12-14-2017

An Automated System for the Assessment and Ranking of Domain Ontologies

Melinda H. McDaniel
Georgia State University

AN AUTOMATED SYSTEM FOR THE ASSESSMENT AND RANKING
OF DOMAIN ONTOLOGIES

by

MELINDA MCDANIEL

Under the Direction of Veda C. Storey, PhD

ABSTRACT

As the number of intelligent software applications and the number of semantic web sites continue to expand, ontologies are needed to formalize shared terms. Often it is necessary to either find a previously used ontology for a particular purpose, or to develop a new one to meet a specific need. Because of the challenge involved in creating a new ontology from scratch, the latter option is often preferable. The ability of a user to select an appropriate, high-quality domain ontology from a set of available options would be most useful in knowledge engineering and in developing intelligent applications. Being able to assess an ontology's quality and suitability is also important when an ontology is developed from the beginning. These capabilities, however, require good quality assessment mechanisms as well as automated support when there are a large number of ontologies from which to make a selection.

This thesis provides an in-depth analysis of the current research in domain ontology evaluation, including the development of a taxonomy to categorize the numerous directions the research has taken. Based on the lessons learned by the literature review, an approach to automatic assessment of domain ontologies is selected and a suite of ontology quality assessment metrics grounded in semiotic theory are presented. The metrics are implemented in a Domain Ontology Rating System (DoORS), which is made available as an open source web application. An additional framework is developed that would incorporate this rating system as part of a larger system to find ontology libraries on the web, retrieve ontologies from them, and assess them to select the best ontology for a particular task. An empirical evaluation in four phases shows the usefulness of the work, including a more stringent evaluation of the metrics that assess how well an ontology fits its domain and how well an ontology is regarded within its community of users.

INDEX WORDS: Domain ontology, interoperability, metrics, ontology assessment, ontology evaluation, ranking, ontology, semiotics, semiotic ladder, semiotic layers

AN AUTOMATED SYSTEM FOR THE ASSESSMENT AND RANKING

OF DOMAIN ONTOLOGIES

by

MELINDA MCDANIEL

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2017

AN AUTOMATED SYSTEM FOR THE ASSESSMENT AND RANKING OF DOMAIN ONTOLOGIES

by

MELINDA MCDANIEL

Committee Chair:	Veda C. Storey
Committee:	Rajshekhar Sunderraman
	Vijayan Sugumaran
	Balasubramaniam Ramesh

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2017

ACKNOWLEDGEMENTS

I would like first to express my sincere gratitude to my advisor Dr. Veda Storey for the help she has given me with this research. She has taught me to bridge the gap between computer science and computer information systems smoothly, and has taught me what it means to be a researcher. She has also become a friend to me with her many words of encouragement when the work became challenging. Her vast knowledge and her unwavering energy have made me aim to always do my very best.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Vijayan Sugumaran, Dr. Raj Sunderraman, and Dr. Bala Ramesh, for sharing their knowledge with me and for their insightful suggestions.

My sincere thanks also goes to Dr. Mary Hudachek-Buswell, Dr. Olga Glebova, and Dr. Inna Rytsareva, three amazing computer scientists, who became my friends and inspired me by their kindness, enthusiasm, and the fact that they had traveled this path before me.

Last but not the least, I would like to thank my family: My husband, Craig, who helped me through many technical difficulties and taught me about version control, my parents who raised me with a love of learning, and my brother, Ed, who calls me "Professor" just because he knows how much I like it.

Thank you all very, very much. I could never have done it without you.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

As the number of interconnected systems has continued to grow, ontologies have been developed for a variety of domains (Ashburner et al., 2000), (Dos Santos et al., 2016), (Ramaprasad and Syn, 2013), (Welty, McGuinness, and Smith, 2004), (Yan, Ng, and Lim, 2002). A domain ontology is a conceptualization specific to a particular area of interest. Given the availability of such ontologies, one would expect that it would be a straightforward task to identify an ontology to fill a need without having to build a completely new one from scratch. Unfortunately, many of the ontologies found in ontology repositories and libraries are not of adequate quality (Obrst et al., 2007) because they often have syntactic or semantic flaws (Guarino and Welty, 2009). Others are free from errors but not comprehensive, flexible, or robust enough to be used for purposes other than the one(s) for which they were originally designed (Ma et al., 2014).

When searching for an ontology to employ in an application, it would be helpful to have an effective means of evaluating candidate ontologies that correspond to specific criteria and prioritizing the importance of various quality standards. Ideally, one should be able to enter domain terms into an online ontology repository and retrieve a list of potential ontologies that would meet the needs of an application with a corresponding appraisal of the quality of each. The question to be answered is whether it is possible to automatically assess the quality of domain ontologies to improve the selection process for use in tasks that require knowledge about the real world.

1.1 Research Objectives

Ontologies are challenging to build. Not only must the ontology contain no syntactic, semantic or cyclical errors, but its creators must have a well-defined understanding of the domain, logical connections and breadth of its intended use (Obrst et al., 2007). Poorly designed ontologies can cause problems for intelligent applications which use them. However, there is no general agreement on what the attributes are that should be considered when evaluating whether an ontology is of high quality. This lack of consensus regarding quality reduces the usefulness of ontologies (Gangemi et al., 2006a).

Metrics are valuable for assessing the quality of an ontology because they are one way in which an ontology can be assigned a numerical score for a specific attribute (Tartir et al., 2005). A comprehensive suite of metrics which encompasses a broad set of attributes is useful for calculating a quantifiable score representing an ontology's quality (Burton-Jones et al., 2005). This type of score is essential for selecting an ontology when multiple choices are available (Romano and McDonald, 2011).

Two challenges for automated ontology assessment are the lack of available ontology libraries (d'Aquin and Noy, 2012), (Obrst et al., 2007) and the lack of consistency within the few libraries that are available (Gulla, Strasunskas, and Tomassen, 2006), (Hoehndorf et al., 2011) for users to download and use. The repositories that do exist, such as BioPortal (Noy et al., 2009), allow users to upload ontologies to the library but accept these uploaded ontologies without evaluating their quality or format. As a result, the ontologies that are available in libraries are expressed in a variety of formats, which makes it difficult for a system to make an assessment.

The objectives of this research, therefore, are 1) to compose a complete domain ontology assessment suite, 2) to implement the suite in a prototype that will automatically assess some of the attributes of an ontology to assist in the selection of a domain ontology based upon quality criteria, and 3) to develop the prototype into a complete open source web application available for use and extension by others. The contribution of doing so is to partially automate the assessment of domain ontology quality to improve the selection process when an ontology is needed that requires knowledge about the real world, thereby progressing domain ontology use.

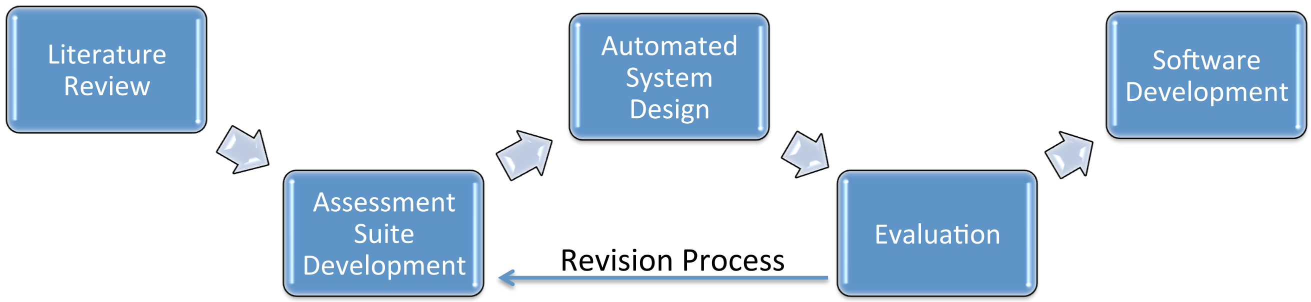

Figure (1.1) Thesis Development Process

1.2 Research Tasks

Tasks involved in carrying out this research include 1) a review of literature on ontology assessment to determine the best approach to automating ontology assessment 2) compiling a suite of ontology metrics corresponding the four layers of semiotic theory of signs 3) formalizing the assessment metrics 4) developing a system to automatically retrieve ontologies from a library and rank them 5) implementing the metrics in the ontology ranking system, 6) test the system by applying the ranking system to existing special-purpose and general-purpose ontology libraries, 7) building a complete open source web application and 8) exploring the use of the system in ontology applications.

The process of developing this research is based upon an iterative life cycle model for software development. The assessment suite has been developed and tested in a prototype system with multiple iterations to ensure that the metrics results are valid. Each time the metrics are updated, the testing is repeated. The full software system will not be completed until the metrics have been proven valid. Figure 1.1 shows a diagram for the system's development procedure emphasizing the repeated testing.

1.3 Thesis Overview

This thesis proceeds as follows. Chapter 2 introduces the notion of a domain ontology and provides an example. Chapter 3 discusses the need for, and challenges pertaining to, ontology assessment. Chapter 4 reviews related work in ontology quality evaluation categorizing them into a six differing approaches and introducing a domain ontology assessment classification scheme and a proposed framework for ontology evaluation using all six approaches. Chapter 5 proposes and formulates a suite of metrics based upon four layers from semiotic theory. A domain ontology ranking system is introduced in chapter 6, with an empirical evaluation in chapter 7. Chapters 8 and 9 focus more specifically on individual metrics from the suite and provide a framework for testing these. Chapter 10 summarizes the work that has been done and proposes future research expanding the system.

CHAPTER 2

DOMAIN ONTOLOGY CONCEPTS

In philosophy, the study of ontology deals with the nature of reality—exploring the similarities, differences and relationships between the types of entities that exist. Researchers in information systems and knowledge-based systems have expanded the definition so that the term ontology refers to, not only the vocabulary itself, but also the concepts the vocabulary is intended to express (Chandrasekaran, Josephson, and Benjamins, 1999). 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, Josephson, and Benjamins, 1999).

Ontological statements are expressed as triples, where two terms are connected by the relationship between them. Ontologies are represented in an ontology language such as OWL, an ontology language that adds formal semantics to the resource description framework (RDF) for representing relationships between entities on the web. In Figure 2.1, the line between Vegetarian Pizza and Pizza represents an *is-a* relationship. The line between Pizza and Pizza Base represents a *has-a* relationship.

The main purpose of a domain ontology is to create an agreement between two entities about what is intended by terms when they are used. It defines a common vocabulary for people or systems who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and the relationships between them (Noy and McGuinness, 2001). There are multiple reasons why an ontology would be needed. For example, 1) to make assumptions explicit, 2) to separate domain knowledge from implementation information, and 3) to allow for domain knowledge to be reused for other purposes (Noy, Fergerson, and Musen, 2000). Figure 2.2 illustrates how ontologies

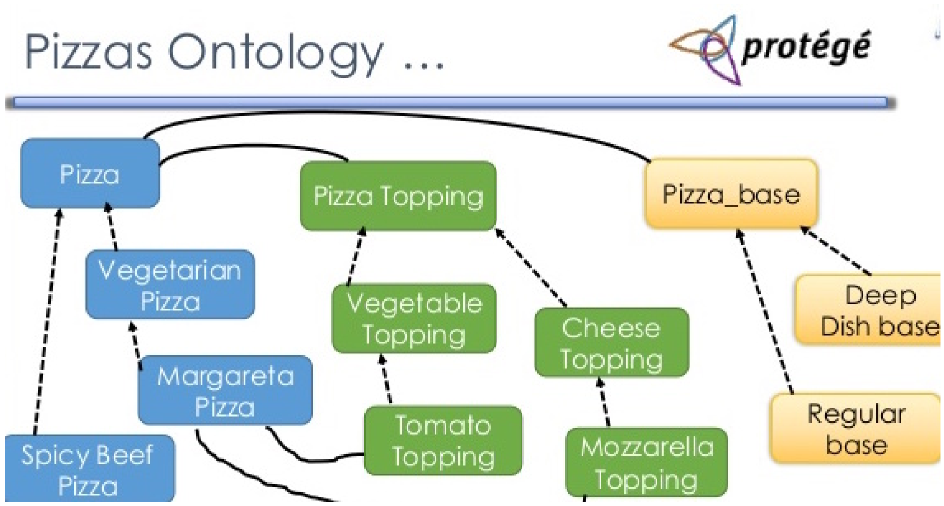

Figure (2.1) Relationships in the Stanford Pizza Ontology Example (Rector et al., 2004)

can formalize the commitment between languages, whether human spoken languages or computer languages (Guarino, 1998). The ideas of a conceptualization are clarified, leading to one or more ontologies. An ontology creates an agreed upon commitment and defines the intended meaning of the information being shared.

Domain ontologies can quickly become extensive and convoluted. Medical ontologies, in particular, often have hundreds of thousands of terms and relationships to express. It is essential to have a tool such as Protégé (Noy, Fergerson, and Musen, 2000) to help a user explore the concepts included. Figure 2.3 shows an example ontology displayed in the Protégé tool using the OntoGraf visualization plug-in to indicate relationships between terms (Falconer, 2010). The ontology shown is a simple pizza ontology developed by the Manchester OWL research group to teach OWL language concepts (Horridge and Bechhofer, 2009) and displays both inheritance relationships and data properties. Figure 2.3 shows a section of the .owl file containing a few ontological statements.

An ontology is developed to define a set of data and data structures for use by humans and software programs. Problem-solving methods, applications, and both human and software agents use ontologies as data to complete their goals (Noy and McGuinness, 2001). As the amount of data available today continues to grow, the need for ontologies

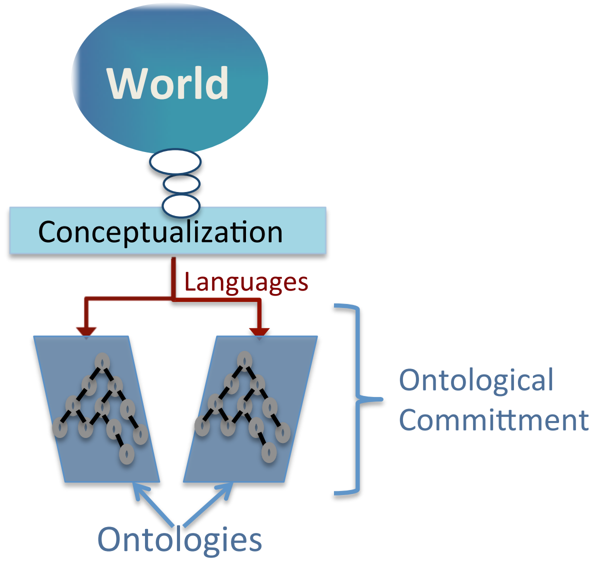

Figure (2.2) Ontological commitment made by two ontologies

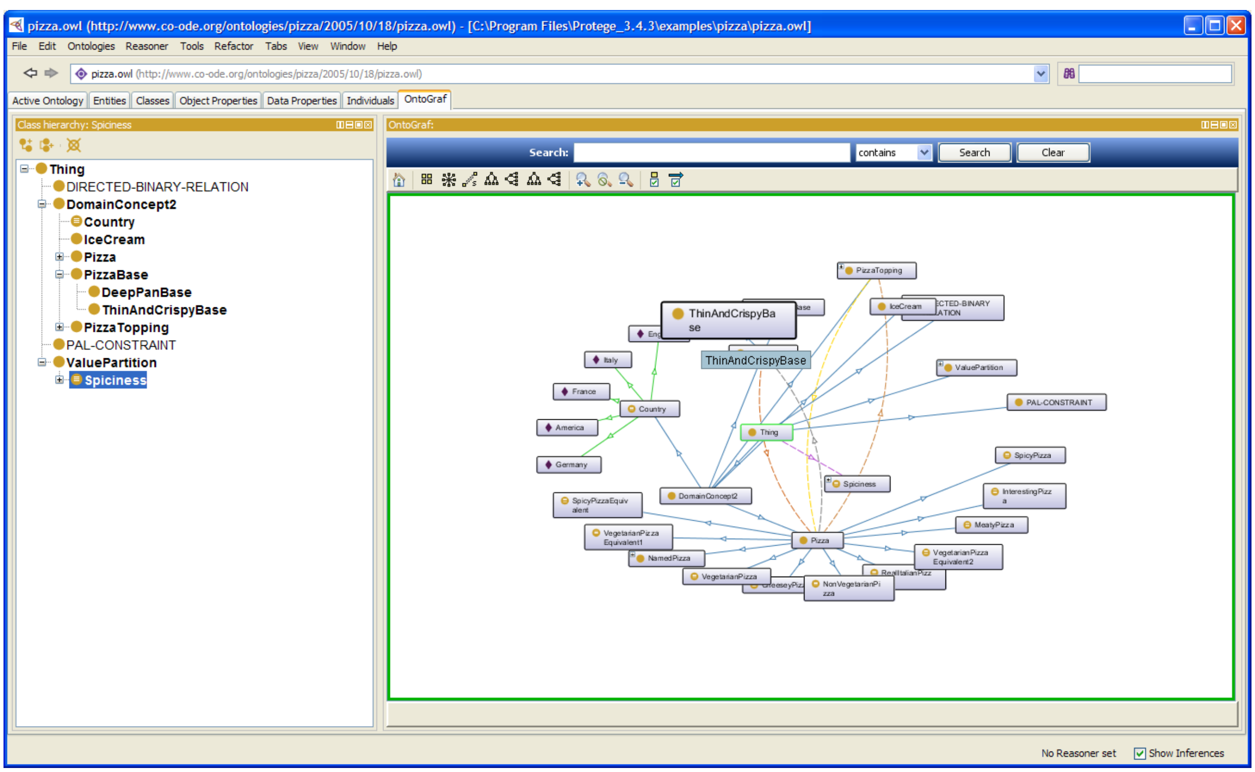

Figure (2.3) Pizza Ontology visualized in Protégé

```xml
<owl:Class rdf:about="#AnchoviesTopping">
  <rdfs:label xml:lang="pt">CoberturaDeAnchovies</rdfs:label>
  <rdfs:subClassOf rdf:resource="#FishTopping"/>
  <owl:disjointWith rdf:resource="#PrawnsTopping"/>
  <owl:disjointWith rdf:resource="#MixedSeafoodTopping"/>
</owl:Class>
<!--
   http://www.co-ode.org/ontologies/pizza/pizza.owl#ArtichokeTopping
-->
<owl:Class rdf:about="#ArtichokeTopping">
  <rdfs:label xml:lang="pt">CoberturaDeArtichoke</rdfs:label>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasSpiciness"/>
      <owl:someValuesFrom rdf:resource="#Mild"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#VegetableTopping"/>
  <owl:disjointWith rdf:resource="#MushroomTopping"/>
  <owl:disjointWith rdf:resource="#AsparagusTopping"/>
  <owl:disjointWith rdf:resource="#LeekTopping"/>
  <owl:disjointWith rdf:resource="#PetitPoisTopping"/>
  <owl:disjointWith rdf:resource="#OnionTopping"/>
  <owl:disjointWith rdf:resource="#SpinachTopping"/>
</owl:Class>
<!--
   http://www.co-ode.org/ontologies/pizza/pizza.owl#AsparagusTopping
-->
<owl:Class rdf:about="#AsparagusTopping">
  <rdfs:label xml:lang="pt">CoberturaDeAspargos</rdfs:label>
  <rdfs:subClassOf rdf:resource="#VegetableTopping"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasSpiciness"/>
      <owl:someValuesFrom rdf:resource="#Mild"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <owl:disjointWith rdf:resource="#PetitPoisTopping"/>
</owl:Class>
```


Figure (2.4) Portion of Pizza Ontology represented in OWL will continue to expand.

An example application of ontologies is the Semantic Web —an extension of the web in which the semantics of terms are explicitly defined using online ontologies (Berners-Lee, Hendler, and Lassila, 2001). The intention is for systems to be able to communicate with each other with little or no human intervention (Obrst et al., 2007), (Tolk and Muguira, 2003). Ontologies are crucial to this process because, for systems to communicate, they must share a common vocabulary (Romano and McDonald, 2011), but in order for these ontologies to be valuable, they must meet a defined standard of quality. Therefore there needs to be a way to assess their quality and a way to select the highest quality ontology when multiple options are available.

CHAPTER 3

MOTIVATION

The World Wide Web has grown to include many intelligent applications (Debattista, Auer, and Lange, 2016). The Semantic Web is an extension of the World Wide Web in which entities of the web share information and work together without dependence on human intervention (Berners-Lee and Kagal, 2008).

The Semantic Web requires the ability to express information in a precise, interpretable form, so that software sharing such data can also gain an understanding of the meaning of the terms describing the data (Hendler and Berners-Lee, 2010). Referred to as the third component of the Semantic Web (Berners-Lee, Hendler, and Lassila, 2001), ontologies are the means by which separate web components can share a common language and communicate to work together efficiently (Guarino, 1997a), contributing to automated reasoning (Romano, Horesh, and Dreyfuss, 2014). Interoperability between ontological resources is required to automatically analyze data across different data repositories and to allow for the automatic reasoning necessary for knowledge discovery (Noy and Musen, 1999).

3.1 Task Ontology Fit

Task ontology fit refers to determining the domain ontology that will produce the best results on a given task. If measuring task ontology fitness were a simple task, it would be a trivial task to find an ontology on the web and reuse it when a new purpose has been determined. Unfortunately, the calculation of which ontology would be the best depends on many factors relating to the task itself, the domain to be represented, and the users of the application (Guarino and Welty, 2002).

3.2 Ontology Development and Use Challenges

High-quality ontologies are difficult to build, because developing a good ontology requires, not only understanding of a particular domain, but also logic, reasoning, and clarity about the intended use of the data (Sugumaran and Storey, 2002), (Uschold, 2011). Over the past few decades many ontologies have been developed, both for specialized tasks and as upper ontologies, specifying a general vocabulary to be shared by multiple domains (Al-Yahya et al., 2010). Attempts have also been made to extract ontologies from existing Web documents (Alani et al., 2003). Since the quality of these ontologies varies greatly, though, prospective users cannot be certain about their clarity, coverage, consistency, or fitness for an intended purpose (Obrst et al., 2007).

