin and Natalya F. Noy. 2012. Where to publish and find ontologies? A survey of ontology libraries. *Web Semantics: Science, Services and Agents on the World Wide Web* 11 (2012), 96–111.

Donna Marie DeCarolis and David L. Deeds. 1999. The impact of stocks and flows of organizational knowledge on firm performance: An empirical investigation of the biotechnology industry. *Strategic Management Journal* 20, 10 (1999), 953–968.

Faezeh Ensan and Weichang Du. 2013. A semantic metrics suite for evaluating modular ontologies. *Information Systems* 38 (2013), 5.

Tamar A. El-Diraby, Lima Celson, and B. Feis. 2005. Domain taxonomy for construction concepts: Toward a formal ontology for construction knowledge. *Journal of Computing in Civil Engineering* 19, 4 (2005), 394–406.

Jérôme Euzenat. 2001, August. Towards a principled approach to semantic interoperability. In *Proceedings of the IJCAI 2001 Workshop on Ontology and Information Sharing*. 19–25.

Nicholas Evangelopoulos, Xiaoni Zhang, and Victor R. Prybutok. 2012. Latent semantic analysis: Five methodological recommendations. *European Journal of Information Systems* 21, 1 (2012), 70–86.

Muhammad Fahad and Muhammad Abdul Qadir. 2008. A framework for ontology evaluation. *ICCS Supplement* 354 (2008), 149–158.

Ricardo de Almeida Falbo, Monalessa Perini Barcellos, Julio Cesar Nardi, and Giancarlo Guizzardi. 2013. Organizing ontology design patterns as ontology pattern languages. In *Extended Semantic Web Conference*. Springer, Berlin, 61–75.

Christiane Fellbaum. 1998. WordNet. In *The Encyclopedia of Applied Linguistics*.

Donna Fritzsche, Michael Grüninger, Ken Baclawski, Mike Bennett, Gary Berg-Cross, Todd Schneider, Ram Sriram, Mark Underwood, and Andrea Westerinen. 2017. Ontology summit 2016 communique: Ontologies within semantic interoperability ecosystems. *Applied Ontology* 12, 2 (2017), 91–111.

Salvatore Gaglio and Giuseppe Lo Re. 2014. *Advances onto the Internet of Things*. Springer, 349.

Aldo Gangemi, Nicolar Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schneider 2002. Sweetening ontologies with DOLCE. In *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*. 166–181.

Aldo Gangemi, Carola Catenacci, Massimiliano Ciaramita, and Jos Lehmann 2006. Modelling ontology evaluation and validation. In *The Semantic Web: Research and Applications*. 140–154.

Jort F. Gemmeke, Daniel P. W. Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R. Channing Moore, Manoj Plakal, and Marvin Ritter. 2017. Audio Set: An ontology and human-labeled dataset for audio events. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'17)*. 776–780.

John H. Gennari, Mark A. Musen, Ray W. Fergerson, William E. Grosso, Monica Crubézy, Henrik Eriksson, and Samson W. Tu. 2003. The evolution of protégé: An environment for knowledge-based systems development. *International Journal of Human-Computer Studies* 58, 1 (2003), 89–12.

Toader Gherasim, Giuseppe Berio, Mounira Harzallah, and Pascale Kuntz. 2012. Problems impacting the quality of automatically built ontologies. In *Knowledge Engineering and Software Engineering (KESE8'12)*. 22.

Asunción Gómez-Pérez. 1996. Towards a framework to verify knowledge sharing technology. *Expert Systems with Applications* 11, 4 (1996), 519–529.

Asunción Gómez-Pérez. 2001. Evaluation of ontologies. *International Journal of Intelligent Systems*. 16, 3 (2001), 391–409.

Asunción Gómez-Pérez. 2004. Ontology evaluation. In *Handbook on Ontologies*. 251–273.

Rafael S. Gonçalves, Martin J. O'Connor, Marcos Martínez-Romero, John Graybeal, and Mark A. Musen. 2017. Metadata in the biosample online repository are impaired by numerous anomalies. *ArXiv Preprint Arxiv*:1708.01286 2017.