Much work has been carried out that attempts to evaluate ontology quality at both high and low levels of detail (Duque-Ramos et al., 2011),(Gangemi et al., 2002), (Hlomani and Stacey, 2014), (Lozano-Tello and Gómez-Pérez, 2004), (Vrandečić, 2009). Attempts have been made to remove errors, (Guarino and Welty, 2002), (Poveda-Villalón, Gómez-Pérez, and Suárez-Figueroa, 2014), rank (Alani, Brewster, and Shadbolt, 2006), (Tartir et al., 2005), and assess them on a variety of aspects (Gangemi et al., 2006a), (García, JoseâĂŹGarcía-Peñalvo, and Therón, 2010). However, there is still a lack of consensus regarding ontology validation (García, JoseâĂŹGarcía-Peñalvo, and Therón, 2010), (Neuhaus et al., 2013), which slows down the transition of ontologies from simply a theoretical symbolic structure, to a reliable system component (Gangemi et al., 2006a).

3.3 Quality Assessment needs and requirements

Research has also been conducted on knowledge reuse, recognizing the need for a user to find and use an existing ontology rather than creating a new one for a specific purpose (Fernández et al., 2009), (Ma et al., 2014), (Normann and Kutz, 2010). This would be advantageous because there are difficulties related to ontology creation, whereas reuse has the potential to further data interoperability between heterogeneous information systems by allowing them to share a common ontology (Musen et al., 2012).

The Linked Open Vocabularies Project (LOV), for example, provides an extensive repository of ontologies containing, not only the ontologies themselves, but also useful information about them. The latter includes the links between the ontologies, their revision history, and author contact information (Vandenbussche et al., 2017). BioPortal (Whetzel et al., 2011) and OntoHub (Grüninger, 2013) are two other repositories providing ontologies and information about them to support knowledge reuse. These extensive domain ontology repositories make it even more difficult for a user to select the best ontology for a specific purpose due to the large number of choices available.

3.4 Need for improved quality assessment

To facilitate ontology selection, a system should be used to partially automate the selection of the best ontology for a specific purpose. This requires a means to assess ontology quality based upon a specified set of criteria (Brank, Grobelnik, and Mladenić, 2007). These criteria must be quantifiable so a system, rather than a human, can complete it. The use of assessment metrics contributes to interoperability between systems because of the ability to provide quantifiable information about attributes of an ontology. Metrics can be used to guide the development of a new ontology, and to assist a user in selecting an ontology from multiple choices (Tartir et al., 2005).

CHAPTER 4

LITERATURE REVIEW

Both theoretical and applied research efforts have recognized 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 (Gruber, 1993) proposed the definition of an ontology for practical use as "an explicit specification of a conceptualization," Dahlgren (Dahlgren, 1995) suggested a naïve approach to ontology development and Berners-Lee et al. (Berners-Lee, Hendler, and Lassila, 2001) called for the development of ontologies as an integral part of the Semantic Web.

There are two distinct ways of considering the ontology evaluation problem. The first, traditionally called "glass box" or "component" evaluation, examines the 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. 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 the 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. For this type of evaluation, it is essential to be able to identify criteria for measuring whether an ontology is suitable for a particular need (Brank, Mladenic, and Grobelnik, 2006). 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. Figure 4.1 summarizes relevant terms used in this thesis.

Term	Citation	Definition
Adaptability	[Vrandečić 2009]	Measures how well an ontology anticipates its future uses and whether it provides a secure foundation which is easily extended and flexible enough to react predictably to small internal changes
Alignment	[Obrst et al. 2006]	The process of evaluating 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 which 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 which 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]	Defined as a conceptualization specific to a particular domain
Essence	[Guarino and Welty 2002]	Refers to how essential the property is to an entity, and only includes properties which 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 intended users need 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. 2015]	Refers to an extension of Linked Data in which the data is expanded by providing definitions of the terms used within the data sets.
Ontology	[Gruber 1995]	Defined as an explicit specification of a conceptualization
Pragmatic	[Stamper et al. 2000]	Defined as the relationships between signs and their consequences
Pruning	[Maedche 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 in order 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
Semantic	[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

Figure (4.1) Domain ontology evaluation terminology

4.1 Ontology Assessment Approaches

Domain ontology evaluation initiatives have emerged over time as the field of domain ontology engineer has matured. Figure 4.2 highlights many of the important developments. Because there is such a large body of research projects with diverse approaches, it would be useful to organize the work to understand which approaches have been successful and which require additional work. From examining these efforts, four major approaches to ontology assessment can be identified: 1) divergent ideas such as efforts to remove errors; 2) efforts to modularize ontologies; 3) ways to score ontologies on specific attributes; and 4) ways to assess ontologies based on their fitness for a specific task.

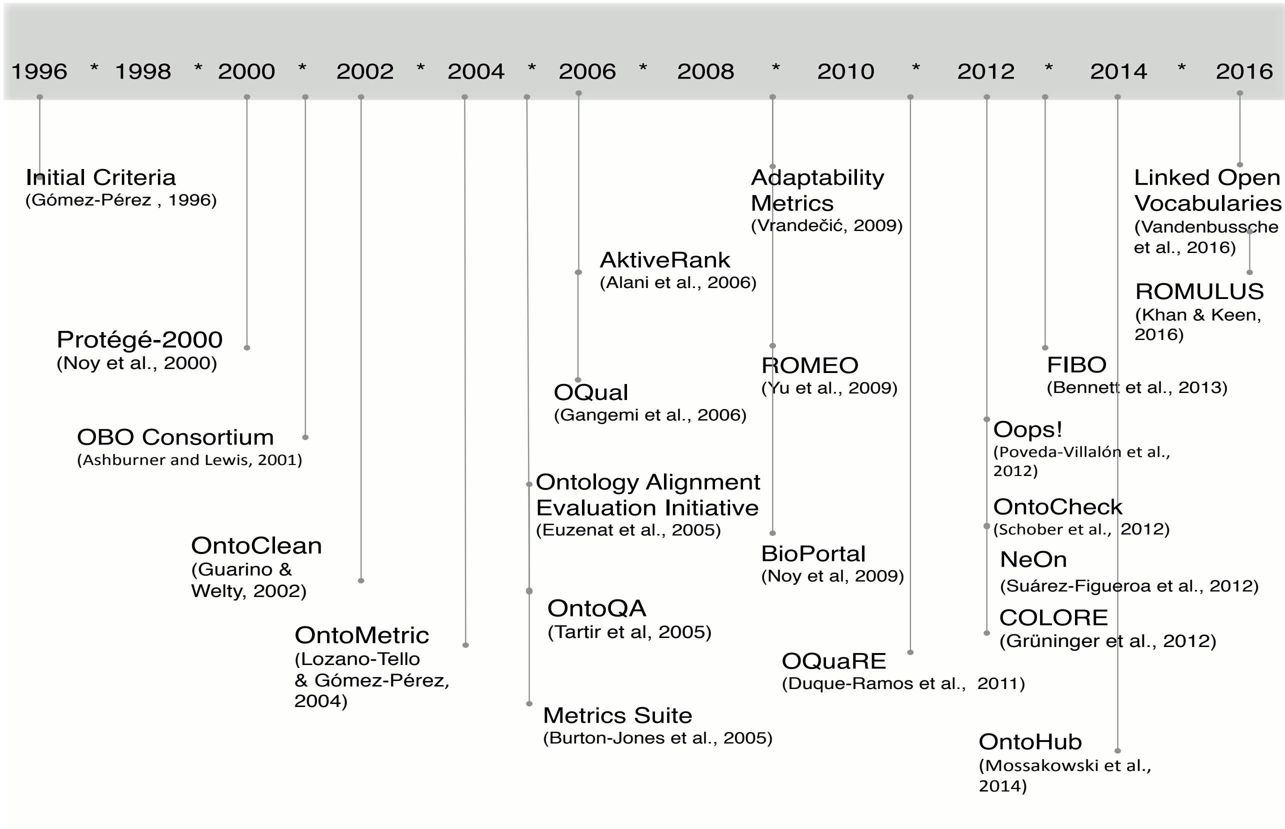

Figure (4.2) Ontology evaluation timeline

Discovering that ontologies can be used to promote interoperability between software systems, leads to the recognition that ontologies must be of a certain, acceptable quality. 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 the quality of the stored ontologies. Recent research has attempted to assess the ontology task 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 aligning ontologies with each other and modularizing them so their components can be assessed separately. Some research efforts have tried to combine approaches but research efforts have generally taken differentiated tactics for solving the ontology assessment problem. Therefore, a classification scheme would be useful for understanding how these different approaches to ontology evaluation differ, and whether there are points of congruence among them.

Research efforts on ontology evaluation over the last several decades have taken many different directions. Nevertheless, there have been occasional points of agreement and shared results. To identify areas of overlap, it would be useful to identify the main direction each effort has taken, as well as any minor directions.

I, therefore, reviewed over 150 research projects to identify the first, second, and in some cases, the third tactic employed by each. An agglomerative clustering algorithm (Zhao and Karypis, 2002) was used to identify similarities between the tactics.The stopping condition we used was when the number of groups became manageable and could be represented in a table. This clustering method was determined to be the most appropriate choice because our goal was to find similarities between research studies.

4.2 Clustering algorithm for category determination

Agglomerative hierarchical clustering is useful for combining large amounts of information into related clusters of information for more manageable examination. These algorithms determine the clusters by initially assigning each object to a separate cluster and then repeatedly merging pairs of clusters until an ending condition is reached (Zhao and Karypis, 2002). For our analysis, we repeatedly merged the clusters based upon similarity of keywords pertaining to the research study. The keywords were ranked based upon their level of importance to the study. For example, in Guarino and Welty (Guarino and Welty, 2009) research "An Overview of OntoClean", the most common concepts are validation, evaluation, consistency, and pitfalls. The Poveda-Villalon et al. (Poveda-Villalón, Gómez-Pérez, and Suárez-Figueroa, 2014) 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 of clustering, there were still over 40 clusters which was not manageable for a classification scheme. Therefore, for the second round, the keyword lists of each research study was expanded to include synonyms of the original terms in the keyword list. For example, the term pitfalls was expanded to include error, mistake, drawback, and difficulty. At this point, the research by Schober et al. (Schober et al., 2012), "OntoCheck: Verifying Ontology Naming Conventions and Metadata Completeness in Protege 4" 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 therefore the synonyms for each of the newly expanded keywords were added. After this round, six clusters emerged, a manageable number for a classification scheme. Figure 4.3 illustrates how the clustering methodology works for a small set of ontologies with keywords. For large numbers of ontologies, many more rounds would be necessary to achieve a manageable number of clusters.

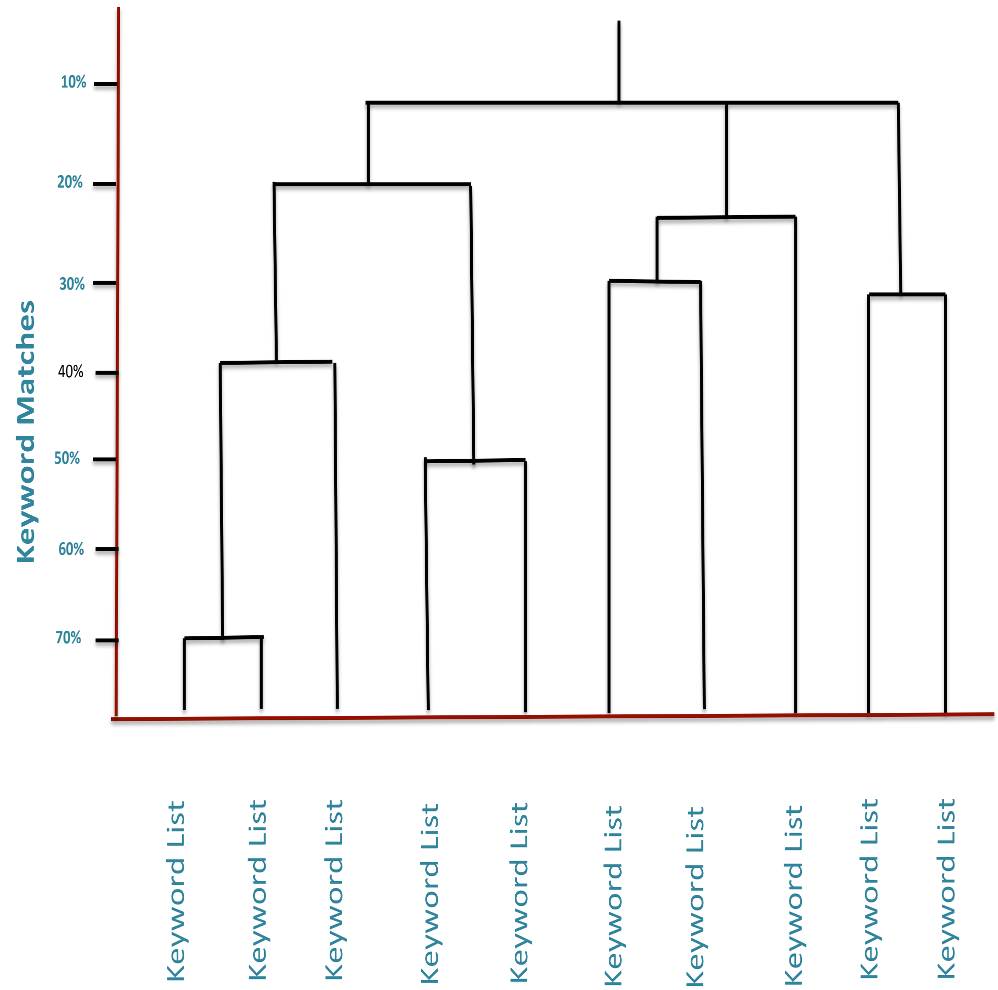

Figure (4.3) Sample of Agglomerative Clustering Algorithm

4.3 Classification Scheme

From the cluster analysis, six divergent approaches to ontology selection and evaluation emerged: 1) error checking and ontology cleaning; 2) alignment with a gold standard and ontology matching; 3) using metrics to quantify an ontology's quality; 4) modularization of ontologies to streamline the assessment task; 5) creating domain-specific libraries in which to store and maintain the ontologies; and 6) determining the task fitness of an ontology. Table 4.1 summarizes these approaches.

Table (4.1) **Classification Scheme**

Direction	**Focus**
Alignment	Determination of congruence between ontologies
Error Checking	Identification and removal of errors in ontologies
Libraries	Establishment of repositories sharing a common domain or language
Metrics	Development of metrics to assess ontology quality based on specific attributes
Modularization	Subdivision of ontologies into smaller modules with specific purposes
Task Fit	Identification of an ontology for a specific need

Most research focuses on a single method determine quality. For example, the Cyc ontology (Matuszek et al., 2006) and the Ontology Alignment Initiative (Cheatham et al., 2015) assess their alignment with other ontologies. OntoClean (Guarino and Welty, 2002) and Oops (Poveda-Villalón, Gómez-Pérez, and Suárez-Figueroa, 2014) assess the errors or lack or errors in each. Colore, Bioportal and OntoHub 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 determine an overall evaluation of ontology quality. The Financial Industry Business Ontology (FIBO) project (Fritzsche et al., 2017) and the Requirements Oriented Repository for Modular Ontologies (ROMULUS) (Khan, 2016)[both focus on the modularization of ontologies, splitting them into submodules that can be used separately. The NeOn project (Suárez-Figueroa, Gómez-Pérez, and Fernández-López, 2012) focuses on the goals required by the task for which an ontology is being developed and evaluates whether that ontology is meeting those goals.

Although the six approaches share some evaluation criteria, but each has a focus that is different from the others. Figure 2 illustrates representative studies for each of these six approaches.

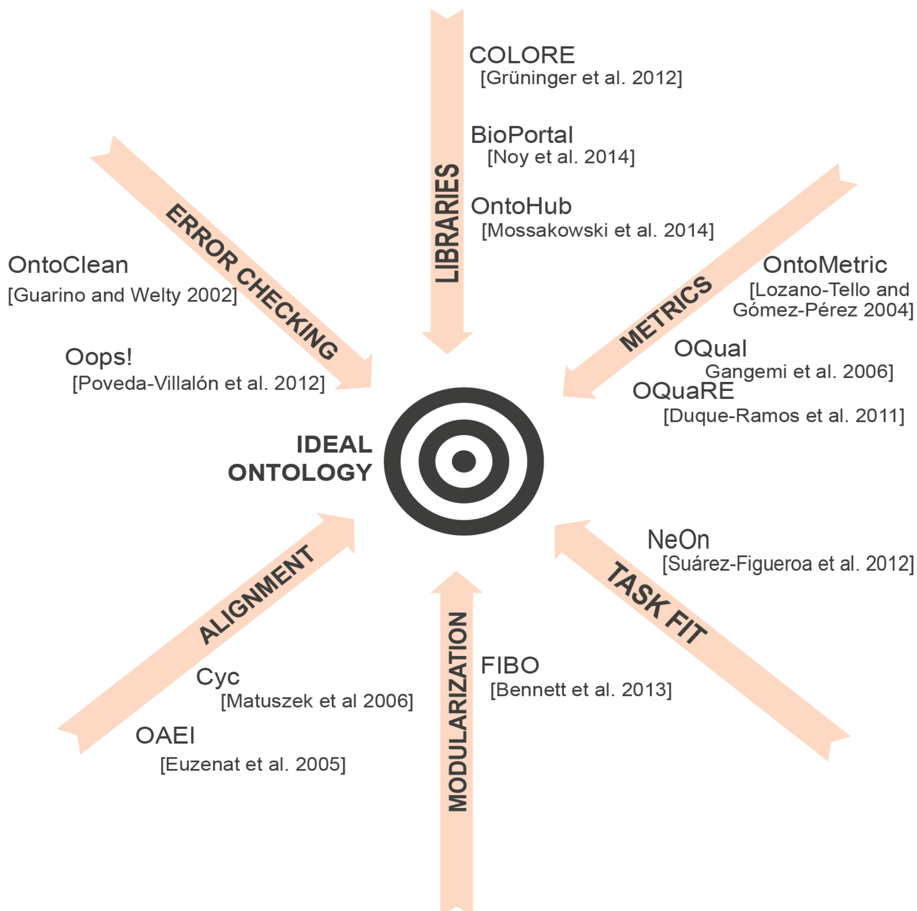

Figure (4.4) Ontology evaluation approaches

Each of the different approaches to ontology assessment research has the same objective, namely, selecting a high-quality ontology for a specific intended use. Each approach, however, has inherent advantages and disadvantages. These six directions serve a classification scheme to organize the enormous amount of work being carried out in ontology evaluation to establish a common foundation upon which to build further research. Each of the six approaches is outlined below and a table illustrating how the taxonomy can be applied follows.

4.3.1 Alignment approach

Ontology alignment refers to the comparison of two ontologies where one is the reference ontology, and the other is Ontology alignment refers to the comparison of two ontologies where one is the reference ontology, and the other, the candidate ontology. Ontology alignment is used to evaluate the quality of the candidate ontology if the quality of the reference ontology has already been established. If, for each concept in an ontology, there exists a corresponding concept in another ontology with the same intended meaning, then they are considered to be in alignment. Methods for matching the concepts include term comparison methods and structure-based comparison methods (Obrst et al., 2007). In the example in Figure 4.5, a concept matching approach is used to assign a similarity score to pairs of terms that are close in meaning (Abolhassani, Hariri, and Haeri, 2006).

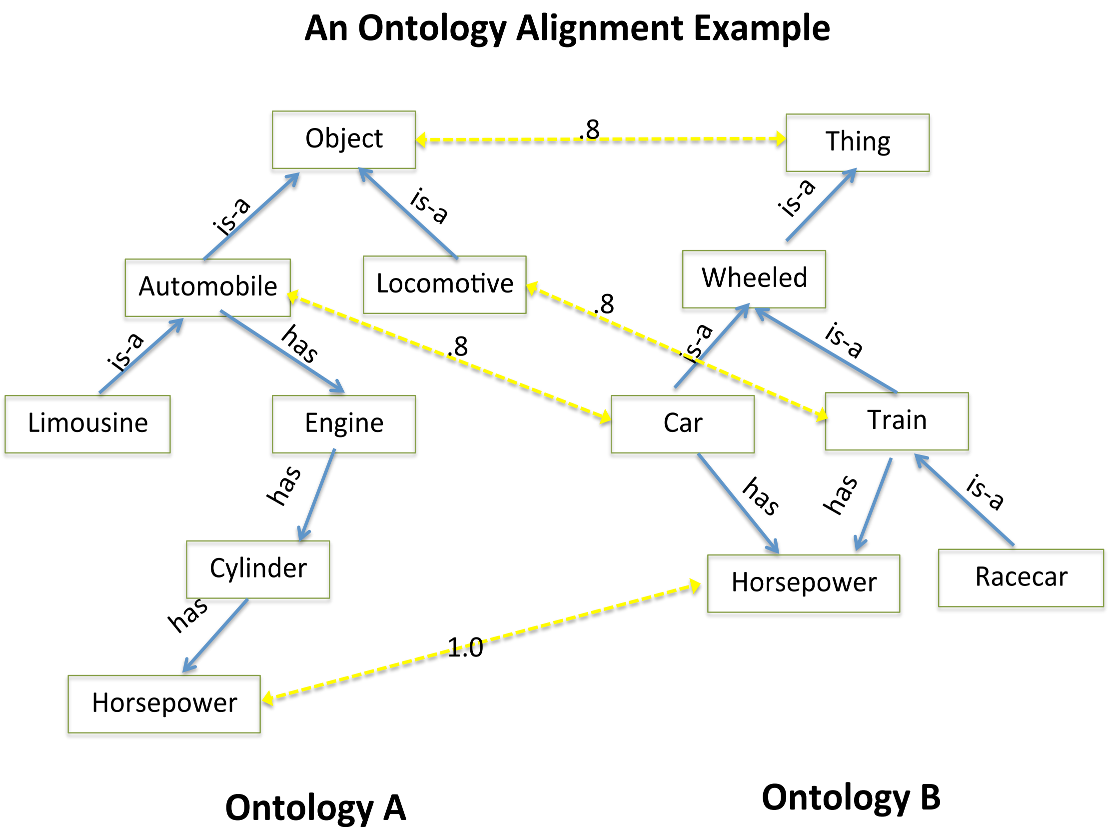

Figure (4.5) Ontology alignment example

Numerous tools, both automatic and semiautomatic, have been developed for ontology concept matching. Since the number of such tools increased rapidly, the Ontology Alignment Evaluation Initiative (OAEI) was established in 2004 to assess the quality of the results of these tools. OAEI is a coordinated international effort that provides predefined test cases as benchmarks to compare the performance of ontology matching tools and algorithms. An annual workshop and competition yields assessment data that can be analyzed, encouraging improved accuracy among the competing systems (Cheatham et al., 2015). Advantages: If an ontology can be shown to be in alignment with a gold-standard ontology, then the likelihood of it being of high quality is increased. If an ontology is in alignment with an upper ontology, such as Sumo (Niles and Pease, 2001) or

Cyc (Matuszek et al., 2006), the quality assessment is also likely to be high. Because of these factors, and the fact that it is possible to automate much of the matching process, a significant amount of research has been conducted to assess alignment (Abolhassani, Hariri, and Haeri, 2006)(Abolhassani 2006, Achichi 2016, Cheatham 2015, Shvaiko 2013). Challenges: The ability to use ontology alignment to evaluate the quality of an ontology has two significant challenges: 1) the lack of reference ontologies that meet the gold standard for comparison (Obrst 2006); and 2) the lack of consistent evaluation or proof of the matching algorithm itself (Shvaiko and Euzenat 2013). Because of these challenges, few researchers have taken this approach.

4.3.2 Error checking approach

Research has been conducted to identify errors and "clean" them (Guarino and Welty, 2002), (Gómez-Pérez, 2001). Error types range from simple syntax errors to complicated semantic and structural problems that might be difficult to identify. Figure 4.6 illustrates an example, from a construction domain ontology, of a cyclical error in which 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.

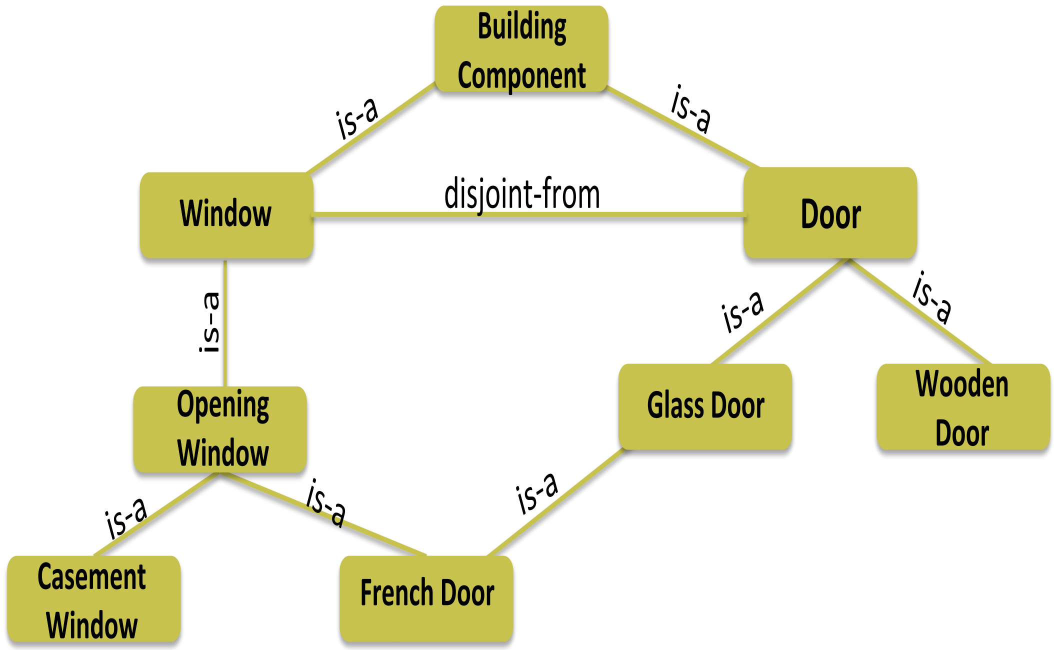

Figure (4.6) Building Construction Ontology with Consistency Problem

Early work on ontology cleaning evaluated ontologies to ensure they met certain basic requirements for validity. GómezPérez [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 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) to reduce costs and to improve interoperability between software systems.

Oops! (Poveda-Villalón et al. 2012) is a simple to use, web-based tool which provides automatic checking for common errors, such as naming conflicts or consistency problems to ontologies uploaded by users. The error list can be easily expanded to other types of errors. 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). Not all types of errors or potential errors are easily located by software, but the removal of common errors and structural problems from ontologies can be very 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 ontology selection problem. This method would need to be combined with other approaches to provide a valid selection.

4.3.3 Library approach

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 libraries store and maintain ontologies related to a specific domain, such as the BioPortal ontology for biomedical ontologies (Noy et al., 2009). Other libraries are multi-purpose allowing, not only domain ontologies from many different domains to reside there, but also high-level ontologies and other types of vocabularies or schema. Table 4.2 provides representative examples of ontology libraries and their objectives.

Table (4.2) **Library Examples**

Ontology Library	**Objective**
BioPortal	Search able 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)

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, that monitor the World Wide Web for newly published ontologies that match a particular 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 levels of quality (Buitelaar, 2005). The Open Biomedical Ontologies (OBO) consortium (Ashburner et al., 2000) 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 that it also includes the OBO foundry, which serves as a repository for biomedical ontologies designed using the established guidelines (Ashburner et al., 2000). An extension to ontology libraries, are 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 included. 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., 2017).

Advantages: Assessing ontology quality within a 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 (Gruninger and Katsumi 2014), most ontology libraries are domain-specific. One of the largest of these is BioPortal, `http://www.sharelatex.com`, an open repository of biomedical ontologies. It provides access to existing ontologies, has the capability for the user to add new ontologies, add notes, contribute mappings between terms, and reviews ontologies based on criteria such as usability, coverage of the domain, accuracy, and level of documentation available. BioPortal also includes a recommender system that provides users with a list of ontologies that match a particular domain assessing their quality based on domain coverage, community acceptance, detail of knowledge and amount of specialization (Noy et al., 2008).

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. This lack of consistency, though, creates challenges for carrying out ontology evaluation within the libraries. Another challenge is the lack of general-purpose ontology libraries. Also beneficial would be the establishment of more libraries with consistent languages and(d'Aquin and Noy, 2012).

4.3.4 Metric-based approach

Objective rather than subjective measurements must be used in order for ontology assessment to be carried out by software rather than by humans. In particular, it is desirable to have numerical scores based on objective information about an ontology. Unfortunately, no consensus has yet been reached about which attributes of an ontology correlate to high quality for all domains and applications.

Gómez-Pérez (1996) identified clarity, consistency, conciseness, expandability and correctness as being among the most important qualities an ontology must possess. These qualities relate to the verification of the definitions and axioms that are explicitly included in the ontology as well as those that can be inferred. These basic attributes have continued to appear in many evaluation frameworks and tools developed over the past twenty years (Vrandečić, 2009), (Lozano-Tello and Gómez-Pérez, 2004) and (Gomez-Perez, Fernández-López, and Corcho, 2006),(Völker et al., 2008).

Metric-based techniques provide a way to quantify the quality of a given attribute of an ontology (Tartir et al., 2005). A numerical score for an ontology's quality can assist a user in making 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. It might also be feasible to automate the calculation of such a score (Obrst et al., 2007), (Alani, Brewster, and Shadbolt, 2006).

Accurate, well-defined, and easy-to-apply metrics are needed for systematic evaluation of ontologies. Metrics can be used to evaluate the quality of a particular ontology or to compare ontologies when multiple candidates fit particular requirements (Tartir et al., 2005). Metrics, rather than measuring an ontology as simply effective or ineffective, can describe a specific aspect of an ontology (Tartir et al., 2005). For a given domain, some attributes of an ontology might be more significant than others, with suites of metrics giving the user the ability to weigh each aspect differently. Ontology assessment metrics have been developed that evaluate: 1] the overall quality of an ontology using multiple attributes, or 2] only one attribute of a high quality metric or a small set of related attributes. Specific approaches to quality assessment are reviewed below, selected for their alignment with the 2013 Summit results (Neuhaus et al., 2013) and established, earlier ontology evaluation research.

Several studies attempt to provide a broad assessment of an ontology's quality from multiple perspectives. Lozano-Tello and Gómez-Pérez (Lozano-Tello and Gómez-Pérez,

2004) developed a hierarchical framework, OntoMetric, which identifies 160 characteristics along five dimensions to evaluate the quality of an ontology and its suitability to a user's requirements (Lozano-Tello and Gómez-Pérez, 2004). The five dimensions relate to the content of the ontology, the language used, the method of development, the building tools, and the associated costs. This effort corresponds to software as part of a larger system (Neuhaus et al., 2013).

The semiotic metrics suite by Burton-Jones et al. (Burton-Jones et al., 2005) is built upon the semiotic framework for sign quality assessment developed by Stamper et al. (Stamper et al., 2000), that measures whether symbols are good or bad, clear or unclear. Ontologies use symbols to describe terms and the relationships among them, making semiotic theory an appropriate measurement tool.

OQual (Gangemi et al., 2006a) evaluates ontologies on three dimensions: 1) structural (the syntax and semantics of an ontology); 2) functional (the relationship between the ontology and its intended task); 3) and usability (the annotation context of the ontology). OQual, itself, is an ontology of ontology validation which can be used to determine the best set of criteria for choosing an ontology for a given task.

Vrandečić (Vrandečić, 2009) proposes a list of eight quality criteria, compiled from prior research, that could to be used as goals to guide the development and evaluation of an ontology. These eight categories are accuracy, adaptability, clarity, completeness, computational efficiency, conciseness, consistency, and organizational fitness. He recognized, however, that "None of them can be directly measured, and most of them cannot be perfectly achieved" [Vrandečić 2009, page 53].

OntoQA (Tartir, Arpinar, and Sheth, 2010) separates its metrics into two separate 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. Knowledgebase metrics check whether a populated ontology is a rich and accurate representation of the real world by evaluating its content. Together this set of metrics can assist a user in determining whether a particular ontology would be suitable to meet individual needs.

OQuaRE (Duque-Ramos et al., 2011)provides both a model and set of metrics to assess ontologies based upon established standards for software quality evaluation. Acknowledging that an ontology can be thought of as 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 sub-characteristics each with a metric provided for determining its value.

Because of the quantifiable nature of metrics, much work has been done on 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 being used, even though, as discussed above, there is no generally accepted procedure for doing so (Neuhaus et al. 2013). The main objective is to design, implement, and apply a valid and useful metrics suite, but "proving" that the metrics used are valid and appropriate requires, most likely, repeated, empirical studies.

4.3.5 Modularization Approach

The idea of modularization of ontologies is derived from software engineering where it 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 is modular in nature, is easier to understand and apply— especially if more than one person is involved in the software's development and use. This advantage of modular software is especially significant in the case of 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. This ability to reuse parts of an ontology is only possible if the portions are truly completely separate modules that are able to be extracted without loss of meaning (Khan and Keet, 2016). Therefore, it is important that domain ontologies be created in such a way that there are extractable parts of an ontology that can be reused outside the original context of the complete ontology. The use of individual parts, or modules, of an ontology can only be done effectively if the modules are of high quality, fit the intended task, and can be used separately 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 as a way to evaluate the overall modularization of an ontology (Kutz and Hois, 2012).

The concept of pruning, which refers to reducing the size of an ontology or module by removing elements outside of a specific application domain, is closely linked to modularization. The goal of pruning is to create a balance between the completeness and preciseness of the ontology. Attempting to create 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 the domain that is too narrowly focused could lead to an ontology that is limited in expressiveness. Figure 4.7 illustrates the process of pruning an ontology.

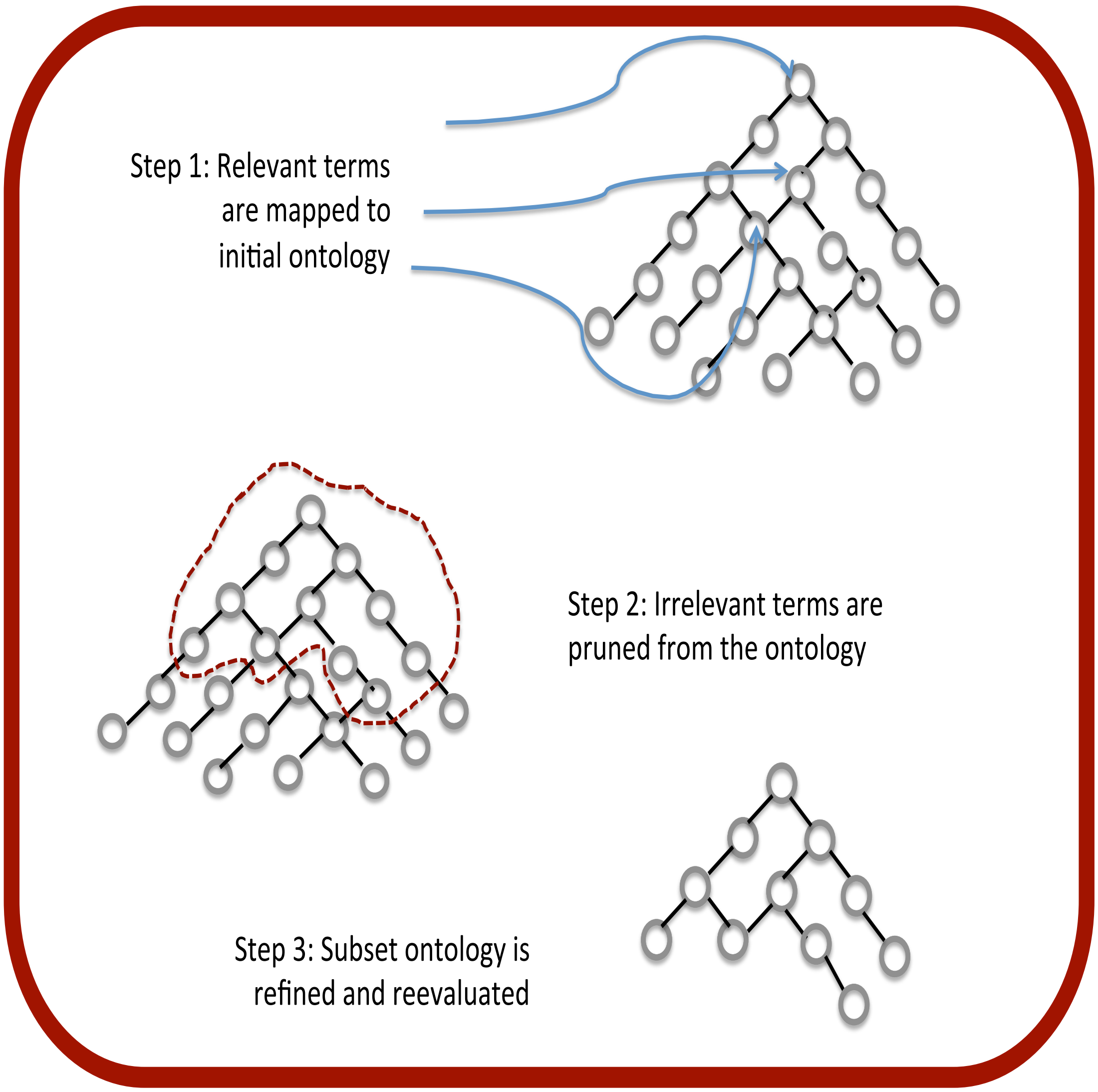

Figure (4.7) Refining and pruning an ontology for a specific domain

The goal of pruning is to create a single ontology or set of ontology modules that provide a rich conceptualization of the target domain, but exclude any parts that are outside of the specific focus (Maedche and Staab, 2004). A system to measure how closely this balance is attained should be a part of any broad ontology assessment.

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

Creating ontologies from more specific, focused and self-contained modules would go a long way toward improving ontology evolution and reuse. The combination and extension of smaller modules and patterns could lead to the formation of larger ontologies containing only the most relevant information for the domain. Less human intervention would be necessary if the individual modules were drawn from ontologies in which usefulness and quality had already been proven. However, the very act of extracting modules must be shown to not subtract from the syntactic, semantic and pragmatic quality of the individual modules. Further ontology quality assessment would need to be done to determine whether this is indeed the case.

4.3.6 Task-fit Approach

Task fit refers to the evaluation or improvement of a given ontology in relation to its performance on a specific set of tasks. The field of software requirements engineering has included research on determining the goal of the software first rather than on the software and how it can achieve the goal. This approach to software creation can also be applied to ontology creation and selection. If it is known what task the ontology will be used for, we can choose or build an ontology specific to that goal. For this method to work, however, it is necessary that ontology engineers have a precisely defined and realistically

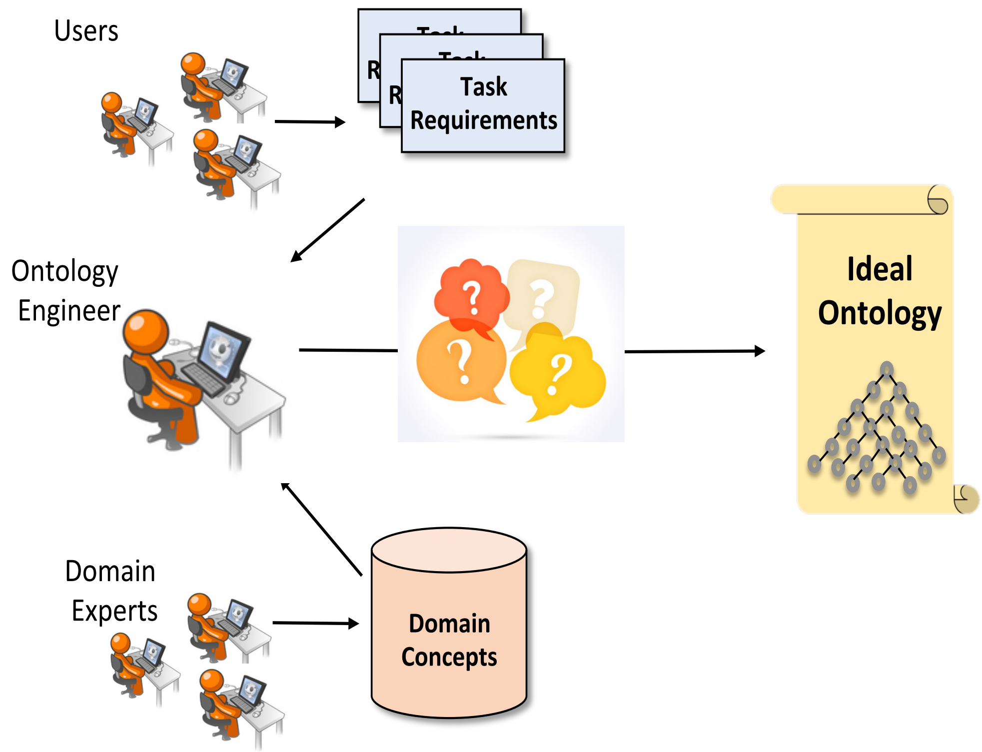

Figure (4.8) Creating an ontology for a specific task requires experts in the domain and users of the task.

achievable goal. Much research has gone into solving the task to domain fitness process, but the question still remains as illustrated in Figure 4.8.

Another approach is to evaluate a preexisting ontology's fitness related to a specific goal, determining a measure of how effective the ontology is to achieve the goal. This effectiveness can be quantified only if there is a measurable way to assess the performance of the ontology for the given task (Porzel and Malaka, 2004) Porzel and Malaka created a methodology for evaluating ontology performance on given tasks and augmenting them to better fit the task requirements.

Brewster et al. (Brewster et al., 2004) developed data-driven techniques for ontology assessment. Their method assesses how well a particular ontology fits a given corpus by examining the internal structure of the ontology. By determining how closely the terms of the corpus are clustered together in the ontology, a measurement can be determined of the ontology's fitness.

More recently, the Neon methodology for ontology creation has been presented that also takes into consideration specific goals when an ontology is being developed, as well as taking into consideration the input, output and specific constraints of a task (Suárez-Figueroa, Gómez-Pérez, and Fernández-López, 2012). The NeOn methodology includes procedures for ontology selection, reuse and re-engineering to fit a particular goal. Each process that is part of the Neon framework includes ontology assessment to ensure that the goals are being met.

Automation of domain ontology selection is needed if machines are going to be capable of selecting ontologies for complex information systems (Obrst et al., 2007). For ontology-task selection to be carried out by a machine, a machine needs to be able to calculate a metric that is an aggregate number, representing an overall quality evaluation. The problem of assigning a value for the fitness metric 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, 2010). This type of concept matching is an over-simplification of the complex nature of matching a domain to a particular ontology. The relationship between an ontology and a conceptualization is dependent on the agent that conceives the conceptualization and on the means by which it is encoded. Therefore, at best, a fitness measurement can only be an approximation (Gangemi et al., 2006b).

4.3.7 Combined approaches

To overcome the challenges pertinent to each individual approach and to improve the effectiveness of the assessment several research efforts have attempted to combine two or more approaches. Advancements have been made by adding metrics to modularity methods, to ontology libraries and to task fit assessments. There has also been progress made in adding error checking and metrics to ontology building (Noy and McGuinness,

2001) and to ontology matching tools (Paulheim, Hertling, and Ritze, 2013). The Combinations column of the classification scheme shown in Figure 4.9 outlines examples of methodologies in which two approaches together contribute a more accurate assessment of ontology quality than one alone would be able to provide. Further combinations are still needed to utilize the advantages of each approach as outlined in the Advantages column of Table 4.9, and to overcome the challenges outlined in the Challenges column. It is observable in the classification scheme that there are several approaches that have not been combined with any another approaches.