Sotirios K. Goudos, Vassilios Peristeras, and Konstantinos Tarabanis. 2006. Mapping citizen profiles to public administration services using ontology implementations of the governance enterprise architecture GEA models. In *Proceedings of 3rd Annual European Semantic Web Conference*. 25–37.

Thomas R. Gruber 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition* 5, 2 (1993), 199–220.

Thomas R. Gruber. 1995. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies* 43, 5 (1995), 907–928.

Michael Grüninger, Torsten Hahmann, Ali Hashemi, Darren Ong, and Atalay Ozgovde. 2012. Modular first-order ontologies via repositories. *Applied Ontology* 7, 2 (2012), 169–209.

Michael Grüninger and Megan Katsumi. 2012. Specifying ontology design patterns with an ontology repository. In *WOP*. 929.

Nicola Guarino and Christopher Welty. 2002. Evaluating ontological decisions with Ontoclean. *Communications of the ACM* 45, 2 (2002), 61–65.

Nicola Guarino, Daniel Oberle, and Steffen Staab. 2009. What is an ontology? In *Handbook on Ontologies*. 1–17.

Nicola Guarino and Mark A. Musen. 2005. Applied ontology: Focusing on content. *Applied Ontology* 1, 1 (2005), 1–5.
Nicola Guarino and Mark Alan Musen. 2015. Applied ontology: The next decade begins. *Applied Ontology* 10, 3–4 (2015), 1–4.
Nicola Guarino and Giancarlo Guizzardi. 2016. Relationships and events: Towards a general theory of reification and truth-making. In *Conference of the Italian Association for Artificial Intelligence*. Springer, Cham, 237–249.
Nicola Guarino. 2017. BFO and DOLCE: So Far, so close. In *Cosmos Taxis* 4.
Giancarlo Guizzardi. 2007. On ontology, ontologies, conceptualizations, modeling languages, and meta models. *Frontiers in Artificial Intelligence and Applications* 155 (2007), 18.
Giancarlo Guizzardi. 2010. Theoretical foundations and engineering tools for building ontologies as reference conceptual models. *Semantic Web* 1, 2 (2010), 3–10.
Giancarlo Guizzardi, Ricardo de Almeida Falbo, and Renata S. S. Guizzardi. 2008. Grounding software domain ontologies in the unified foundational ontology UFO: The case of the ODE software process ontology. In *CIbSE*. 127–140.
Giancarlo Guizzardi, Luis Ferreira Pires, and Marten J. Van Sinderen. 2002. On the role of domain ontologies in the design of domain-specific visual modeling languages. In *Proceedings of the 2nd Workshop on Domain-Specific Visual Languages*.
Giancarlo Guizzardi, Gerd Wagner, João Paulo Andrade Almeida, and Renata S. S. Guizzardi. 2015. Towards ontological foundations for conceptual modeling: The unified foundational ontology (UFO) story. *Applied Ontology* 10, 3–4 (2015), 259–271.
Byeong-Jun, Han, Seungmin Rho, Sanghoon Jun, and Eenjun Hwang. 2010. Music emotion classification and context-based music recommendation. *Multimedia Tools and Applications* 47, 3 (2010), 433–460.
Jens Hartmann, Peter Spyns, Alain Giboin, Diana Maynard, Roberta Cuel, Mari Carmen Suárez-Figueroa, and York Sure. 2005. D1.2.3 methods for ontology evaluation. *EU-IST Network of Excellence NoE IST-2004-507482 KWEB Deliverable* D1.