Figure (4.9) Application of the Domain Ontology Assessment Taxonomy

Approach Taken	Naïve efforts	Advanced efforts	Combinations	Advantages	Challenges
Alignment	Simple comparison to a gold standard [Grefenstette 1994]	Ontology Alignment Evaluation Initiative [Achichi et al. 2016]	**Lacking**	Automation possibilities [Noy and Musen 1999]	Lack of gold standard reference ontologies [Obrst 2006]
Error Checking	Basic syntax checking [Gómez-Pérez 1996] OntoClean [Guarino and Welty 2002]	Advanced semantic checking Oops! [Poveda-Villalón et al. 2012]	Added to a development tool [Musen, 2015]	Easy to apply if requirements are stringent [Solskinnsbakk et al. 2012]	Identification of the error's level of urgency [Gómez-Pérez, 2004]
Libraries	Domain specific libraries BioPortal [Noy et al. 2009]	General purpose libraries Linked Open Vocabularies [Vandenbussche et al. 2015] OntoHub [Mossakowski et al. 2014] COLORE [Grüninger et al. 2012]	Added metrics to libraries BioPortal [Noy et. al. 2009]	Increases likelihood of finding an appropriate ontology	Lack of consistency in ontology quality within the library [Burton-Jones et al. 2005]
Metric-based	Individual metrics [Yang et al. 2006] [Orme et al. 2006] [Fernández et al. 2009] [Ma et al. 2010] [Ouyang et al. 2011]	Metrics suites OQuaRE [Duque-Ramos et al. 2011] OntoQA [Tartir et al. 2005] AKTiveRank [Alani et al. 2006] [Burton-Jones et al. 2005] OntoMetric [Lozano-Tello and Gómez-Pérez 2004]	Added to a development tool [Noy et al. 2001]	Quantifiable [Tartir et al. 2005]	Proving the validity of the metrics [Kitchenham et al., 1995]
Modularity	Modular ontologies FIBO project [Bennett et al. 2013]	Framework for ontology modularization [Khan and Keet 2015]	Added metrics to assess modularity Cohesion [Yao et al. 2005] Coupling [Orme et al., 2006]	Smaller pieces are easier to apply [Grau et al. 2006]	Extracting the individual modules [Kutz and Hois 2012]
Task Fit	Requirement matching [Porzel and Malaka 2004] ROMEO [Yu et al. 2009]	Complete goal and assessment toolkit NeOn [Suárez-Figueroa et al. 2012]	**Lacking**	Ultimate goal of a domain ontology [Kim and Storey 2012]	Mapping an ontology to a conceptualization [Gangemi et al. 2006]

4.3.8 Combination Examples

There have been efforts made to create metrics that measure the quality of the individual modules when a modular approach is employed. The attributes of cohesion, coupling, and complexity have been the focus of much research due to their contribution to determining whether an ontology is easily adapted for a new purpose (Orme et al. 2006; Yang et al. 2006; H. Yao et al. 2005; Ouyang et al. 2011). H. Yao et al. (2005), for example, established cohesion as a fundamental characteristic of an ontology and proposed a set of metrics to measure it for modular ontologies. Oh et al. (2011) added additional ontology modularity metrics to find a way to quantify the relationship between entities in an ontology. Ma et al. (2011) proposed four additional ontology cohesion metrics based upon ontological semantics, and Orme et al. (2006) defined coupling metrics based upon commonly accepted software engineering measurements, which assess how well the modules work together in systems of ontologies.

One excellent use 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 developing or modifying ontologies [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.

The ROMEO methodology combines the use of metrics with task fit to match the task for which the ontology will be needed to specific metrics that evaluate the ontology's suitability to that application. The methodology consists of three steps: asking the user questions about the task, mapping the task to specific metrics, and assessing the ontology's quality based on the specific metrics identified as relevant (Yu, 2011).

4.3.9 Evaluation throughout the life cycle

Gómez-Pérez [1996] was the first to acknowledge that 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 verification of complete, consistent and concise definitions at all stages of the ontological development process [Gómez-Pérez, 1996]. At least one of the ontology assessment approaches should be applied at each step of the process to ensure 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, 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 which 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 4.10 illustrates how this cycle should occur. Combining domain ontology assessment throughout the life cycle with more exploration of the combinations of approaches as described later in this chapter, would do much to progress the field of domain ontology assessment.

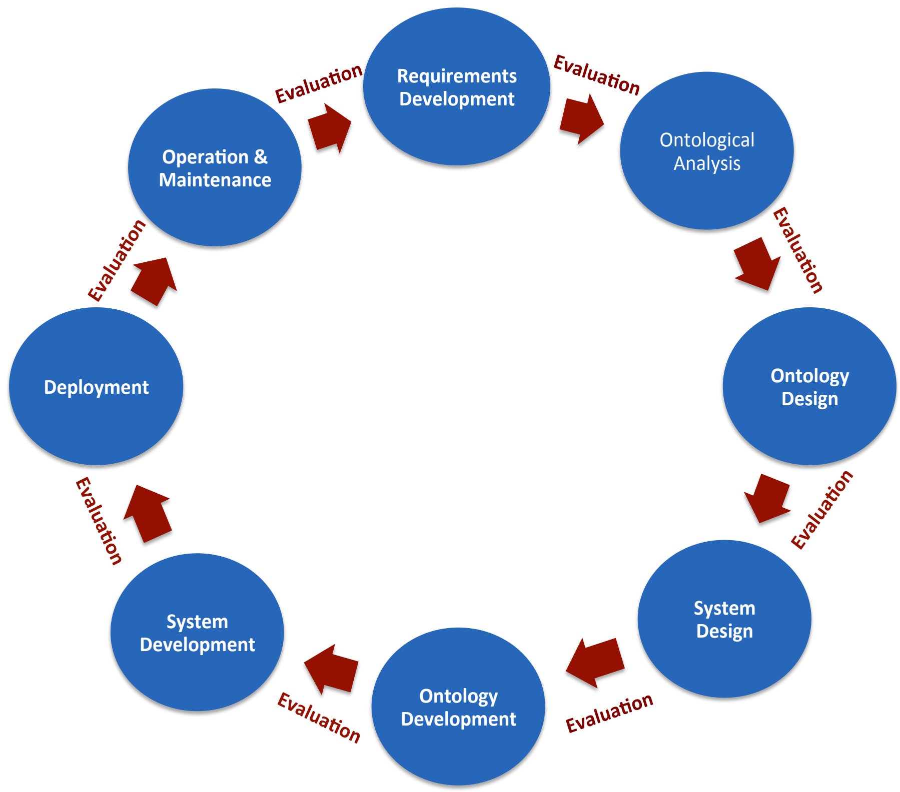

Figure (4.10) Ontology Life Cycle Model for Ontology Evaluation - adapted from (Neuhaus et al. 2013)

4.4 Proposed Solution to the Ontology Selection Problem

From the above, it is clear that, over the last few decades, a vast amount of research has focused on finding ways to evaluate ontologies based on their quality and suitability for a particular task, with encouraging results emerging. However, problems related to ontology evaluation still 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; including both an overall evaluation and a specific evaluation dependent on task and domain requirements.

Selecting the best ontology for a specific task requires that the ontology is free of errors, modular in nature, and stored in an ontology repository where it can be easily

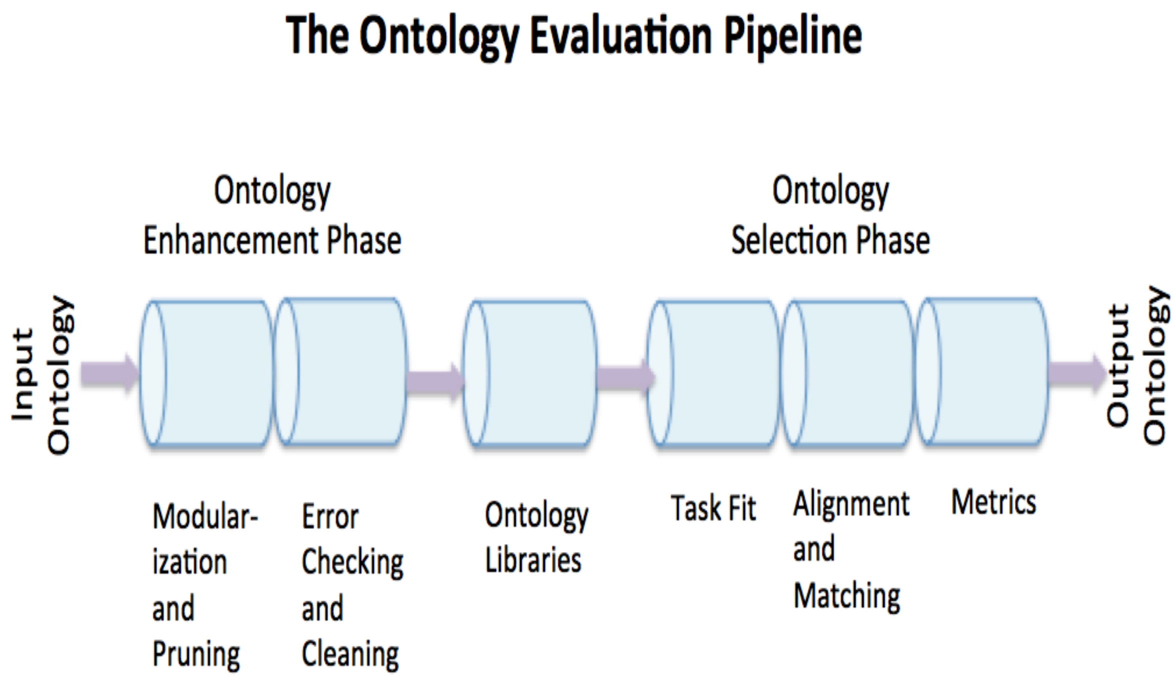

Figure (4.11) Pipeline Framework for Ontology Assessment

found. It also requires that the ontology scores high on specific attributes, that it aligns well with the required domain, and that it fits the task for which it is needed. These requirements can best be addressed if existing research efforts are combined. This combination of work to be carried out can be represented by the sequential pipeline in Figure 6, where the ontology is the input, and the resulting output, an ontology for a specific purpose. This pipeline framework envisions work on ontology selection as containing two distinct phases. The first is to ascertain that an ontology is of high quality before its placement into an ontology repository so it can be easily located. The second phase ensures that any ontology selected from within an ontology repository is appropriate for the needed task and domain. Figure 4.11 illustrates the Pipeline Framework for Ontology Assessment and identifies the research approaches that contribute to each of the framework's phases.

4.4.1 Ontology Enhancement Phase

Although many ontologies created by novice ontology developers or automatically created by software are problematic, they still contain valuable information. This phase of the ontology evaluation process, therefore, tries to improve ontologies based on findings in the research areas of ontology cleaning (to remove errors), and ontology modularizing (so that ontologies can be subdivided into more easily employed pieces). These tasks need to be completed before the ontologies can be placed into libraries. The order in which they are performed can vary depending on the ontology that is input. For example, if the ontology is unable to be read, the error checking and cleaning phase would need to be done before the modularization process takes place. The final task of this phase is to select an appropriate, well-maintained ontology repository in which to place the ontology so it is easily accessible for future use.

4.4.2 Ontology Selection Phase

Selecting an ontology for a specific use requires that the ontology be: 1) fit for the intended task; 2) in alignment with the intended domain; and 3) of high quality. Therefore, this phase of the ontology evaluation pipeline is devoted to helping the user make these choices. The ontology repositories should be of high quality, be error free and contain modular ontologies. The user, therefore, needs to make a decision based on domain and task requirements. Some research efforts have found ways to assess an ontology's fitness for a particular task (Porzel and Malaka 2004, Suarez-Figueroa et al. 2012). Other research efforts have provided ways to determine whether an ontology is in alignment with a specific domain (Achichi et al., 2016). Still other research efforts have created metrics to assess an ontology's quality on multiple attributes (Duque-Ramos et al. 2011, Gangemi et al. 2006, Gomez-Perez 2004). This phase of the assessment pipeline should combine all three of these methods to make the best ontology selection decision. The order in which the selection is made is dependent upon the task for which the ontology will ultimately be used, therefore the assessment for an ontology's task fitness should be the first assessment done for this phase of the ontology selection process. However, as with the first phase of the pipeline framework, the order in which these assessments are done could vary depending on a user's needs.

4.4.2.1 Pipeline Applicability The two phases of the ontology evaluation process (ontology enhancement and ontology selection), as well as the ontology library that stores the ontologies, form a sequential pipeline. The pipeline is an appropriate way to ensure that all relevant criteria for the selection of an ontology are included in the decision-making process.

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. What is essential is for ontology engineers to recognize that one approach to ontology evaluation is not sufficient for a complete ontology assessment, but that by combining methods the accuracy of the results can be improved.

CHAPTER 5

DOMAIN ONTOLOGY ASSESSMENT SUITE

The identification of attributes to include in an ontology assessment or selection framework is challenging. Gómez-Pérez (Gómez-Pérez, 1996), for example, initially identified a set of characteristics that an ontology should possess to be useful including: clarity, consistency, conciseness, expandability and correctness (Gómez-Pérez, 1996). Other work has expanded this list to include craftsmanship, modularity, stability and many others (Neuhaus et al., 2013), (Orme, Tao, and Etzkorn, 2006), (Orme, Yao, and Etzkorn, 2007), (Rector, 2002).

A suite of metrics to assess the quality of an ontology should include applicable, recognized quality attributes that, together, give a complete appraisal without redundancy. In prior research, many attributes have been shown to be valid, although some are not easily obtained without human intervention, and many have overlapping definitions. Table 5.1 identifies noteworthy candidate attributes for ontology quality assessment.

Although a large number of individual assessment metrics and metrics suites have been developed for assessing the quality of ontologies, very few have been implemented. Although researchers have performed case studies to prove the effectiveness of their metrics, the frameworks they employ are often not publicly available. Notable exceptions are: Protégé [46] and OntoQA [73], which are available for download, and Oops! [57], which has a web interface.

The most commonly employed tool for creating an ontology is the Protégé development tool parencitemusen2015protege. The Protégé environment is a free open source platform for creating, modifying and studying ontologies (Horridge et al., 2004). It is customizable, enabling the addition of features such as OntoCheck [67], a plug-in that checks for naming and completion errors. Even without any additional options, Protégé

TABLE 5.1
ATTRIBUTES FOR ONTOLOGY QUALITY ASSESSMENT

Attribute	Source	Definition
Adaptability	Vrandečić [80]	Measures how well an ontology anticipates how its future uses and whether it provides a secure foundation which is easily extended and flexible enough to react predictably to small internal changes
Authority	Stvilia [75]	Measures the reputation of an ontology within a particular community
Clarity	Gruber [27]	Refers to whether an ontology effectively communicates the intended meaning of its defined terms and contains objective definitions which are independent of a particular context
Cohesion	H. Yao et al. [79]	Refers to the degree to which the elements of a module belong together
Completeness	Gómez-Pérez [25]	Refers to whether an ontology has sufficiency in its definitions to all possible domains
Conciseness	Gómez-Pérez [25]	Refers to the absence of redundancies including redundancies which could be inferred from its definitions and axioms
Consistency	Gómez-Pérez [25]	Refers to whether it is impossible to retrieve conflicting output results from all valid input terms
Correctness	Gómez-Pérez [25]	Refers to whether the concepts, instances, relationships and properties modeled correlate with those in the world being modeled
Coupling	Orme et al. [52]	Assesses how well the modules work together in systems of ontologies
Craftsmanship	Neuhaus et al. [44]	Refers to whether the ontology is built carefully, including its syntactic correctness and consistent implementation
Deployability	Neuhaus et al. [44]	Refers to whether the deployed ontology meets the requirements of the information system in which it will be used
Expandability	Gómez-Pérez [25]	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
Extendibility	Gruber [27]	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. [44]	Refers to whether the axioms and the annotations of an ontology represent the intended domain correctly
Fitness	Neuhaus et al. [44]	Refers to whether the ontology meets the requirements of its intended use
Intelligibility	Neuhaus et al. [44]	Refers to the ability of all intended users need to understand the intended interpretation of the ontology elements

Maintainability	Rector [59]	Inversely proportional to the amount at which changes cause unanticipated side effects or make unrealistic demands on ontology maintainers.
Modularity	Rector [59]	Refers to whether parts of the ontology can be built separately and later combined without overlarge combinatorial complexity
Relevance	Burton-Jones et al. [10]	Refers to the extent to which the ontology provided the type of information needed by a particular application
Reusability	Duque-Ramos et al. [16]	Refers to the degree to which an ontology, or a portion of an ontology, can be reused for a different purpose or in order to build other ontologies
Richness	Burton-Jones et al. [10]	Refers to the proportion and type of features in the ontology language that have been used in a particular ontology
Sensitiveness	Gómez-Pérez [25]	Relates to how much a small change in a given definition can alter the existing structure of an ontology
Stability	Orme et al. [52]	Measures the extent and type of changes made over time to an ontology, and does not necessarily measure whether these changes are desirable or not

includes simple counters for classes, subclasses, properties and axioms of an ontology and can provide a general idea about the structure of an ontology (**musen20115protege**).

OntoQA (Tartir, Arpinar, and Sheth, 2010) is a Java application that includes metrics for evaluating both populated and unpopulated ontologies, and is also publicly available although it has not been updated since 2010. OntoQA contains a comprehensive suite of metrics that be used to assess and rank ontologies by installing and running it on a user's system. OntoQA does not work with current versions of OWL although its source code is freely available.

Oops! (Poveda-Villalón, Gómez-Pérez, and Suárez-Figueroa, 2014) is probably the easiest to use of the automated ontology assessment systems. It is available on the web and only requires accessing a website and uploading an ontology or its URL to obtain an indication of its quality. Oops! checks for common pitfalls in an ontology and generates a report of the problems the ontology contains. The list of pitfalls that it is able to identify is updated regularly to include new possible errors (Poveda-Villalón, Gómez-Pérez, and Suárez-Figueroa, 2014).

Both Protégé and Oops! are easy to use, but only provide limited information regarding the quality of an ontology. OntoQA gives a more in-depth assessment of the quality of an ontology, but requires more skill from the user to determine the results. The intent of the research reported in this paper, then, is to develop a system that is easy to use, incorporates a web interface, and is comprehensive with respect to quality assessment.

5.1 Semiotic Theory

Semiotics, the study of signs, is a philosophical area dealing with the examination of a sign on multiple levels. Historically, the three layers upon which a sign was evaluated were syntactical (considering formal structure and language), semantic (determining meanings) and pragmatic (regarding intentions and usefulness). Stamper et al. (Stamper et al., 2000) added social as an additional division to the layers because signs and language cannot be evaluated without taking their social context into consideration.

Ontologies contain tightly interconnected collections of signs. Therefore, semiotics should be useful for analyzing their quality (Sowa, 2000). A semiotic approach, similar to that of Burton-Jones et al. (Burton-Jones et al., 2005), but modified and extended, has been developed and employed in this research to ensure that all aspects of an ontology are assessed without redundancy. Semiotics allows us to divide our appraisal into four main areas that concern major invariants in the examination of effective signs: their structure, their meaning, their intentions and their social consequences (Stamper et al., 2000).

Each layer of the semiotic framework is dependent upon having obtained at least a minimal quality level for the layer below it, creating a semiotic ladder (Stamper et al., 2000). If a sign cannot be read, then it cannot have meaning. If it does not have meaning, it cannot fulfill its intended use. If it is not useful, it cannot have value in a community. If the minimal requirements are met at each layer, it is then possible to reach a higher quality assessment at each layer and weight these to obtain an overall quality assessment score. Figure 1 shows the semiotic layers that are relevant for ontology evaluation.

Using semiotics to assess the quality of an ontology requires that two assumptions be

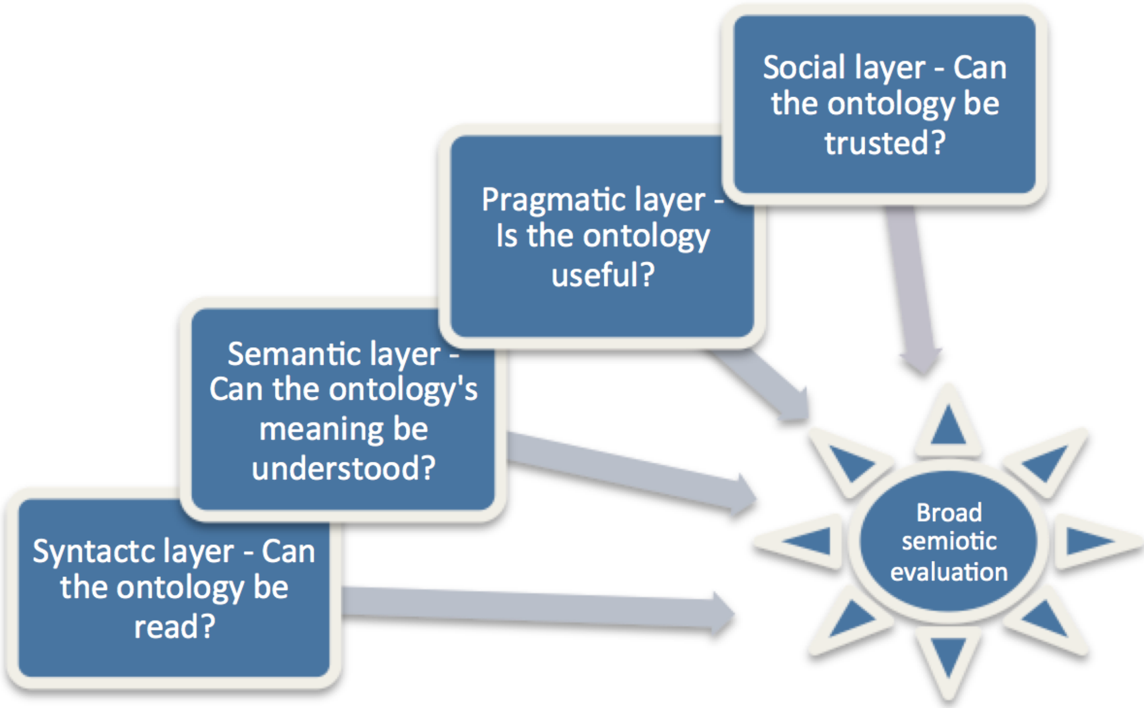

Figure (5.1) Ontology Assessment based on Stamper et al.'s (2001) semiotic ladder

made: 1) the ontology must be represented in a recognized language; and 2) there must be an independent, exterior source of semantic meanings that can be used to assess the semantics of the ontology [10]. If the ontology is not represented in a language that can be parsed, then it cannot be read. Additionally, if the ontology represents a unique domain without a definitive authority of semantic meanings, any system created using semiotic theory would not be a complete evaluation, but would only be partially able to assess the ontology.