Jim Hendler and Tim Berners-Lee. 2010. From the semantic web to social machines: A research challenge for AI on the world wide web. *Artificial Intelligence* 174, 2 (2010), 156–161.
Martin Hepp, Daniel Bachlechner, and Katharina Siorpaes. 2006. OntoWiki: Community-driven ontology engineering and ontology usage based on wikis. In *Proceedings of the 2006 International Symposium on Wikis*. ACM.
Martin Hepp. 2007. Possible ontologies: How reality constrains the development of relevant ontologies. 2007. *IEEE Internet Computing* 11 (2007), 1.
Amanda Hicks. 2017. Metrics and methods for comparative ontology evaluation. *Ciência da Informação* 46 (2017), 1.
Julia Hirschberg and Christopher D. Manning. 2015. Advances in natural language processing. *Science* 349, 6245 (2015), 261–266.
Hlomani Hlomani and Deborah Stacey. 2014. Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey. *Semantic Web and Information Systems* 1, 5 (2014), 1–11.
Mike Hobbs, Cristina Luca, Arooj Fatima, and Mark Warnes. 2014. Ontological analysis for dynamic data model exploration. *Electronic Journal of Applied Statistical Analysis: Decision Support Systems and Services Evaluation* 5, 1 (2014), 42–56.
Robert Hoehndorf, Michel Dumontier, and Georgios V. Gkoutos. 2012. Evaluation of research in biomedical ontologies. In *Briefings in Bioinformatics*. bbs053.
James Huang, Stephanie Rogers, and Eunkwang Joo. 2014. Improving restaurants by extracting subtopics from yelp reviews. In *iConference 2014 (Social Media Expo)*.
Khairul Nurmazianna Ismail. 2014. Development of durian ontology from unstructured text and external knowledge source. PhD diss., Universiti Teknologi MARA.
Marzanah A. Jabar, Mustafa Salah Khalefa, Rusli H. Abdullah, and Salfarina Abdullah. 2014. Meta-analysis of ontology software development process. *International Review on Computers and Software (IRECOS)* 9, 1 (2014), 29–37.
Dean Jones. 1998. Developing shared ontologies in multi-agent systems. In *ECAI'98 Workshop on Intelligent Information Integration*.
Yong-Bin Kang, Yuan-Fang Li, and Shonali Krishnaswamy. 2012. Predicting reasoning performance using ontology metrics. In *The Semantic Web (ISWC'12)*. 198–214.
Pavandeep Kataria, Radmila Juric, Shamimabi Paurobally, and Kambiz Madani. 2008. Implementation of ontology for intelligent hospital wards. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*. IEEE, 253–253.
Megan Katsumi and Michael Grüninger. 2017. Choosing ontologies for reuse. *Applied Ontology* 12, 3–4 (2017), 195–221.
C. Maria Keet, Mari Carmen Suarez-Figueroa, and Maria Poveda-Villalon. 2013. The current landscape of pitfalls in ontologies. In *Proceedings of the 5th International Conference on Knowledge Engineering and Ontology Development*.
Zubeida Casmod Khan and C. Maria Keet. 2015. An empirically-based framework for ontology modularisation. *Applied Ontology* 10, 3–4 (2015), 171–195.
Zubeida Casmod Khan and C. Maria Keet. 2016. ROMULUS: The repository of ontologies for MULtiple USes populated with mediated foundational ontologies. *Journal on Data Semantics*. 5, 1 (2016), 19–36.
M. Rahamatullah Khondoker and Paul Mueller. 2010. Comparing ontology development tools based on an online survey. In *World Congress on Engineering (WCE 10)*.

Jongwoo Kim and Veda C. Storey. 2012. Ontologies: Sourcing the World Wide Web. In *Organizational Efficiency through Intelligent Information Technologies*. 65.

Caroline Chepkoech Kiptoo, Aurona Gerber, and Alta Van der Merwe. 2016. Towards citizen-expert knowledge exchange for biodiversity informatics: A conceptual architecture. *African Journal of Information and Communication* 18 (2016), 33–54.

Philip Kitcher. 1995. The advancement of science: Science without legend, objectivity without illusion. Oxford University Press, 432 pages.

John Krogstie. 2012. *Model-Based Development and Evolution of Information Systems: A Quality Approach*. Springer.

Oliver Kutz and Joana Hois. 2012. Modularity in ontologies. *Applied Ontology* 7, 2 (2012), 109–112.

Axel van Lamsweerde. 2001. Goal-oriented requirements engineering: A guided tour. In *Proceedings of the 5th IEEE International Symposium on Requirements Engineering*. IEEE, 249–262.

Christoph Lange. 2013. Ontologies and languages for representing mathematical knowledge on the Semantic Web. *Semantic Web* 4, 2 (2013), 119–158.

Kai R. Larsen, David E. Monarchi, Dirk S. Hovorka, and Christopher N. Bailey. 2008. Analyzing unstructured text data: Using latent categorization to identify intellectual communities in information systems. *Decision Support Systems* 45, 4 (2008), 884–896.

Kai R. Larsen, Susan Michie, Eric B. Hekler, Bryan Gibson, Donna Spruijt-Metz, David Ahern, and Heather Cole-Lewis. 2017. Behavior change interventions: The potential of ontologies for advancing science and practice. *Journal of Behavioral Medicine. Springer* 40, 1 (2017), 6–22.

Dong Joon Lee and Besiki Stvilia. 2017. Practices of research data curation in institutional repositories: A qualitative view from repository staff. *PloS One* 12, 3 (2017), e0173987.

Douglas B. Lenat. 1995. CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM* 38, 11 (1995), 33–38.

Danfeng Liu, Antonis Bikakis, and Andreas Vlachidis. 2017. Evaluation of semantic web ontologies for modelling art collections. In *Advances in Databases and Information Systems*. Springer, Cham, 343–352.

Adolfo Lozano-Tello and Asunción Gómez-Pérez. 2004. Ontometric: A method to choose the appropriate ontology. *Journal of Database Management* 2, 15 (2004), 1–18.

Roman Lukyanenko and Jeffrey Parsons 2018. Beyond micro-tasks: Research opportunities in observational crowdsourcing. *Journal of Database Management (JDM)* 291 (2018), 1–22.

Yinglong Ma, Beihong Jin, and Yulin Feng. 2010. Semantic oriented ontology cohesion metrics for ontology-based systems. *Journal of Systems and Software* 83, 1 (2010), 143–152.

Yinglong Ma, Ling Liu, Ke Lu, Beihong Jin, and Xiangje Liu. 2014. A graph derivation based approach for measuring and comparing structural semantics of ontologies. *Transactions on Knowledge and Data Engineering* 26, 5 (2014), 1039–1052.

Alexander Maedche and Raphael Volz. 2012. The ontology extraction and maintenance framework Text-To-Onto. In *Proceedings on the Workshop on Integrating Data Mining and Knowledge Management*.

Marcos Martínez-Romero, Clement Jonquet, Martin J. O'Connor, John Graybeal, Alejandro Pazos, and Mark A. Musen. 2017. NCBO Ontology Recommender 2.0: An enhanced approach for biomedical ontology recommendation. *Journal of Biomedical Semantics* 8, 1 (2017), 21.

Natalia Maslova and Vsevolod Potapov. 2017. Neural network doc2vec in automated sentiment analysis for short informal texts. In *International Conference on Speech and Computer*. 546–554.

Cynthia Matuszek, John Cabral, Michael J. Witbrock, and John DeOliveira. 2006. An introduction to the syntax and content of cyc. In *AAAI Spring Symposium: Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering*. 44–49.

Melinda McDaniel, Veda C. Storey, and Vijayan Sugumaran. 2018. Assessing the quality of domain ontologies: Metrics and an automated ranking system. *Data and Knowledge Engineering* 115 (2018), 32–47.

Rada Mihalcea and Paul Tarau. Textrank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*.

David Milne and Ian H. Witten. 2013. An open-source toolkit for mining Wikipedia. *Artificial Intelligence* 194 (2013), 222–239.

Till Mossakowski, Oliver Kutz, and Mihai Codescu. 2014. Ontohub: A semantic repository for heterogeneous ontologies. In *Proceedings of DACS'14*.

Sandeep Sricharan Mukku, Nurendra Choudhary, and Radhika Mamidi. 2016. Enhanced sentiment classification of Telugu text using ML techniques. In *SAAIP@ IJCAI*. 29–34.

Mark A. Musen. 2015. The protégé project: A look back and a look forward. *AI Matters* 1, 4 (2015), 4–12.

Fabian Neuhaus, Amanda Vizedom, Ken Baclawski, Mike Bennett, Mike Dean, Michael Denny, and Peter Yim. 2013. Towards ontology evaluation across the life cycle: The ontology summit 2013. *Applied Ontology* 8, 3 (2013), 179–194.