Although other work has used semiotics to assess ontologies (Dividino, 2007), none has proposed a broad assessment that is: 1) generalizable enough to be applied to any domain; 2) inclusive of current research related to knowledge reusability and modularity; and 3) easily implemented.

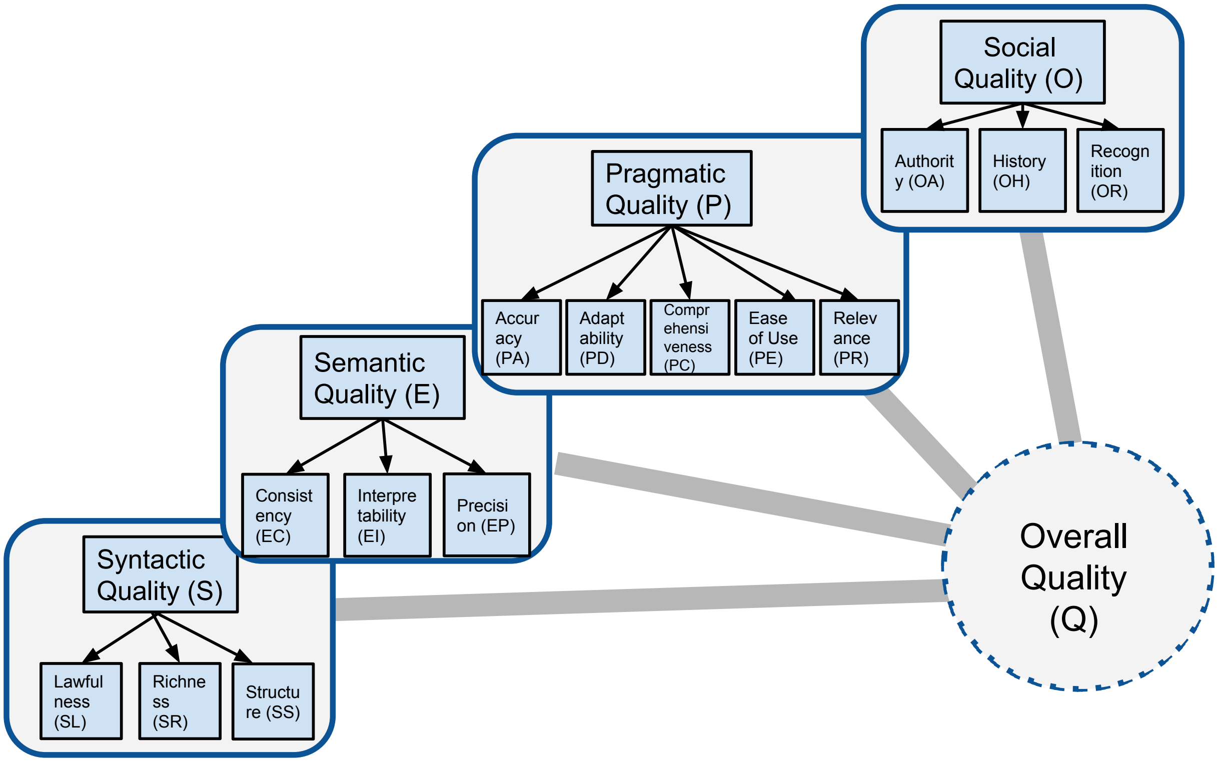

Figure (5.2) Layered Ontology Quality Metrics

5.2 Semiotic-based Layered Assessment

The semiotic ladder requires that each layer of a sign's evaluation is built upon the preceding layer (Stamper et al., 2000). This requirement applies to the evaluation of an ontology's quality as well. If an ontology cannot be read it cannot be understood. If an ontology cannot be understood it cannot be used. If it cannot be used, it does not have value within a community. Therefore the Layered Ontology Quality Metrics Suite is designed so that each layer while defined separately, contributes to an overall quality assessment. Figure 5.2 illustrates the layered nature of the suite.

5.3 Overall Quality Assessment

To provide a comprehensive assessment of an ontology each of the semiotic layers pertinent needs to be adequately evaluated with quantifiable metrics. The Layered Ontology Metrics Suite contains fifteen metrics designed to assess syntax, semantics, pragmatics and social quality. The metrics were derived from prior work (Burton-Jones et al., 2005), (Tartir et al., 2005), (Welty, McGuinness, and Smith, 2004), (Yao, Orme, and Etzkorn, 2005) and extended employing recommendations from the IEEE standard for a software quality metric methodology (Committee, 1998). Each layer of the suite consists of metrics that can be weighted to satisfy a user's particular need. The overall layer assessments can also be weighted. The ability to recalculate outcomes under alternative weights to determine the impact of a particular variable, as part of sensitivity analysis, is a distinct advantage of an automated system in which calculations are performed quickly.

The Layered Ontology Metrics Suite incorporates evaluation of all but two of the attributes for ontology assessment identified previously in Table 5.1. The attributes that are currently not included are deployability and fitness, both of which relate to how well an ontology fits the needs of the user, thus requiring more input from the user than the other metrics. Table 5.2 shows the complete suite of metrics organized by semiotic layer. The metrics that comprise each layer are defined. In many cases, the metrics themselves are composed of sub-metrics that can also be weighted to improve the accuracy of the assessment.

Table 5.2

LAYERED ONTOLOGY METRICS SUITE

Attributes	Calculation
Overall quality (Q)	$Q = w_1 \cdot S + w_2 \cdot E + w_3 \cdot P + w_4 \cdot O$
Syntactic quality (S)	$S = w_{s1} \cdot SL + w_{s2} \cdot SR + w_{s3} \cdot SS$
Lawfulness (SL)	Let b be the total number of breached rules in the ontology. Let s be the number of statements in the ontology. Then $SL = b/s$.
Richness (SR)	Let a be the average number of attributes per class in the ontology. Let r be the ratio of non-inheritance relationships to all relationships in the ontology. Then $SR = w_{sr1} \cdot a + w_{sr2} \cdot r$.
Structure (SS)	Let s be the number of subclasses in the ontology. Let c be the total number of classes in the ontology. Then SS = $w_{ss1} \cdot s + w_{ss2} \cdot c$.
Semantic quality (E)	$E = w_{e1} \cdot EC + w_{e2} \cdot EI + w_{e3} \cdot EP$
Consistency (EC)	Let t be the total number of terms used to define classes and properties in the ontology. Let i be the number of terms with inconsistent meanings. Then $EC = i/t$.
Interpretability (EI)	Let t be the total number of terms used to define classes and properties in the ontology. Let w be the number of terms that have a sense listed in an independent authority. Then $EI = w/t$.
Precision (EP)	Let t be the total number of terms used to define classes and properties in the ontology. Let d be the total number of definitions for terms in an independent authority that occur in the ontology. Then $EP = t/d$.
Pragmatic quality (P)	$P = w_{p1} \cdot PA + w_{p2} \cdot PD + w_{p3} \cdot PC + w_{p4} \cdot PE + w_{p5} \cdot PR$
Accuracy (PA)	Let s be the number of statements in the ontology. Let f be the number of false statements. $PA = f/s$. Requires evaluation by a domain expert or truth maintenance system.
Adaptability (PD)	Let a leaf be defined as a class having no subclasses. Let a be the average number of ancestors for the leaves in an ontology. Let r be the ratio of the number of leaves to the total number of classes in an ontology. Then $PD = w_{pd1} \cdot a + w_{pd2} \cdot r$.
Comprehensiveness (PC)	Let c be the total number of classes and properties in the ontology. Let a be the average value for c across the entire library. Then $PC = c/a$.
Ease of use (PE)	Let S be the number of statements in the ontology, let D be the number that have annotations, then PE=D/S.

5.4 Syntactic Quality Layer

Syntax is concerned with the rules used to model the relationships between terms, as well as for generating and parsing expressions. These rules create a structure or model of the relationships between the terms. Inheritance relationships, such as class and subclass relationships create a tree structure with a measurable depth and breadth. When a phrase can be generated or parsed in more than one way, syntactic ambiguity occurs making the meanings of the individual signs less clear. It is possible to classify and rank syntaxes based on their ability to generate non-ambiguous structures (Stamper et al., 2000). Therefore, this layer of the suite contains metrics to evaluate 1) the lawfulness of an ontology; 2) the richness of the non-inheritance relationships be-tween the terms; and 3) the tree structure representing the class and subclass relationships in the ontology.

Symbols can be mapped to one another according to syntactical rules deter-mined by the language in which the ontology is represented. The Lawfulness metric accesses the degree to which these rules have been followed. This metric is important because, without correct syntax, an ontology cannot be read or used (Burton-Jones et al., 2005)

The richness of the ontology's structure is assessed by two separate metrics, Richness and Structure, each of which provides information about the complexity of the ontology's internal links. It is important that an ontology takes advantage of the richness of the language used to express more than simply class and subclass relationships, such as particular attributes possessed by the class or subclass. Therefore, the Richness metric is determined by the ratio of the non-inheritance relationship to the inheritance relationships within the ontology and the overall number of attributes applied to the terms of the ontology.

The Structure metric assesses the ratio of subclasses to total classes in order to assess the structure's depth and breadth. A syntactical structure that is too deep, in which each class only has one subclass, is lacking syntactical relationship information.

The complete assessment of Syntactic Quality is performed using the weighted average of Lawfulness, Richness, and Structure to derive an overall evaluation of the relationship structure represented by the ontology.

5.5 Semantic Quality Layer

A measure of semantic quality evaluates the vocabulary used to convey information. Guarino (Guarino, 1997a) recommends that the terms adopted for concepts, relations, and attributes in an ontology should clearly express the underlying assumptions that they implicitly contain. The Semantic Quality metric measures the quality of the interpretations of term meanings by comparing them to an independent authority. By default, the assessment of semantic quality employs WordNet (Miller, 1995), a large lexical database for English, to find the meanings of the terms used. By identifying the number of potential meanings for each term, it is possible to assess whether the meanings of the terms used in the ontology are clearly defined. Many scientific areas include a lexical database of definitions that can be used in place of WordNet to define terms specific to their domain. The three semantic metrics measured by the suite are Consistency, Interpretability, and Precision, each of which provides information regarding the quality of the vocabulary and relationships expressed by the ontology.

Consistency assesses the number of terms in an ontology that have semantic conflicts. A semantic conflict is an ambiguity caused when a sign can represent more than one potential meaning. An ontology with perfect consistency will guarantee that all included terms have only a single meaning, and that no rules are conflicted by other relationships.

Ideally, it should be possible to map an ontology's terms to real world concepts. Checking whether the terms are contained in an independent semantic source is a useful way to perform this mapping. Therefore, the Interpretability metric measures the percentage of the terms in the ontology that are present in the selected reference dictionary.

For clarity of meaning, it is best to use the most precise terms available. Therefore, Precision assesses whether terms in the ontology have only unambiguous definitions related to the domain under consideration.

5.6 Pragmatic Quality Layer

In semiotic theory, pragmatics is the measure of a sign's usefulness in conveying the meaning intended by its creator (Stamper et al., 2000). To be useful, an ontology must communicate domain information as accurately and as completely as feasible. Other research has shown that the ability to adapt an ontology so it can be used for a new purpose, other than that for which it was originally designed, also indicates a useful ontology (Gangemi et al., 2006a), (Vrandečić, 2009). Guarino has long asserted that even when modeling knowledge for an individual task, it is important to make an effort to foresee how the model could be used for other purposes (Guarino, 1997b).

The Layered Ontology Metrics Suite assesses ontologies on five pragmatic attributes: accuracy, comprehensiveness, relevance, adaptability and ease of use. The Accuracy metric is the most challenging of these to assess because most ontologies are intended for a particular domain. To determine whether the attributes and inheritance relationships within the ontology accurately model its domain requires evaluation by a domain expert. The other four pragmatic attributes, however, can be assessed or at least approximated by examining the structure and contents of the ontology.

Comprehensiveness measures the breadth of coverage provided by the ontology. Relevance, on the other hand, measures how specific the ontology classes are to the required search terms. Together these two metrics give valuable information about the ability of the ontology to cover its domain accurately, completely and specifically.

Reusability refers to the degree to which an ontology, or a portion of one, can be reused for a different purpose or in order to build other ontologies (d'Aquin and Noy, 2012). Adaptability and Ease of Use both assess reusability, but take different approaches. Adaptability measures whether the ontology provides a secure foundation that is easily extended and flexible enough to react predictably to small internal changes (Vrandečić, 2009). Ease of Use, on the other hand, assesses the level of documentation in the form of annotations included. Annotations can consist of comments, descriptions, labels or other useful in-formation, that is not read by the system, but guide humans in understanding and adopting an ontology to use for a new purpose.

Because pragmatic quality is related to the task for which an ontology is designed, or the specific domain or sub-domain for which it will be used, the metrics for Pragmatic Quality require additional user input for their assessment. The usefulness of these metrics have been evaluated in two separate phases of the empirical evaluation of the metrics suite. These metrics are more clearly defined and their specific testing procedures are covered in detail in Chapter 8.

5.7 Social Quality Layer

Social Quality assesses how well an ontology is accepted in the community of which it is a part. Domain ontologies usually reside in a community with common interests and a specific purpose. The Social Quality metrics, therefore, are intended to capture the fact that agents and ontologies do not exist in isolation but instead are part of knowledge communities (Burton-Jones et al., 2005). Three weighted metrics, Authority, History, and Recognition comprise the Social Quality assessment.

Stvilia et al. (Stvilia et al., 2007) defines authority as the “degree of reputation of an ontology in a given community or culture” (p.10). The Authority metric for a given ontology is determined by counting how many other ontologies link to it, and within each link, how much information is shared [87]. A higher assessment for Authority indicates that a greater number of other ontologies define classes and relationships using its definitions (Burton-Jones et al., 2005).

It is important that the definition of a concept not be separated from the discussion that lead to the shaping of the concept because the history of its conceptualization is an important part of the definition (Hepp, 2007). Therefore, an ontology, which is made up of concepts and rules, should not be separate from the history of those rules. The History metric assesses how long the ontology has been in a specified library, as well as the number of revisions it has undergone, suggesting that an ontology with a longer history is more likely to be dependable (Burton-Jones et al., 2005).

An ontology's recognition assesses the level of use it has received within its community. The Recognition metric, therefore, is determined by the number of times an ontology has been downloaded, the number of reviews it has received, and the ratio of positive to negative reviews. Each of these three components of the Recognition metric is determined by its comparison to the average value for that metric with-in the community.

Because social quality is dependent upon an ontology's value within its particular community, the metrics for Social quality were given particular care in their formalization. These metrics are more clearly defined and their specific testing procedures are covered in detail in Chapter 9.

CHAPTER 6

DOMAIN ONTOLOGY RANKING SYSTEM

To demonstrate the effectiveness of the suite of metrics, they were implemented in a Domain Ontology Ranking System (DoORS), the purpose of which is to rank ontologies for a given set of user criteria. The architecture of DoORS is shown in Figure 6.1.

The main component is the Ontology Quality Assessment Module, which calculates the values for each of the metrics. The other components manage the user interface, search for candidate ontologies, rank the ontologies, and select a subset to recommend to the user. The web interface allows the user to specify an ontology library to be searched, as well as the desired search terms and the weights to be applied to the individual metrics to determine the best task-domain fit. Because of the modular nature of the Layered Ontology Metrics Suite, DoORS is constructed to have an assessment component with four separate modules, one for each semiotic layer: syntactic, semantic, pragmatic and social.

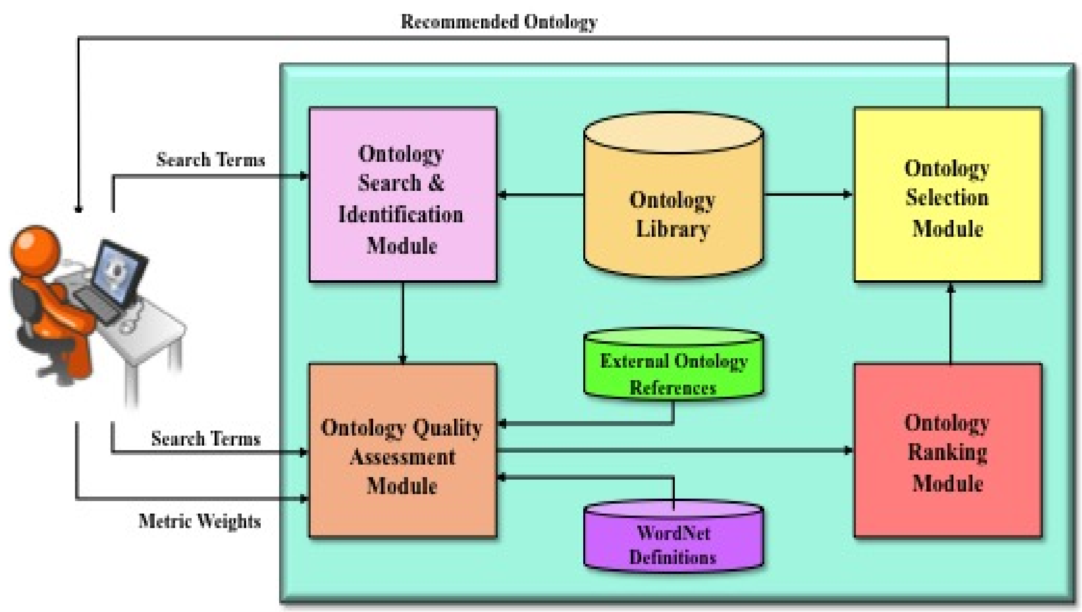

Figure (6.1) Architecture of the Domain Ontology Ranking System

6.1 Architecture of DoORS

The Domain Ontology Ranking System consisting of an: a) Ontology Search and Identification Module, b) Ontology Quality Assessment Module, c) Ontology Ranking Module, and d) Ontology Selection Module. Each module serves an explicit purpose in the overall task of providing a user with a list of recommended ontologies for a specific need.

The Ontology Search and Identification Module handles user input, supports the ontology library search methodology, and retrieves the initial set of candidate ontologies for consideration. It allows a user to provide the URL for the ontology library to be searched, terms needed for a domain or task, and weights to reflect which metrics are considered most important. The user has control over the ontology selected by the choice of input parameters. Repeated use of the system can give the user different output results based on the parameters entered each time. DoORS does not make the final decision of which ontology to select, but instead, provides helpful information to enable the user to make that decision. This module identifies an initial set of candidate ontologies to be assessed.

The initial set of candidate ontologies is input to the Ontology Quality Assessment Module, which calculates a score for each ontology. To assess ontological quality, this module utilizes WordNet Definitions (Miller, 1995), external ontology references, and the ontology structure. This module is the most crucial portion of the system because it applies the Layered Ontology Metrics Suite to make the determination of ontology quality. The module is composed of four submodules, one for each of the Semiotic theory layers pertinent to ontologies, as described below.

The Ontology Ranking Module reverse orders the ontologies based on the overall score computed by the Ontology Quality Assessment Module. The resulting list of ontologies with their corresponding scores serves as input to the Ontology Selection Module. The list is ordered in reverse score order so that the first ontology output is deemed the best fit.

The main task of the Ontology Selection Module is to identify the threshold upon which a candidate has received a score high enough to be presented to the user. The task characteristics specified by the user and the ontology characteristics are used to make this assessment. This module ensures that at least one ontology is suggested, but does not always give the same number of ontologies to the user if very few are of high enough quality. This module also handles the output, providing the user with the ontology names, along with the ranking scores and other information about the recommended ontologies.

6.2 Ontology Quality Assessment Module

The Ontology Quality Assessment Module of DoORS is comprised of four individual submodules—one for each layer of the semiotic framework. Each submodule has prescribed input required for calculating the assessment value of the metric for that particular layer. Figure 6.2 shows the Ontology Quality Assessment Module which consists of the: a) Syntactic Quality Assessment Submodule, b) Semantic Quality Assessment Submodule, c) Pragmatic Quality Assessment Submodule, d) Social Quality Assessment Submodule, and e) Overall Quality Assessment Submodule. Because each of the metrics is computed individually, running the four submodules in parallel could reduce the time required for processing.