José Ángel Noguera-Arnaldos, Mario Andrés Paredes-Valverde, María Pilar Salas-Zárate, Miguel Ángel Rodríguez-García, Rafael Valencia-García, and José Luis Ochoa. 2017. IM4 Things: An ontology-based natural language interface for controlling devices in the internet of things. In *Current Trends on Knowledge-Based Systems*. 3–22.

Natalya F. Noy, Nigam H. Shah, Patricia L. Whetzel, Benjamin Dai, Michael Dorf, Nicholas Griffith, and Mark A. Musen. 2009. BioPortal: Ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research* 29, 3 (supp. 2) (2009), W170–173.

Leo Obrst, Pattrick Cassidy, Steven R. Ray, Bradford Smith, Dagobert Soergel, Matthew West, and Peter Yim. 2006. The 2006 Upper Ontology Summit Joint Communiqué. *Applied Ontology* 1, 2 (2006), 203–211.

Leo Obrst, Werner Ceusters, Inderjeet Mani, Steve Ray, and Barry Smith. 2007. Evaluation of ontologies toward improved semantic interoperability. *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, Vol. 9780387484389. Springer US, 139–58.

Leo Obrst, Mark Musen, Barry Smith, Fabian Neuhaus, Frank Olken, Mike Grüninger, M. Raymond, Patrick Hayes, and Raj Sharma. 2008. Ontology Summit 2008: Towards an open ontology repository. http://ontolog.cim3.net/wiki/OntologySummit2008_Communique.html.

Leo Obrst and Patrick Cassidy. 2011. The need for ontologies: Bridging the barriers of terminology and data structure. *Geological Society of America Special Papers* 482 (2011), 99–123.

Leo Obrst, Arturo Sanchez, Amanda Vizedom, Peter Yim, and Barry Smith. 2011. Ontology summit 2010: Creating the ontologists of the future. *Applied Ontology* 6 (2011), 91–98.

Leo Obrst, Michael Grüninger, Kenneth Baclawski, Mike Bennett, Dan Brickley, Gary Berg-Cross, and Peter Yim. 2014. Semantic web and big data meets applied ontology: The Ontology Summit 2014. *Applied Ontology* 9, 2 (2014), 155–170.

Sunju Oh, Heon Y. Yeom, and Joongho Ahn. 2011. Cohesion and coupling metrics for ontology modules. *Information Technology and Management* 12, 2 (2011), 81–96.

Ontology Evaluation Across the Ontology Lifecycle. 2013. Retrieved from https://www.zotero.org/groups/ontologysummit2013.

Anthony M. Orme, Haining Tao, and Letha H. Etzkorn. 2006. Coupling metrics for ontology-based system. *IEEE Software* 23, 2 (2006), 102–108.

Liubo Ouyang, Beiji Zou, Miaoxing Qu, and Chengming Zhang. 2011. A method of ontology evaluation based on coverage, cohesion and coupling. In *8th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'11)*. IEEE, 2451–2455.

Jinsoo Park, Wonchin Cho, and Sangkyu Rho. 2007. Evaluation framework for automatic ontology extraction tools: An experiment. In *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops*. Springer, Berlin, 511–521.

Jeffrey Parsons and Yair Wand. 2008. Using cognitive principles to guide classification in information systems modeling. *MIS Quarterly* 32, 4 (2008), 839–868.

Heilo Paulheim, Sven Hertling, and Dominique Ritze. 2013. Towards evaluating interactive ontology matching tools. In *The Semantic Web: Semantics and Big Datapp*. Springer, Berlin, 31–45.

Charith Perera, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. 2014. Context aware computing for the Internet of Things: A survey. *IEEE Communications Surveys and Tutorials* 16, 1 (2014), 414–454.

Perrine Pittet and Jérôme Barthélémy. 2015. Exploiting users' feedbacks: Towards a task-based evaluation of application ontologies throughout their lifecycle. In *International Conference on Knowledge Engineering and Ontology Development*, 2.

Robert Porzel and Rainer Malaka. 2004. A task-based approach for ontology evaluation. In *ECAI Workshop on Ontology Learning and Population*.

María Poveda-Villalón, Mari Carmen Suárez-Figueroa, and Asunción Gómez-Pérez. 2012. Validating ontologies with Oops! knowledge engineering and knowledge management. In *18th International Conference on Knowledge Engineering and Knowledge Management*. Springer, 267–281.