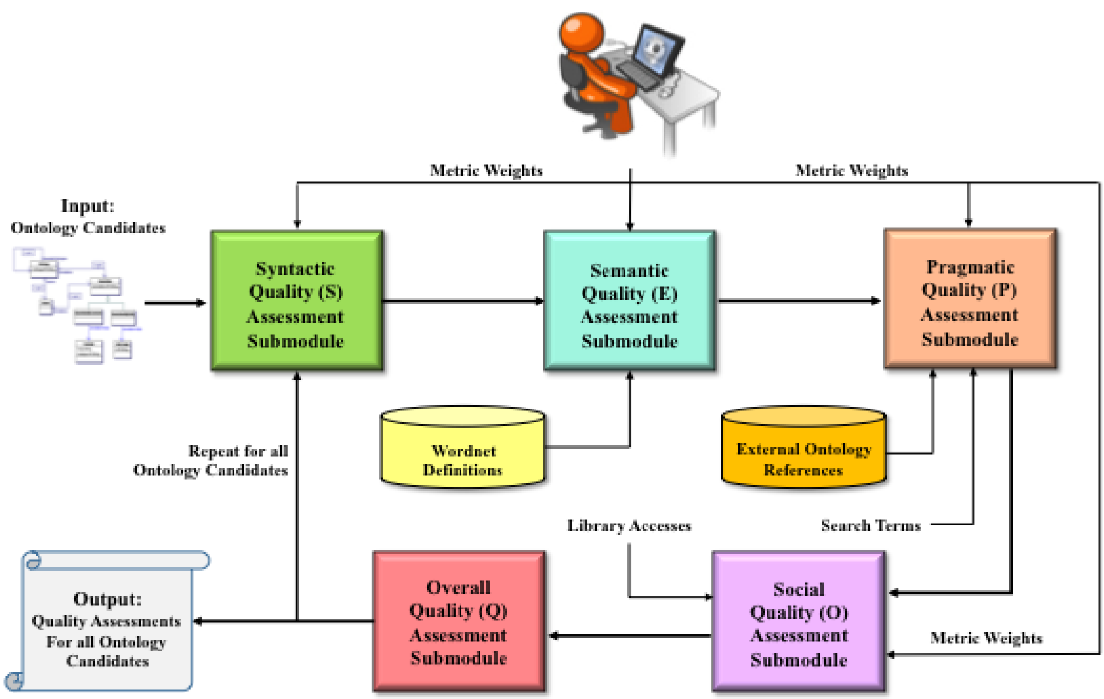

Figure (6.2) Architecture of the Ontology Quality Assessment Module

The Overall Quality Assessment Submodule computes an overall quality score for an ontology by aggregating the individual scores for the syntactic, semantic, pragmatic and social quality metrics. The input parameters determine the weights for each metric. By default, all metrics are equally weighted.

6.3 Implementation Details

An prototype of the Domain Ontology Assessment Module of the ranking system has been developed using the Python Language Version 2.7 according to the design decisions and constraints outlined in Table 6.1. Python was chosen as the language to use because of the availability of the Python Natural Language Toolkit [7] for natural language parsing and the ElementTree module (Robinson, Aumann, and Bird, 2007) for determining the shape and size of the relationship structure represented by the parsed ontology. The OWL language used as the predicted language that the ontology would be represented in is Manchester OWL (Horridge et al., 2006) although the system will convert the ontology to that language by accessing the Manchester OWL API (Horridge and Bechhofer, 2009) if the ontology is represented in a different RDF format.

Table 6.1 Design Constraints

Constraint	**Rationale**
Scalability	Large Ontologies
Generalizable	All domains
Automatable	System Interoperability
Modular	Easy to maintain
Web Application	Ease of use
Open Source	Contribute to other research

Algorithms used include linear search, quick-sort and chunking (streaming by splitting the ontologies into 8 kilobyte segments). For the final version of the DoORS system, each of these algorithms will be explored more thoroughly to analyze their running times and to determine whether more efficiency is possible. In particular, the option of running the DoORS system on a GPU cluster to take advantage of parallel processing with the pyCuda library will be explored.

This implementation is available as a web application for users to input individual ontologies and receive an assessment based on ten attributes of the suite. All of the attributes from the Syntactic and Semantic layers have been implemented. From the Pragmatic layer, all attributes except Accuracy have been implemented since the Accuracy metric requires an approximation or the participation of a domain expert. The metrics from the Social layer require for the user to enter the URL of the on-line community of which their ontology is a member. These metrics have been shown to be feasible in other research (McDaniel, Storey, and Sugumaran, 2016) and will be added to the complete DoORS application at a later date. Figure 6.3 displays the currently implemented module of the DoORS architecture.

Figure 6.4 shows a screenshot of the user interface for the current version of DoORS, located at https://owlparser.herokuapp.com/. The source code for the ontology converter is available at http://github.com/MelindaMcDaniel/ontconverter. The source code for the ontology parser is available at http://github.com/MelindaMcDaniel/owlparser The source code for the social quality assessment prototype is available at

http://github.com/MelindaMcDaniel/socialquality.

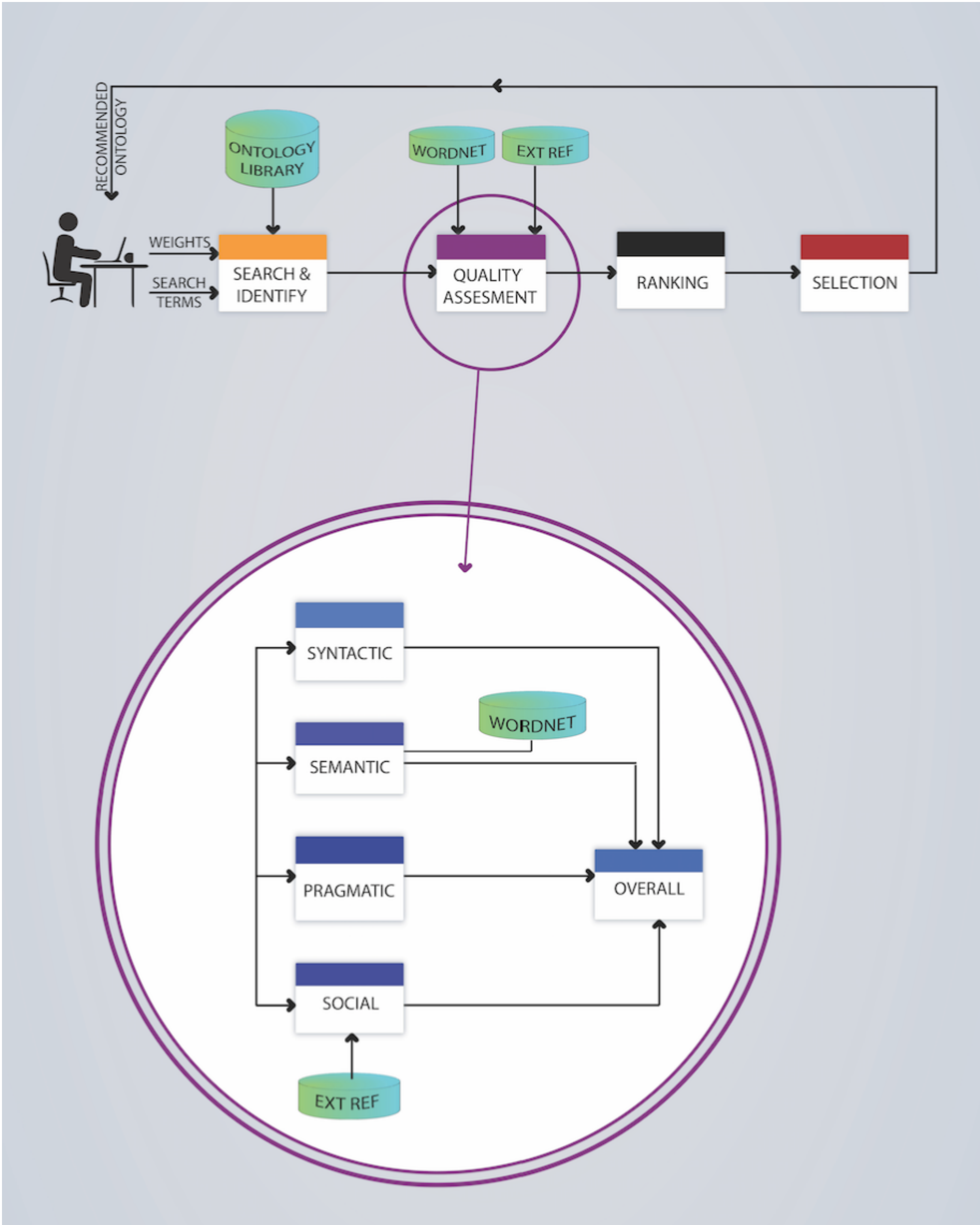

Figure (6.3) Currently implemented module of the DoORS architecture

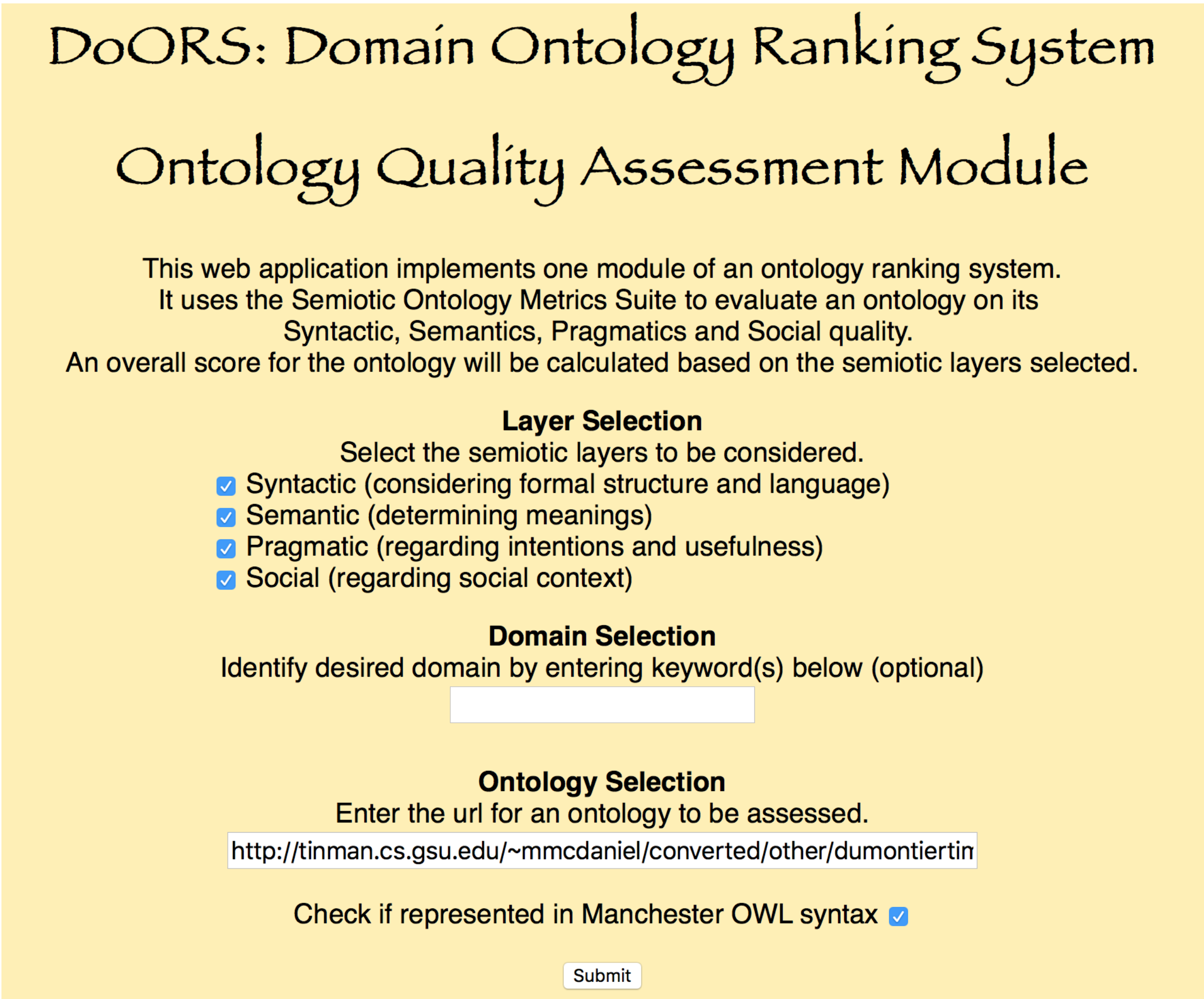

Figure (6.4) Screenshot of the DoORS User Interface

CHAPTER 7

EVALUATION OF METRICS AND THE DOORS APPLICATION

7.1 Evaluation

Both qualitative and quantitative measures were employed to prove the usefulness and validity of the system. We determined that these four specific requirements for the DoORS system must be met for the system to be considered valid: 1) the system should be usable by human ontology selectors, 2) the system's metrics should provide meaningful numbers, 3) the domain must be accurately represented by the numbers and 4) the variations between the provided scores should be different enough to be meaningful when used as part of the ontology selection process. To ensure that all requirements were met, four phases of evaluation were done. Figure 7.1 outlines the evaluation process.

Phase	Description	Purpose
1	Compare DoORS to other ontology quality assessment systems.	Evaluate practical usefulness of the system
2	Demonstrate DoORS in one on one meetings with experienced ontologists	Evaluate Face Validity
3	Compare DoORS results when run on Good, Average, and Poor quality ontologies identified from literature.	Evaluate level of distinction between scores
4	Obtain domain expert opinions of which of three ontologies are the most useful for a specific domain	Evaluate modeling of a domain

Figure (7.1) Phases of Empirical Evaluation

Qualitative evaluation with visual interaction was considered appropriate for the usability and meaningfulness of the system to show face validity and usefulness. Face Validity is a subjective evaluation of the performance of a system by people knowledgeable about the application domain. For our test case, the people were ontology engineers who had created, selected and examined ontologies previously to show whether the interaction process with the system is usable.

Quantitative evaluation for content validity was considered appropriate to prove domain representation and meaningfulness of the scores provided. This was done in two separate experiments; one experiment compared the results of the system to other ontology assessment systems and the other experiment compared the results of the system to determinations made by human domain experts.

Before any other evaluation was done, a simple preliminary experiment with a few test subjects was carried out to make sure that the metrics being used by the system were appropriate. This experiment consisted of three web developers examining four sample ontologies with five ontology attributes being evaluated.

7.2 Phase 0 - Preliminary Experiment

The BioPortal ontology repository was first used in the evaluation. BioPortal, an extensive medical and scientific ontology library, contains approximately 500 ontologies with over 6 million classes (Whetzel et al., 2011). Each of these libraries contains ontologies stored in a variety of formats including xrdf and Owl (Antoniou and Van Harmelen, 2004). BioPortal includes statistics on the ontologies, including the number of classes, subclasses, and properties, as well as a visualization of each ontology in the library. Some of the ontologies also provide information about projects currently using the ontology and independent reviews. Users of the BioPortal website can obtain information about each of the ontologies to help users decide whether to download the ontologies for their use (Whetzel et al., 2011).

The first task of the preliminary testing was to simulate a scenario where a novice user needs to select an ontology to use in his or her task/application. A web developer is an example of a user who would have technical knowledge about ontologies and their internal structure and relationships in order to be able to develop an intelligent agent-based application. While these users may not have extensive knowledge about a particular domain, they are familiar with the overall structure of ontologies and can select an appropriate one based on the requirements of the application they are developing.

The objective of the preliminary testing is to have such users select and rate several ontologies from the BioPortal ontology repository that contain a particular concept. These results are compared to the set of ontologies suggested by the DoORS system, based upon the overall quality metric that the system computes. The experiment was initially carried out with a single web developer who was given 1) a sample task for which an ontology would be needed and 2) information regarding five ontologies that could be used for completing the task.

The test subject was instructed to rate how well each of the ontologies possessed each of the attributes that are implemented in DoORS: Adaptability, Interpretability, Precision, Relevance, and Richness. Using a Likert Scale (Likert, 1932) from 1 to 5 where 1 is strongly disagree and 5 is strongly agree, the tester recorded how well she thought each of the ontologies reflected each of the five attributes. Based upon comments recorded during the experiment (Kuusela and Pallab, 2000), the form was modified slightly to include definitions of the five attributes.

Next, two other web developers rated the same ontologies again using the Likert Scale for each of the five attributes. These test subjects were also encouraged to talk aloud as they worked through the form and their comments were recorded.

An advantage of selecting BioPortal ontologies for this research is the availability of user reviews and other information about the ontologies stored there. The test subjects were provided with this information to help them make their assessments. The comments from each of the three test subjects made it clear that the decision of which ontology would be the best was difficult, even with the vast amount of information provided by BioPortal. The subjects stated that they appreciated having the list of five attributes to help them determine the best dimensions upon which to measure an ontology. The test subjects judged the attribute of adaptability as the hardest to measure because the ontologies were lengthy.

The time required for the three subjects to complete the ratings for the five ontologies ranged from 30 to 40 minutes. On the BioPortal website alone there are over 60 ontologies that include the term blood, for example [83]. Therefore, selecting from all possible ontologies, rather than simply five ontologies, would be a more time-consuming task. In contrast, the DoORS tool can quickly perform the metric calculations and recommend ontologies.

Although each of the three test subjects rated the ontologies differently on different attributes, both their comments and the scores revealed a clear consensus that some ontologies were clearly 'better' than others. Comments made when examining the RadLex radiology ontology (Rubin, 2008), in particular, were negative from all three subjects, conveying the opinion that the content did not seem to be well-connected, and that the relationships between the terms in the ontology were not clear.

The NCIT ontology provided by the National Cancer Institute (Golbeck et al., 2011), on the other hand, elicited many positive comments including that it contained a great deal of useful information, had good reviews from other users, and was being currently used in other projects. Unfortunately, not all the ontologies have reviews on the BioPortal library website (Whetzel et al., 2011), so this information is not always available. The three other ontologies, Uberon Ontology [42], Environmental Ontology (Buttigieg et al., 2013), and the Foundational Model of Anatomy (Rosse and Mejino, 2003), each received middle range scores, mostly 2, 3 or 4 out of five, on the five attributes. Thus, each was found to have some aspects that the subjects valued.

All of the test subjects found it challenging to make this type of decision. Even though the BioPortal website contained extensive amounts of information about each ontology, the test subjects were unclear about how this information could be put to use for assessing which ontology would be the best for a given task. They also expressed frustration at how long it took to study all the information and make a selection.

The automated assessment portion of DoORS, on the other hand, was able to determine the scores for these five metrics almost instantly. The NCIT ontology, which had been favored by the three test subjects, received the highest marks on two of the attributes using the system. The RadLex ontology, which had been unpopular with the human testers, received the lowest marks on four of the five attributes using the system. The other three ontologies, each received the highest score on one of the attributes, ranging somewhere in the middle of the other attributes. Although the actual attribute scores determined by the DoORS prototype were not the same as those of the human testers, the overall assessment of which would be a "good" ontology for creating a website pertaining to blood, were markedly similar.

Figure 7.2 displays a comparison between the results from the human test subjects and the results determined by DoORS. The results for the human testers were calculated

Preliminary Experiment Results

Ontology Name	DoORS Normalized	{Rank}	Human Normalized	{Rank}
UBER – Uberon Anatomy Ontology [42]	.82	{5}	.69	{3}
NCIT – National Cancer Institute Thesaurus [24]	.71	{4}	.88	{5}
ENVO – Environmental Ontology [11]	.68	{3}	.69	{4}
FMA – Foundational Model of Anatomy [64]	.61	{2}	.64	{2}
RADLEX – Radiology Lexicon [65]	.47	{1}	.32	{1}

The results of the DoORS assessments for the five ontologies and the determinations made by the human testers were similar on the five attributes measured. A rank of 5 is the highest rank for both systems.

Figure (7.2) Results of preliminary comparison between human users and DoORS

by adding the total scores that each ontology earned on all five attributes. Because the scale ranged from 1 to 5 on each of the metrics, the total possible score ranged from 5 to 25. Each total is the average of the determinations of the three test subjects. The totals for the automated system, however, are based on the scores between 0 and 1 for each of the five attributes. Therefore, the highest possible score an ontology could receive from DoORS is 5; the lowest is 0. Both sets of scores were normalized between 0 and 1 for comparison.

7.3 Phase 1 - Comparison to other systems

The next evaluation phase was designed to determine whether the metrics and the system that implemented them would provide comparable information to what is determined by other currently available software. It was hoped that the system would achieve similar results, but would also provide additional information not supplied by other ontology assessment tools. Only three domain ontology assessment tools are currently available as web applications. The BioPortal recommender (Noy et al., 2009) provides a broad domain ontology assessment including numeric values for a range of ontology attributes. Therefore we decide d to do an explicit comparison between the BioPortal results and the DoORS results. The Prot'eg'e development tool (Noy, Fergerson, and Musen, 2000) and the Oops! Pitfall Scanner (Poveda-Villalón, Gómez-Pérez, and Suárez-Figueroa, 2014) on the other hand, do not provide specific numeric values, but instead provide other information, such as a descriptive error list, that can be compared to the DoORS numbers. For these two systems an implicit comparison was considered appropriate.

7.3.1 Explicit comparison

The BioPortal Recommender provides an overall assessment of an ontology based on four attributes: coverage, acceptance, detail of knowledge and specialization. Although the BioPortal recommender system only evaluates ontologies stored in its online repository, it provides a broad assessment of an ontology and allows weighting of the metrics. It compares the scores between selected subsets of the ontologies, and includes a keyword search [47]. The BioPortal Recommender produces a number between 0 and 100, which it uses to rank the ontologies by quality. The DoORS system produces a number between 0 and 1 for each of the five attributes it assesses. The final scores of each system were normalized between 0 and 1 for comparison.

7.3.2 Implicit comparison

The Oops! online web service allows a user to submit an ontology that Oops! will assess for potential problems, or provide a link to an ontology or copy and paste the ontology itself in order to search for problems. The list of problems, called 'pitfalls' that Oops! is able to detect, range from minor problems such as missing annotations, to major problems such as recursive definitions (Poveda-Villalón, Gómez-Pérez, and Suárez-Figueroa, 2014). DoORS was compared to Oops! because they are both intended to carry out ontology evaluation and provide a broad assessment of an ontology's quality.