Maria Poveda-Villalón, Asunción Gómez-Pérez, and Mari Carmen Suárez-Figueroa. 2014. Oops! ontology pitfall scanner!: An on-line tool for ontology evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS)* 102 (2014), 7–34.

Sandeep Purao and Veda C. Storey. 2005. A multi-layered ontology for comparing relationship semantics in conceptual models of databases. *Applied Ontology* 1, 1 (2005), 117–139.

Radim Řehůřek and Petr Sojka. 2011. Gensim—statistical semantics in python. In *Statistical Semantics; Gensim; Python; LDA; SVD*.

David Riaño, Francis Real, Joan Albert López-Vallverdú, Fabio Campana, Sara Ercolani, Patrizia Mecocci, Roberta Annicchiarico, and Carlo Caltagirone. 2012. An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients. *Journal of Biomedical Informatics* 45, 3 (2012), 429–446.

Mariela Rico, María Laura Caliusco, Omar Chiotti, and María Rosa Galli. 2014. OntoQualitas: A framework for ontology quality assessment in information interchanges between heterogeneous systems. *Computers in Industry* 65, 9 (2014), 1291–1300.

Peter Rittgen. 2010. Quality and perceived usefulness of process models. In *Proceedings of the 2010 ACM Symposium on Applied Computing*. ACM, 65–72.

Dumitru Roman, Uwe Keller, Holger Lausen, Jos de Bruijn, Rubén Lara, Michael Stollberg, and Dieter Fensel. 2005. Web service modeling ontology. *Applied Ontology* 1, 1 (2005), 77–106.

Rosetta Romano and Craig McDonald. 2011. Assessing the quality of ontology. In *MCIS 2011 Proceedings*. Paper 50.

Rajendra Kumar Roul and Sanjay Kumar Sahay. 2014. An effective approach for web document classification using the concept of association analysis of data mining. *Arxiv Preprint Arxiv:1406.5616*.

Ahmed R. Sadik and Bodo Urban. 2017. An ontology-based approach to enable knowledge representation and reasoning in worker–Cobot agile manufacturing. *Future Internet* 9, 4 (2017), 90.

Andreas Scheuermann and Joerg Leukel. 2014. Task ontology for supply chain planning—a literature review. *International Journal of Computer Integrated Manufacturing* 27, 8 (2014), 719–732.

Daniel Schober, Ilinca Tudose, Vojtěch Svatek, and Martin Boeker. 2012. OntoCheck: Verifying ontology naming conventions and metadata completeness in protégé 4. *Journal of Biomedical Semantics* 3 (2012), S2–S4.

Todd Schneider, Ali Hashemi, Mike Bennett, Mary Brady, Cory Casanave, Henson Graves, Michael Grüninger, Nicola Guarino, Anatoly Levenchuk, Ernie Lucier, Leo Obrst, Steve Ray, Ram D. Sriram, Amanda Vizedom, Matthew West, Trish Whetzel, and Peter Yim. 2012. Ontology for big systems: The ontology summit 2012 communique. *Applied Ontology* 7, 3 (2012), 357–371.

David G. Schwartz. 2014. Enhancing knowledge marketplaces through the theory of knowledge measurement. In *Handbook of Strategic e-Business Management*. 735–748.

Dudley Shapere. 1984. Scientific theories and their domains. In *Reason and the Search for Knowledge*. Springer, 273–319.

M. Shoaib, K. Kalsoom, S. Majid, and F. Majeed. 2011. Software metrics for an efficient design of ontologies. *Pakistan Journal of Science* 63 (2011), 1.

Anna Sidorova, Nicholas Evangelopoulos, Joseph S. Valacich, and Thiagarajan Ramakrishnan 2008. Uncovering the intellectual core of the information systems discipline. *MIS Quarterly* 32, 3 (2008), 467–482.

Ayrton Silva, Luiz Carlos Chaves, and Damires Souza. 2013. A domain-based approach to publish data on the web. In *Proceedings of International Conference on Information Integration and Web-Based Applications & Services*. ACM, 344.

Bhaskar Sinha, Megha Garg, and Somnath Chandra. 2016. Identification and classification of relations for Indian languages using machine learning approaches for developing a domain specific ontology. In *Computational Techniques in Information and Communication Technologies (ICCTICT'16)*. IEEE, 415–420.

Barry Smith, Michael Ashburner, Cornelius Rosse, Johnathan Bard, William Bug, Werner Ceusters, Loius J. Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J. Mungall, and Neocles Leontis. 2007. The OBO foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology* 25, 11 (2007), 1251–1255.

John F. Sowa. 2000. Ontology, metadata, and semiotics. In *International Conference on Conceptual Structures*. Springer, 55–81.

John F. Sowa. 2006. The challenge of knowledge soup. *Research Trends in Science, Technology and Mathematics Education* (2006), 55–90.

John F. Sowa 2010. The role of logic and ontology in language and reasoning. In *Theory and Applications of Ontology: Philosophical Perspectives*. Springer, 231–263.

John F. Sowa (Ed.). 2014. Principles of semantic networks: Explorations in the representation of knowledge. Morgan Kaufmann publishers Inc.

John F. Sowa. 2015. Signs and reality. *Applied Ontology* 10, 3–4 (2015), 273–284.

John F. Sowa. 2018. Ontology standards: History, lessons, and warnings. Retrieved from http://jfsowa.com/temp/ontostan.pdf.

Ronald Stamper, Kecheng Liu, Mark Hafkamp, and Yasser Ades. 2000. Understanding the roles of signs and norms in organizations—a semiotic approach to information systems design. *Behaviour and Information Technology* 19, 1 (2000), 15–27.

Veda C. Storey and Bernhard Thalheim. 2017. Conceptual modeling: Enhancement through semiotics. In *Proceedings of the International Conference on Conceptual Modeling*. Springer, Cham, 182–190.

Darijus Strasunskas and Stein L. Tomassen. 2008. The role of ontology in enhancing semantic searches: The EvOQS framework and its initial validation. *International Journal of Knowledge and Learning* 4, 4 (2008), 398–414.

Vijayan Sugumaran and Veda C. Storey. 2006. The role of domain ontologies in database design: An ontology management and conceptual modeling environment. *ACM Transactions on Database Systems (TODS)* 31, 3 (2006), 1064–1094.

Besiki Stvilia. 2007. A model for ontology quality evaluation. *First Monday* 12 (2007), 12.

Mari Carmen Suárez-Figueroa, Asuncion Gomez-Perez, and Mariano Fernández-Lopez. 2012. The neon methodology for ontology engineering. In *Ontology Engineering in a Networked World*. Springer, Berlin, 9–34.

Matt Taddy. 2015. Document classification by inversion of distributed language representations. *Arxiv Preprint ArXiv:1504.07295*.

Samir Tartir, I. Budak Arpinar, Michael Moore, Amit P. Sheth, and Boanerges Aleman-Meza. 2005. OntoQA: Metric-based ontology quality analysis. In *Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM'05) Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources*. https://corescholar.libraries.wright.edu/knoesis/660.

Samir Tartir, I. Budak Arpinar, and Amit P. Sheth. 2010. Ontological evaluation and validation. In *Theory and Applications of Ontology (TAO), Volume II: Ontology: The Information-Science Stance*, R. Poli (Ed.). Springer.

Olga Tatarintseva, Vadim Ermolayev, Brita Keller, and Wolf-Ekkehard Matzke. 2013. Quantifying ontology fitness in OntoElect using saturation- and vote-based metrics. In *International Conference on Information and Communication Technologies in Education, Research, and Industrial Applications*. Springer, 136–162.

Bernhard Thalheim. 2013. *Entity-Relationship Modeling: Foundations of Database Technology*. Springer Science & Business Media.

Andreas Tolk and James A. Muguira. 2003. The levels of conceptual interoperability model. In *Proceedings of the 2003 Fall Simulation Interoperability Workshop*. 7.

Mark Underwood, Michael Grüninger, Ken Baclawski, Mike Bennett, Gary Berg-Cross, Torsten Hahmann, Leo Obrst, and Ram Sriram. 2015 Internet of Things: Toward smart networked systems and societies. In *Ontology Summit 2015*. Retrieved from http://ontolog.cim3.net/OntologySummit/2015/.