Protégé is a tool used primarily for building and modifying ontologies that includes built-in metrics that provide information about the structure of an ontology. Specifically, Protégé itemizes and provides counts for the number of classes, properties, and axioms (Noy, Fergerson, and Musen, 2000). Because Protégé does not produce a numeric score for an ontology's quality rating, for the purposes of this study, a general ranking can be

Explicit Evaluation Results

	DoORS Results		BioPortal Results	
Ontology Name	Normalized	{Rank}	Normalized	{Rank}
UBER – Uberon Anatomy Ontology [42]	.82	{1}	.93	{2}
NCIT – National Cancer Institute Thesaurus [24]	.71	{2}	1.0	{1}
ENVO – Environmental Ontology [11]	.68	{3}	.76	{3}
FMA – Foundational Model of Anatomy [64]	.61	{4}	.59	{4}
RADLEX – Radiology Lexicon [65]	.47	{5}	.56	{5}

Results of DoORS are compared to the results of BioPortal's Recommender System. Norm. indicates Normalized between 0 and 1 to compare the two systems. {Rank} indicates overall ranking where 1 is the highest rank.

Figure (7.3) Explicit Comparison of DoORS to BioPortal Recommender

performed based on the presumption that ontologies with more information are better and thus ranked higher.

7.3.3 Results of Phase 1

Many similarities are noted between the results obtained by DoORS and each of the three online ontology assessment tools. In particular, the numerical results determined by the BioPortal recommender system show close correlation to the numerical results of DoORS. The two other systems, Oops! and Protégé, both of which do not provide an actual numeric comparison, also show patterns that indicate similarities in ontology quality preferences.

Both DoORS and the BioPortal Recommender provide a broad overview of an ontology's quality, and give numeric results to provide an explicit comparison as shown in the table shown as Figure 7.3. The ranking from 1 (the highest ranked) to 5 (the lowest ranked) of the five ontologies is also displayed. The ranking is even more meaningful than the numeric scores because the intended purpose of the DoORS tool is to help a user make a selection when multiple choices are available. Although the results were not identical between the BioPortal Recommender and DoORS, their identification of which of the five ontologies were the first and second best choices are in agreement.

The BioPortal Recommender can only assess ontologies stored in its own repository and is dependent on information stored there to make the assessment. DoORS, on the other hand, can handle any ontology stored online, regardless of whether it is part of an ontology repository. The other two systems in the evaluation, Oops! and Protégé, do not provide explicit scoring for an ontology. Instead, they each provide useful information that can help a user choose from a list of possible ontologies. Therefore, their comparison with DoORS is appropriate.

The Oops! system identified possible flaws in ontologies and found a few such errors in the ontologies used for this evaluation. Although all of the ontologies contained at least one error, the first and second choice ontologies selected by DoORS contained fewer critical errors than the other three ontologies, as illustrated in Figure 7.4. Because a high-quality ontology should be as error-free as possible, the table also shows the rank ordering determined by a smaller number of pitfalls. A rank of 1 indicates the highest rank. The Oops! online tool is well-maintained, frequently updating its list of pitfalls. Oops! also provides a REST service so it would be possible to incorporate a pitfall check into the DoORS system as a future enhancement (Poveda-Villalón, Gómez-Pérez, and Suárez-Figueroa, 2014).

Protégé provides counts regarding the number of classes, properties and axioms when an ontology is uploaded into the application. A ranking using the average of these three counts is included in the table shown as Figure 7.4. Because these counts only measure the size and complexity of the ontology, and not its other attributes, only a partial correlation can be made.

Implicit Evaluation Results

Ontology	DoORS		Oops!		Protégé			
	Norm	{Rank}	Important Pitfalls	{Rank}	Classes	Properties	Axioms	{Rank}
UBER [42]	.82	{1}	11, 24, 41	{2}	18,831	189	263,157	{2}
NCIT [24]	.71	{2}	4, 8, 11, 13	{3}	118,167	173	1,508,597	{1}
ENVO [11]	.68	{3}	11,41	{1}	8,969	15	126,552	{5}
FMA [64]	.61	{4}	10,11,38,41	{4}	78,978	7	715,965	{3}
RADLEX [65]	.47	{5}	3,4,8,10,11,12,13,22	{5}	46,060	95	483,789	{4}

Results of DoORS are compared to the results of the Oops! and Protégé ontology systems. {Rank} indicates implicit overall ranking where 1 is the highest. Rank is determined by least pitfalls in the case of Oops!, and by greatest amount of information available in the case of Protégé.

Figure (7.4) DoORS results are compared to Oops! and Protégé

7.4 Phase 2 - Evaluating User Interface

Qualitative evaluation with visual interaction was considered appropriate for the usability and meaningfulness of the system to show face validity and usefulness. Face Validity is a subjective evaluation of the performance of a system by people knowledgeable about the application domain. For our test case, the people were ontology engineers who had created, selected and examined ontologies previously to show whether the interaction process with the system is usable. One of the experts has developed and used ontologies as a part of the semantic web for his work with Delta. The other two experts are academics who have developed and explored ontologies for their academic research areas.

A thirty-minute demonstration was given to each of three ontology engineers to get feedback on the potential usefulness of the system for selecting an ontology when multiple choices are available. The sessions provided insight into the advantages and disadvantages of the DoORS interface as it currently exists. The three ontology experts in general, thought the user interface was well-designed, but the options available were limited. The comments from the ontology developers is summarized in Table 7.1. From these comments, it can be concluded that the system is easy to use and provides useful

Table (7.1) **Face Validity Comments**

Comments about Current Implementation	**Improvements Desired**
The system is able to process an ontology quickly and provides valuable assessments of many attributes of ontology quality.	It would be nice to enter several ontologies at the same time to have them compared on particular feature or set of features.
The instructions are clear. The user input boxes are easy to use.	I would like to be able to prioritize the attributes importance to my preferences without having to enter percentages.
I like having the ability to choose only one area for ranking.	The numbers don't mean much to me except in comparing one ontology to another. It would be better if the numbers had ranking scales.
In my area, most of the ontologies are extremely large. I'm impressed at how fast this system works.	What if none of the ontologies is good? How would I know that? Would it still tell me which one is the best?

information for ontology ranking although more information could improve the system even further.

7.5 Phase 3 - Evaluating Score Distinction

The second quantitative evaluation that was done compared the results produced by the system when ranking ontologies that had been shown in literature to be high quality to the rankings of other ontologies. The ontologies selected for this experiment were ontologies from three separate ontology repositories. The ontologies retrieved from the W3C good ontologies (Berrueta et al., 2008) repository were specifically selected as examples of ontologies of high quality.

The ontologies retrieved from the Protégé repository (Noy, Fergerson, and Musen, 2000) represent average ontologies. The repository stores ontologies recommended by at least one source as being good ontologies, but no additional validation has been made. These ontologies, however, have been used for at least one task, therefore a certain level of quality can be assumed. The ontologies retrieved from the OntoHub sandbox repository (Mossakowski, Kutz, and Codescu, 2014) are ontologies that have no expectation of

Phase 3 Results

Category	**Standard Deviation**	**DoORS Rating**
Good Ontologies	.08	.71
Protégé Ontologies	.08	.53
Sandbox Ontologies	.06	.46

Table (7.2) DoORS ratings for good, average, and poor ontologies

quality. They are stored in the OntoHub sandbox for experimental purposes and have no expectation of quality. Many of the OntoHub sandbox ontologies are no longer accessed.

7.5.1 Phase 3 Results

Each of the ontologies from these repository categories was processed by the DoORS system. The results, shown in Table 7.2, show that the scores are meaningful. The average score received by the set of ontologies endorsed as "good" was slightly higher than that of the "average" ontologies, and significantly better than the set of ontologies from the OntoHub sandbox. To prove whether these numbers were significantly different than could have occurred from an aberration in the data, a T-test was done. For this test, a p-value was determined for each null hypothesis. For the null hypothesis that average ontologies are no better than poor ontologies, the p-value was shown to be 0.158655. For the null hypothesis that good ontologies are no better than average ontologies, the p-value was shown to be even more significant, only .012224. Therefore there is significant difference between the quality of the ontologies in the sandbox, in the Protégé library of average level ontologies, and in the selection of ontologies identified by literature as being exemplary.

7.6 Phase 4 - Evaluating Domain Modeling

The DoORS system evaluates ontologies based upon four layers of the theory of semiotics. Semiotics, the evaluation of signs, is appropriate for ontologies, which are made up of signs. Therefore the system's metrics are valid in theory, but in order to prove their validity it was necessary to compare the results of the DoORS system to the results determined by humans knowledgeable in a particular domain. The domain selected for this experiment was the domain of computer science; specifically computer science nouns used in introductory computer science classes. Three ontologies were created following the guidelines stipulated by the W3C consortium for ontology development in varying degrees of adherence. The three ontologies were then tested using the DoORS system to assess their score based on the semiotic ontology metrics suite.

Three ontologies were especially designed to receive high, medium and low rankings on the semiotic levels of meanings of signs. One ontology was designed to receive high score on syntax, semantics and pragmatics. Another was designed to receive medium scoring on the same three levels of semiotics. The third was designed to receive low scores on each of there assessment levels. To assure that they met the criteria, they were each assessed by the DoORS software interface and the scores were as expected. The three ontologies, assigned letters A, B and C, were then sent to potential participants knowledgeable in computer science terminology. These participants were members of a mailing list for teachers of high school Advanced Placement Computer Science courses; therefore they met the criteria of being knowledgeable in the domain.

The results of the evaluation, outlined in Table 7.3, show that the DoORS system provides results comparable to those of the human domain experts. Ontology A, the ontology preferred by the human test subjects, was also given a higher assessment score by the semiotic ontology assessment system. Ontology B, which was less preferred by the human test subjects, received the second highest score from the system. Ontology C, which most of the human test subjects rated as the ontology of the three that performed

Phase 4 Results

Ontology	DoORS score	Positive Votes	Negative Votes
A	.75	.85	.04
B	.74	.11	.13
C	.69	.07	.80

Table (7.3) DoORS ratings compared to human domain expert preferences

the worst in covering the domain, also received the lowest assessment from the rating system.

Although the human test subjects did not assign numeric values to the three ontologies they rated, their ontology preferences closely matched those of the semiotic ontology rating system. In particular, ontology A was vastly preferred over the other two ontologies. Because of this correlation, it follows that content validity has been proven.

7.7 Evaluation Summary

The challenges caused by the size of the ontologies and the difficulty of finding testing subjects knowledgeable in a specific domain and also experienced with domain ontology use led to the decision to evaluate the semiotic metrics suite and the DoORS implementation in multiple phases to ensure that all evaluation goals were met. Each of the phases of the empirical evaluation achieved its specific goal as outlined in figure 7.5.

Phase	Description	Purpose	Results
1	Compare DoORS to other ontology quality assessment systems.	Evaluate practical usefulness of the system	DoORS ratings were consistent with those of Oops!, Protégé, and BioPortal.
2	Demonstrate DoORS in one on one meetings with experienced ontologists	Evaluate Face Validity	Comments made by ontologists showed that the system was useful.
3	Compare DoORS results when run on Good, Average, and Poor quality ontologies identified from literature.	Evaluate level of distinction between scores	T-test analysis showed that the differences between the three sets of scores were statistically significant.
4	Obtain domain expert opinions of which of three ontologies are the most useful for a specific domain	Evaluate modeling of a domain	Experts' ontology preferences corresponded to high scores determined by DoORS.

Figure (7.5) Results of Empirical Evaluation

This evaluation has shown that the Semiotic Ontology Metrics Suite provides useful information for ontology engineers, and that the DoORS interface is easy to apply. Additionally, the speed of the system and the ability to customize the rating parameters, have been shown to be advantageous to users.

CHAPTER 8

METRICS EVALUATING ONTOLOGY USEFULNESS

The Semantic Web is an extension of the World Wide Web in which entities of the web share information and work together without dependence on human intervention (Vandenbussche et al., 2017). For the Semantic Web, these concepts are defined using domain ontologies to model the entities and the relationships between them within a specific knowledge area. Semantic Web pages are tagged with classes from the ontology to allow for interoperability with other web entities. Terms and relationships found in Semantic Web applications use formal ontologies to make the semantics explicit so that the consistency of the knowledge can be assured, contributing to automated reasoning (Nguyen et al., 2014). Interoperability between ontological resources is required to automatically analyze data across different repositories, supporting knowledge discovery kontokostas2014test.

Being able to assess the ability of an ontology to model its domain is dependent on the ontology being of sufficient quality, not only syntactically and semantically, but also pragmatically. Pragmatic quality is the measure of how well an ontology contributes to accomplishing the purposes and goals of an application (**thalheim2014syntax**). When developing or selecting an ontology for a Semantic Web application, it would be helpful to have a means of evaluating candidate ontologies to assess their pragmatic quality in an attempt to ensure that the ontology selected is best fits the goals of the application.

8.1 Semantic Web Interactions with Ontologies

With respect to the Semantic Web, we need domain ontologies that have, in a given vocabulary, the meaning of a term expressed and understood by defining: 1) all the properties that can be used on it; and 2) the types of those objects that can be used as the values of these properties (Yu, 2011). For the Semantic Web, a suitable ontology is needed for a given application. However, selection of ontologies from the many that are available is challenging due, for example, to different ways of representing domain ontologies, and domain ontologies being constructed independently, for different people, using different resources (Gruber, 1993).

There are many advantages to using a domain ontology for a Semantic Web application to model its particular domain. First, domain assumptions are made explicit, allowing for knowledge reuse. Second, the use of an ontology provides a way to encode the knowledge and semantics that a machine can understand, furthering interoperability between systems and making large-scale machine processing easier (Yu, 2011). Data interoperability is facilitated because the use of an ontology as part of a Semantic Web application promotes knowledge reuse and formally represents the knowledge related to a given domain. Web searches are more powerful when web entities use semantic tags to specify term meanings, allowing a search engine to find related concepts and perform reasoning tasks rather than simply searching for specific key terms (Yu, 2011).

Domain ontologies are represented in ontology description languages such as RDFS and OWL that are especially designed to represent the type of complex relationships often found in natural language (Horridge et al., 2006). The OWL language was created to express relationships among classes defined in different documents on the web and to construct new classes based on the unions and intersections of existing classes. The OWL language can also add properties to the terminology used in a web document such as requiring that all members of a class have a particular property, or whether certain properties may not be held by members of a particular class. The knowledge represented in OWL is logic-based so computer programs can interpret the meaning and verify consistency without requiring human interaction (Yu, 2011).

Semantic web entities interact with ontologies through the use of semantic tags assigning meanings to the contents included on the web pages through a process called Semantic Markup. Figure 8.1 illustrates how concepts from an ontology can be added as

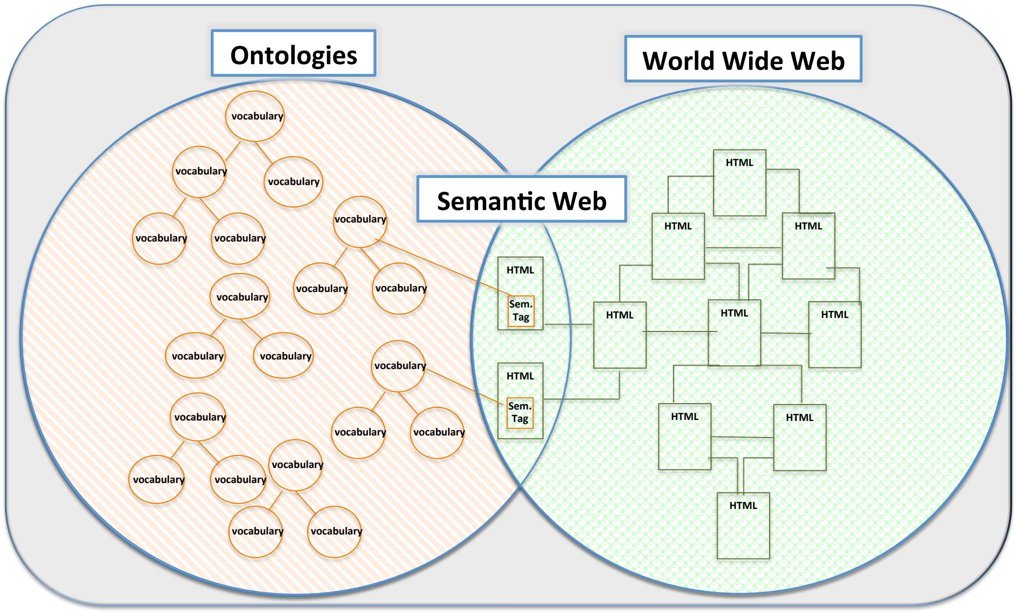

Figure (8.1) Ontologies formalize web page terminology

semantic tags to web pages.

8.2 Pragmatics Defined

Stamper et al. (Stamper et al., 2000) defined pragmatics as a measure of how well signs are able to express the intentions of their user. This definition can also apply to ontologies, which are made up of signs designed to represent concepts in a domain. Thalheim (**thalheim2014syntax**) defines pragmatics as the study of how languages are used for intended functions depending on the purposes and goals within a community of practice. This definition can be expanded to ontologies, which are models of the language used for a specific domain. Together these two definitions express how well an ontology fulfills the intentions of its users. For the Semantic Web, the users are the developers of web applications between which an ontology conveys the intended vocabulary. Assessing the pragmatic quality, therefore, can be thought of as a measurement of how well the ontology fulfills the goals of the applications that employ it, with that goal being interoperability with other applications.

8.3 Evaluating Pragmatics for Domain Ontologies

The assessment of an ontology's quality cannot be separated from the context in which it is intended to be used. For Semantic Web applications, the context is the ability to accurately express shared terms with other web applications or services. Therefore, in this research, three metrics are proposed to assess the pragmatics of an ontology in terms of its usefulness for that context. These metrics, evaluate consistency, coverage and usability, with each measuring an essential aspect of usefulness. Together, they provide an overall evaluation of an ontology's pragmatic quality.

8.3.1 Consistency

An ontology is not useful if it has redundancy or cyclical errors (Weber, 2003). These errors prevent full coverage of a domain because cyclical errors cause some portions of the ontology to be unreachable and other portions to contain more than one conflicting definition.

Figure 8.2 shows an example of an ontology containing a consistency error caused by a class and subclass relationship in which each is in a different disjoint class. In this example, a Soy Sausage Pizza cannot be a subclass of both Sausage Pizza and Soy Pizza because a Soy Pizza is a subclass of Vegetarian Pizza, which is clearly disjoint from a Meat Pizza, and a Sausage Pizza is a subclass of Meat Pizza. These types of inconsistencies are difficult to find, but obviously problematic.

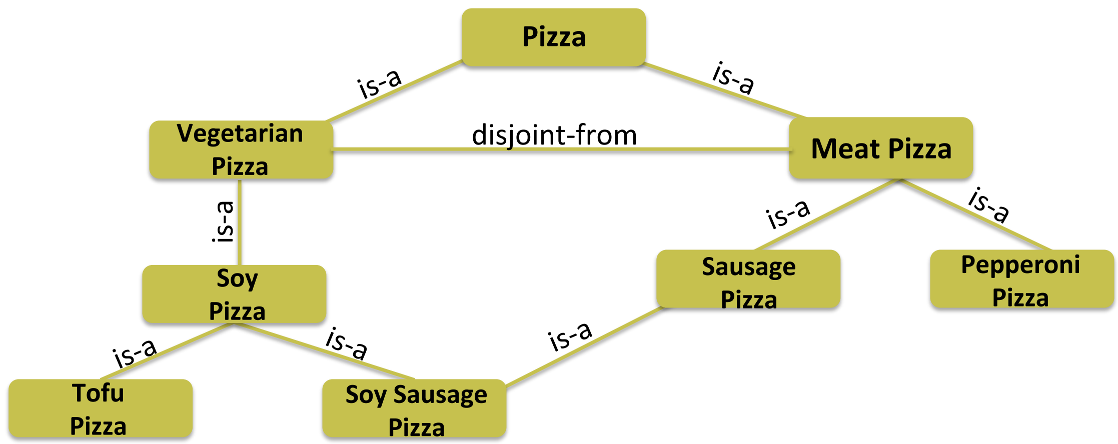

Figure (8.2) Example of consistency error in an ontology

The Consistency metric assesses whether an ontology is free of these type of errors and is computed as the ratio between the number of consistency-error-free relationships and the total number of relationships in an ontology. A perfect consistency score, therefore, would be 1.0. In equation 1, if R represents the total number of relationships in an ontology and E represents the number of consistency errors in the ontology, then: Consistency = (R – E) / R

8.3.2 Coverage

Coverage assesses the balance between covering enough of the domain so that concepts which are part of the domain are not omitted without covering such a broad area that too many irrelevant concepts are also included in the ontology. The sub-metrics for Comprehensiveness and Relevance represent these opposing ideas.

Comprehensiveness is defined as how well an ontology covers all concepts required for a particular domain. In general, it is simply the size of the ontology relative to the size of other ontologies under consideration. Larger ontologies are more likely to cover all the concepts necessary for full coverage of a domain. Let C be the number of classes in this ontology and M the maximum number of classes for an ontology under consideration. Then: Comprehensiveness = C / M

Relevance is the balancing metric to Comprehensiveness in that it assesses whether all concepts of the ontology are relevant to the desired domain. An ontology that receives the highest assessment of relevance, does not include any irrelevant concepts. Let R be the number of classes in an ontology relevant to a set of keywords. Let C be the total number of classes in the ontology. Then: Relevance = R / C

Domain modeling requires that an ontology's coverage is neither too broad nor too narrow, balancing relevance with comprehensiveness to accurately model the desired domain. Figure 3 shows the competing requirements in achieving the optimum concepts to be included in the ontology. It is important that the domain ontology cover the entire domain without missing any concepts, but does not become unwieldy from being overloaded with irrelevant concepts.

The concepts of comprehensiveness and relevance work together to assess how well the balance between too broad and too narrow has been reached; therefore, the assessed values for the metrics of Comprehensiveness and Relevance are weighted to compute the overall value for the Coverage metric. The computation of whether an ontology accurately covers the intended domain is formalized as Coverage = w1 * Comprehensiveness + w2 * Relevance where w1 and w2 are the weights for each of the sub-metrics.