Pierre-Yves Vandenbussche, Ghislain A. Atemezing, María Poveda-Villalón, and Bernard Vatant. 2017. Linked open vocabularies LOV: A gateway to reusable semantic vocabularies on the Web. *Semantic Web* 8, 3 (2017), 437–452.

Johanna Völker, Denny Vrandecic, York Sure, and Andreas Hotho. 2008. AEON–An approach to the automatic evaluation of ontologies. *Applied Ontology* 3, 1–2 (2008), 41–62.

Denny Vrandečić. 2009. Ontology evaluation. In *Handbook on Ontologies*. 293–313.

W3C-SWBPD. 2004. Semantic Web Best Practices and Deployment Working Group. Retrieved from http://www.w3.org/2001/sw/BestPractices/.

Wang Ye, Zhi Zhou, Shan Jin, Debin Liu, and Mi Lu. 2017. Comparisons and selections of features and classifiers for short text classification. *IOP Conference Series: Materials Science and Engineering* 261, 1 (2017), 012018.

Ron Weber. 2002. Ontological issues in accounting information systems. In *Researching Accounting as an Information Systems Discipline*, Sutton S, Arnold V, Eds. American Accounting Association, Sarosota, FL.

Stuart Weibel, John Kunze, Carl Lagoze, and Misha Wolf. 1998. Dublin core metadata for resource discovery. No. RFC 2413.

Patricia L. Whetzel, Natalya F. Noy, Nigam H. Shah, Paul R. Alexander, Csongor Nyulas, Tania Tudorache, and Mark A. Musen. 2011. BioPortal: Enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications. *Nucleic Acids Research* 39, suppl_2 (2011), 541–545.

Alexander Willner, Chrysa Papagianni, Mary Giatili, Paola Grosso, Mohamed Morsey, Yahya Al-Hazmi, and Ilya Baldin. 2015. The open-multinet upper ontology towards the semantic-based management of federated infrastructures. *EAI Endorsed Transactions on Scalable Information Systems* 2, 7 (2015), 1–10.

Zhe Yang, Dalu Zhang, and Chuan Ye. 2006. Evaluation metrics for ontology complexity and evolution analysis. In *e-Business Engineering (ICEBE '06)*. IEEE, 162–170.

Haining Yao, Anthony M. Orme, and Letha Etzkorn. 2005. Cohesion metrics for ontology design and application. *Journal of Computer Science* 1, 1 (2005), 107–113.

Wen Yao, Chao Chu, Akhil Kumar, and Zang Li. 2009. Using ontology to support context awareness in healthcare. In *Proceedings of the 19th Workshop on Information Technologies and Systems*.

Jonathan Yu, James A. Thom, and Audrey Tam. 2009. Requirements-oriented methodology for evaluating ontologies. *Information Systems*. 34, 8 (2009), 766–791.

Songmao Zhang and Olivier Bodenreider. 2006. Law and order: Assessing and enforcing compliance with ontological modeling principles in the foundational model of anatomy. *Computers in Biology and Medicine* 36, 7–8 (2006), 67-4-693.

Hongyu Zhang, Yuan-Fang Li, and Hee Beng Kuan Tan. 2010. Measuring design complexity of semantic web ontologies. *Journal of Systems and Software* 83, 5 (2010), 803–814.

Ying Zhao and George Karypis. 2002. *Comparison of Agglomerative and Partitional Document Clustering Algorithms*. Technical Report TR-02-014. Minnesota Univ. Minneapolis Dept. of Computer Science.

Hang Zhou, Yang Yang, and Hong-Bin Shen. 2017. Hum-mPLoc 3.0: Prediction enhancement of human protein subcellular localization through modeling the hidden correlations of gene ontology and functional domain features. *Bioinformatics* 33, 6 (2017), 843–853.

Pawel Ziemba, Jarosław Jankowski, Jarosław Wątróbski, and Jarosław Becker. 2015. Knowledge management in website quality evaluation domain. In *Computational Collective Intelligence*. Springer, 75–85.

Received May 2018; revised April 2019; accepted April 2019