8.4 Ease of Use

Ease of Use assesses the level of annotation in an ontology (Gangemi et al., 2006a). Annotations and comments in ontologies provide: 1) guidance to human users when examining the ontology; and 2) additional information for ontology matching tools to link the data in the ontology. Because of the complexity of language, meta-data is needed to provide insight into the meaning of the terms in the ontology. Ontologies containing little or no annotations are less useful. Figure 4, shows a comment from the Wyner et al.'s Legal Case Ontology (Wyner and Hoekstra, 2012) clarifying the meaning of the term "Issue" in the intended context of the ontology.

Usability, in this research, is assessed by: 1) the number of comments and annotations in the ontology; and 2) the length and placement of those annotations. Longer comments are more useful in expressing ideas (Wyner, 2010). Comments located close to the class they are describing, as opposed to all in one place, are most useful in discerning meaning (Horridge et al., 2006), so both length and placement are considered when computing Usability. Usability is the weighted average of the ratio of comments to classes and the ratio of locations to comments. The best score an ontology could receive on usability is 1.0 which would be achieved when an ontology has a comment or annotation for each class and those comments are evenly distributed throughout the ontology. Usability is calculated as shown in equation 5. Let A represent the number of annotations in the ontology. Let L represent the number of different locations in the ontology that contain an annotation. Let C represent the total number of classes in the ontology. Then: Usability $= w1 * (A/C) + w2 * (L/A)$

8.5 Exploratory Evaluation

To test the validity of the consistency, coverage, and usability metrics for assessing the quality of ontologies, we conducted an exploratory investigation in two phases using experts from four separate domains.

8.5.1 Phase 1

Participants were experts recruited from: building construction, culinary arts, law, and mathematics. These domains were chosen to incorporate unique terminology and to allow for the different types of tasks that might be required in Semantic Web applications. The participants were given two tasks: 1) rank ontologies related to their domain; and 2) rank the relative importance of each of the metrics in making these rankings. The domain experts were given a simple questionnaire on the metrics that identified the ontologies under consideration and provided definitions for the three attributes. During phase 1, the researcher took notes of the time the experts spent examining each ontology and the comments that were made.

Task 1. The participants were each shown several ontologies related to their area of interest and asked which one they would consider using if they were planning to create a web application and wanted to define the terms. They were encouraged to talk aloud as they compared the ontologies using the Prot'eg'e application (**noy2001creating**) with the OntoGraf (Falconer, 2010) visualization plug-in. They were instructed to rank the ontologies in order of preference and to explain their reasonings for the ranking.

Task 2. The participants were then asked to consider the three pragmatics metrics and asked whether these attributes (consistency, coverage and usability) were qualities important for the selection of an ontology. They were instructed to rank the metrics in order of importance in making their assessment of which ontology would be preferred by them if they were creating a Semantic Web application.

Phase 1 Results. All four of the experts were in agreement that the metrics would provide useful information about ontology quality, but the relative importance of each differed greatly among the domains. Each of the attributes for consistency, coverage, and usability was found to be of paramount importance by at least one expert, but also considered to be of very minor importance by an expert in a different domain.

The Building Construction expert, for example, considered that coverage of the domain is the most important attribute. In particular, not only should all terms be included, but the addition of new terms and features should be easy to perform. In the Building Construction field, the relationships between features and materials is continually changing as new materials and new purposes for those fields continue to be explored. He did not consider annotations to be very necessary in the Building Construction field. He considered consistency important, but not as important as coverage of the domain and the ability to add new terms.

The legal domain expert agreed with how each of the three metrics were obtained

Table (8.1) **Results of Assessment Metric Importance Rankings**

Domain	**Consistency**	**Coverage**	**Ease of Use**
Building Construction	2	1	3
Culinary Arts	3	1	2
Law	2	3	1
Mathematics	1	3	2

and stated that, for his domain, the inclusion of annotations and comments is essential. He considered this to be more important than coverage and consistency because the precise meaning of legal terms is often a subject of dispute between legal experts. Therefore, it is essential that any term be clearly defined and examples be included from actual court cases where possible. Consistency was less important because, with the disagreement between meanings of terms in the legal field, it is not unlikely that consistency errors may be found, even in carefully built ontologies. One challenge is the fact that different legal systems, for example between the United States and the United Kingdom, use different legal terms to mean different things. Therefore, since one term could have a completely different meaning in another legal system, ontologies must be well-annotated in addition to being well-designed.

The mathematics expert considered consistency to be the most significant attribute for an ontology because the mathematics field allows for no possibility of misconstrued terms. The culinary arts expert, on the other hand, rated consistency of minor importance because what is meant by different cooking terms can vary between types of cuisine.

Table 2 shows the disparity between the rankings of metric importance mentioned by the domain experts. For each domain, a value of one indicates the most important attribute, while a value of three indicates the least important attribute.

8.5.2 Phase 2

Given the varying importances of the metrics, an additional phase of testing was needed to ensure that the three pragmatics metrics would be able to accurately assess

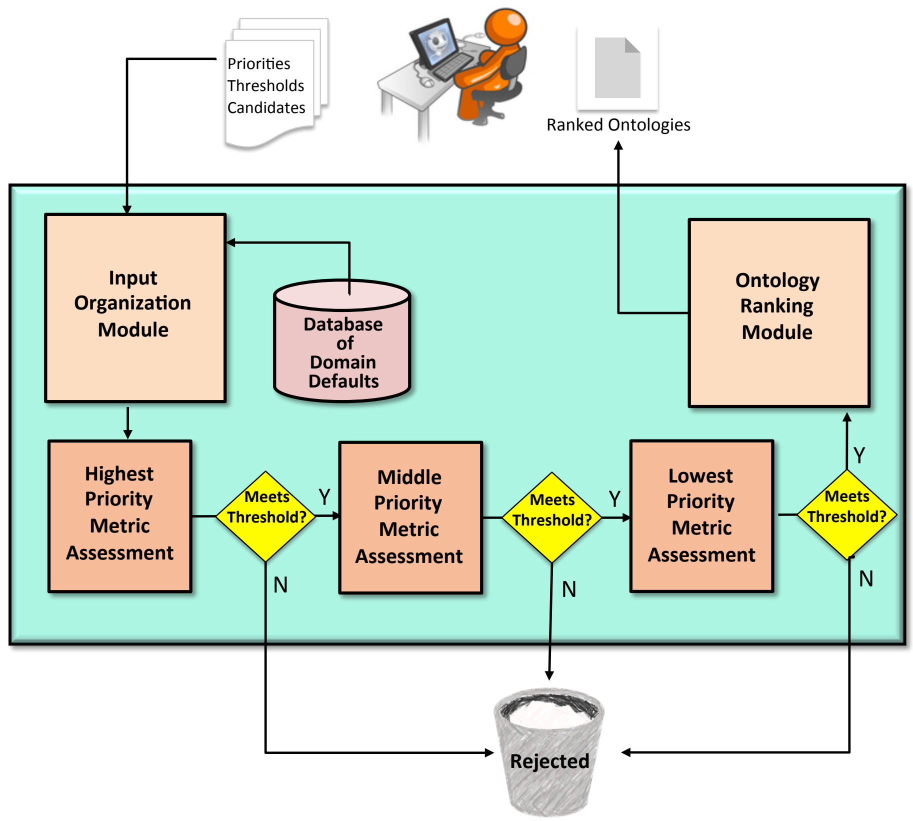

Figure (8.3) Architecture of Ontology Pragmatics Assessment (OPA) framework

the quality of an ontology if their computations were performed in the order of priority deemed most applicable for a specific context. To perform these determinations, an Ontology Pragmatic Assessment (OPA) framework was developed, as shown in Figure 8.3. The framework's purpose is to take a set of ontologies to be considered for a domain and assess the ontologies by applying the three pragmatic metrics in the priority order input by the user, assuring that, at each stage of the assessment, an acceptable level of quality is reached.

8.6 Framework Architecture

The OPA framework accepts input in the form of a set of domain ontologies for consideration, the list of the three metrics with assigned threshold and weighting values, and a set of one or more keywords identifying a specific domain. At each step of the framework an ontology may be rejected if it does not meet an acceptable standard for that metric. as shown in the lower portion of figure 5. The metrics are calculated in order of highest to lowest priority. Each of OPA's modules serves a distinct purpose in the assessment process.

Input Organization Module This module accepts four items of input from the user: 1) a set of ontologies stored in OWL format, 2) a set of importance weights for the three pragmatics metrics, and 3) a set of three threshold values for the quality level associated with each of the pragmatics metrics, and 4) a set of keywords representing the desired domain. This module checks the input data and supplies default values where needed, and then passes these values to the metric assessment modules to ensure that the system evaluates the metrics correctly.

Priority Metric Assessment Modules These three modules assess the quality of the candidate ontologies using the priorities and thresholds received from the Input Organization Module. At each module, the pragmatic metric of that priority is computed and compared to the threshold value for an acceptable level of quality on that metric. If any of the ontologies receive a score less than the acceptable level the ontology is rejected.

Ontology Ranking Module This module determines a final score for each of the non-eliminated ontologies corresponding to the weighted average of the results from each of the three assessment modules. The module then produces a ranking list by score order that is given as output to the user.

Phase 2 Method To test the OPA framework, the original sets of candidate ontologies examined by the four domain experts served as input. Additional information was input in the form of the priority weights of the three metrics as determined by the domain

Table (8.2) **Steps for Framework Development and Evaluation**

Domain expert is shown a set of candidate ontologies and three metrics
Domain expert ranks the metrics in order of domain importance
Domain expert ranks the ontologies in order of preference
Framework is developed to evaluate the ontologies based on priority order
Candidate ontologies are entered into the framework for evaluation
System results are compared to the domain expert's results

experts, and the threshold values input by the domain experts. For each domain, the ontology selected by the system matched the ontology the expert had preferred during manual examination of the ontologies. Table 8.2 summarizes the steps followed in the framework's development and evaluation.

Phase 2 Results In all instances, the ontology selected by the system corresponded to the choice made by the domain expert during the examination of the ontologies. For example, the Building Construction expert selected the FreeClassOWL building materials ontology (Radinger et al., 2013) over the other construction and building material ontologies considered. The OPA framework, by weighting Coverage higher than both Consistency and Usability as designated by input selections, resulted in the same selection after applying the metrics.

As noted by Bertossi et al. the assessment of the quality of a data source is context dependent. That is, the notions of "good" or "poor" data cannot be separated from the context in which the data is produced or used (Bertossi, Rizzolo, and Jiang, 2010). Although the context of the development of a Semantic Web application is obviously important for providing an assessment of "goodness", this research shows that the domain of discourse is also important. What constituted a high-quality ontology for one domain could differ significantly from what would constitute a high-quality ontology for a different domain.

CHAPTER 9

METRICS EVALUATING ONTOLOGIES WITHIN COMMUNITIES

The Semantic Web is "a set of standards for knowledge representation and exchange that is aimed at providing interoperability across applications and organizations" (Berners-Lee and Kagal, 2008). The degree of this interoperability between human and software agents depends upon how many communities they have in common and how many ontologies they share (Berners-Lee and Kagal, 2008). An ontology, which has been called the third component of the Semantic Web, is defined simply as a group of consistent and related terms (Berners-Lee and Kagal, 2008) and more formally as "a formalization of a shared conceptualization" (Gruber, 1995). The latter definition, and the idea that the conceptualization is "shared" is expanded further by Hepp et al. (2006) who asserted that "ontologies are not just formal representations of a domain, but much more community contracts about such formal representations". (Hepp, Bachlechner, and Siorpaes, 2006).

A community consists of a set of relationships between people sharing a common interest (Andrews, 2002). An online community can then be considered as a community that employs the Internet for communication among its members (Andrews, 2002). Berners-Lee and Kagal described the Semantic Web as composed of overlapping online communities of varying sizes and fractal in nature, as membership in these communities changes frequently (Berners-Lee and Kagal, 2008). Many online communities allow members to participate fully in the site through contributing and accessing information, as well as by commenting on the information added by other members. The BioPortal ontology repository (Noy et al., 2009), for example, considers anyone who uses this portal to be a member and allows them to actively contribute to the content in the library—a fact that its designers claim should increase the quality of that content (Noy, Griffith, and Musen, 2008).

This feeling of shared responsibility within a community for the overall improvement of the ontological content is consistent with what Shadbolt and Berners-Lee have asserted will greatly reduce the effort involved in developing an ontology as the size of the community grows (Shadbolt, Berners-Lee, and Hall, 2006). Noy et al. contend that the Wisdom of the Crowd could even replace knowledge experts when a consensus is able to be reached within a community (Noy, Griffith, and Musen, 2008). Reaching this consensus, however, is not always easy, requiring time and effort, and a large number of dedicated participants. Therefore, the degree of participation in the process of revising, adopting, expanding and reviewing of any ontology is a factor in the assessment of that ontology's value.

The selection of an ontology from among the options available in an ontology repository should be made based upon a broad set of attributes that may be weighted depending upon the requirements of each application (Vrandečić, 2009). One of the attributes to include in such a list of criteria should be the acceptance of the ontology within its community. Metrics to assess this acceptance should include measures of how many community members endorse the ontology, how long the ontology has been available, how much active participation has been done by community members in the ontology's development. This community acceptance attribute is difficult to assess, with metrics to measure it not applied successfully in the past (Burton-Jones et al., 2005). While much work has been carried out developing metrics related to syntactic, semantic and pragmatic aspects of ontologies, the social quality of ontologies has not been thoroughly investigated. The objective of this research, therefore, is to do so.

This research introduces new metrics for social quality assessment, defines them formally, applies them to existing ontologies, and analyzes the challenges involved in using them. The result is to show how these attributes provide valuable insight into ontology quality and should, therefore, be included in any rigorous ontology evaluation. The results of this assessment could promote interoperability between systems and help progress the use of ontologies in the Semantic Web.

Term	*Definition*	*Reference*
Authority	"The degree of reputation of an ontology in a given community or culture"	Stvilia et al. [8]
Community	"A set of relationships where people interact socially for mutual benefit."	Andrews [4]
History	"The way that a particular subject or object has developed or changed throughout its existence"	History [9]
Online Community	"A social network that uses computer support as the basis of communication among members instead of face-to-face interaction"	Andrews [4]
Revising	"The act of thinking, comparing, deciding, choosing then taking action"	Sudol [10]
Revision	"The act of making changes to a written document to make it better"	Horning and Becker [11]
Social Network	"A set of people (or organizations or other social entities) connected by a set of socially-meaningful relationships"	Wellman [12]
Social Quality	"The level of agreement among participants' interpretations"	Su and Ilebrekke [13]

Figure (9.1) Definitions of terms related to social quality assessment

Terms related to social quality assessment used in this chapter are defined in the table represented by Figure 9.1.

9.1 Ontology Role in Communities

D'Aquin and Noy (2012) defined an ontology library as "a Web-based system that provides access to an extensible collection of ontologies with the primary purpose of enabling users to find and use one or several ontologies from this collection" (d'Aquin and Noy, 2012). Although ontologies should reside in libraries and be developed and endorsed by communities that share a common interest (Shadbolt, Berners-Lee, and Hall, 2006), little work has been conducted to develop a means for assessing the amount of recognition received by each ontology within a library. To provide a comprehensive picture of an ontology's quality, factors such as how much the ontology is being used, how many other ontologies refer to this one as an authority, and how long the ontology has been in existence, should all be taken into consideration (Burton-Jones et al., 2005).

A community can no longer be considered as a physical place, but, rather, as a set of relationships between people who interact socially for their mutual benefit (Andrews, 2002). An online community is a social network that uses the Internet to facilitate the communication among its members rather than face-to-face meetings (Andrews, 2002). These virtual social networks are frequently used for information sharing and problem solving among members who share common interests (Wellman, 1997).

Ontologies have been defined as formal representations of a domain, but in order for those representations to be meaningful, they must be agreed upon by the members of a community (Shadbolt, Berners-Lee, and Hall, 2006). This type of meaningful discourse between members of a group is a dynamic social process consisting of shared topics being added, expanded, revised or even discarded. Therefore, an ontology representing the shared communication between members should not be static, but should be able to reflect the community consensus of meaning at any particular time (Berners-Lee and Kagal, 2008). When a community shows its approval of an ontology by actively participating in its ongoing evolution, the quality of the ontology is more likely to be high within that community (d'Aquin and Noy, 2012). A way of measuring this type of active participation would be helpful in assessing community endorsement of a particular ontology.

Although communities should support the development, maintenance and endorsement of ontologies (Shadbolt, Berners-Lee, and Hall, 2006), very few assessment systems have a means by which to measure an ontology's value within its community. OntoMetric (Lozano-Tello and Gómez-Pérez, 2004), the BioPortal Recommender (Jonquet, Musen, and Shah, 2010), and the Semiotic Metrics Suite (Burton-Jones et al., 2005) are among the few suites that attempt to assess an ontology's acceptance within a community as one of the factors to measure its quality. Unfortunately, none of these assessment suites are able to fully evaluate the level of acceptance an ontology receives within its community.

OntoMetric (Lozano-Tello and Gómez-Pérez, 2004) contains approximately 160 metrics for assessing ontology quality, which focus primarily on the fitness of an ontology for a particular software project for which it will be used. However, only three of its metrics relate to its relation-ship with other ontologies. The large number of metrics makes the OntoMetric system difficult to employ (Tartir et al., 2005). The OntoMetric system reflects the fact that part of the suitability of an ontology for a given project is the methodology used to create it. It, therefore, assesses the social acceptance of that methodology by counting the number of other ontologies that were created with it, the number of domains that have been ex-pressed with its developed ontologies, and how important the ontologies developed with this methodology have become. Unfortunately, in most situations, a user must attempt to answer these questions (perhaps by conducting additional research) as well as to provide an answer expressed on a scale between “very low” and “very high,” reducing the accuracy of the results in this factor’s assessment.

The BioPortal Recommender system includes Acceptance metrics as part of the ranking system that it provides as a tool for choosing an ontology for a particular purpose (Jonquet, Musen, and Shah, 2010). Users enter desired keywords and the recommender system presents a list of ontologies from the BioPortal repository containing the keywords. The list of applicable ontologies is ranked in order of each ontology’s score on four individually weighted attributes, one of which is the Acceptance of the ontology within the BioPortal community. The other three attributes that are included in the Recommender system are Coverage, Detail of knowledge and Specialization. Unfortunately, the metrics used by the BioPortal recommender system to assess Acceptance are based on factors such as the number of site visits to the BioPortal website, membership in the UMLS database and mentions in the BioPortal journal, so those metrics cannot be used on ontologies in other libraries without access to this information.

The Semiotic Metrics Suite developed by Burton-Jones et al. (Burton-Jones et al., 2005) is based upon the theory of semiotics, the study of signs and their meanings, and builds upon Stamper et al.’s (Stamper et al., 2000) framework for assessing the quality of signs. One of the layers of the framework is the Social layer, which evaluates a sign’s usefulness on a social level by evaluating its “potential and actual social consequences” and asks the question "Can it be trusted?" (Stamper et al., 2000). The Semiotic Metrics Suite includes the Social layer, which measures an ontology's recognition within a community by two metrics: 1) Authority which measures the link from an ontology to other ontologies in the same library; and 2) History which measures the frequency with which these links are employed. Unfortunately the calculations for these measurements require information that is not available for most ontologies. The number of links from other ontologies to a particular one, and the number of times the linking ontologies have been used for other applications are usually not provided by ontology libraries, making these metrics difficult to use for ontology assessment. This research introduces new Authority and History metrics using information available for most ontologies and includes a case study demonstrating their effectiveness.

9.2 Social Quality metrics

Social Quality is "the level of agreement among participants' interpretations" (Stamper et al., 2000) and reflects the fact that, because agents and ontologies exist in communities, agreement in meaning is essential within the community. This research proposes two new metrics to measure the level of an ontology's recognition within its community by measuring its authority within the library and the history of its participation and use in the library. These metrics can be combined to determine the overall assessment for Social Quality within the library.

Stvilia defines Authority as the "degree of reputation of an ontology in a given community or culture" (Stvilia et al., 2007). One way to measure Authority is by the number of other ontologies that link to it as well as how many shared terms there are within those linked ontologies. More authoritative ontologies signal that the knowledge they provide is accurate or useful (Burton-Jones et al., 2005).

Another social metric is the History of an ontology. The history of a conceptualization is a valuable part of its definition (Hepp, Bachlechner, and Siorpaes, 2006). The History metric measures the number of years an ontology has existed in a library, as well as the number of revisions made to it during the course of its residence there. Ontologies with longer histories are expected to be more dependable because each new revision should improve upon the previous version showing a pattern of active participation by community members resulting in additions and modifications.

9.3 Overall Social Quality metric

The Social Quality metric is computed by the combined weighted scores on these two measurements defined as SQ. The weights of the History and Authority metrics could be equivalent, but it is possible for a user to adjust the significance of each for a particular task by varying the values of the weights.

Definition 1: The Social Quality (SQ) of an ontology is defined as the weighted average of Authority (SQa) and History (SQh) where wa represents the percentage assigned by the user to the authority attribute and wh represents the weight assigned to the history attribute. $SQ = w_a * SQa + w_h * SQh$

9.4 Social Authority metric

The Authority of a particular ontology is determined by the number of other ontologies that link to it. By scanning all of the other ontologies in the library looking for links to this ontology, two counts are determined: the number of total links to the ontology; and the number of ontologies which include 1 or more references to it. The two counts are weighted depending on the user's task and the result is normalized between 0, meaning no links at all, to a score of 100, indicating that this ontology is the one in the library with the most links to it. The equation for computing this metric is defined as SQA. External links can also be considered in the determination of SQA if available. Many ontologies, such as the Gene Ontology (Ashburner et al., 2000), are in multiple libraries. SQA should then take into consideration all of the links to the Gene Ontology from all of the libraries for which it is a part.

Definition 2: The Social Quality Authority (SQA) of an ontology is defined as the weighted average of the number of linking ontologies (LO) and the number of total linkages (LT) where w_o represents the percentage assigned by the user to the number of linking ontologies and wt represents the weight assigned to the total number of links. SQA $=w_o * LO + w_t * LT$

9.5 Social History metric

History is determined by calculating the number of years that an ontology has been a member of a community as well as the number of revisions to the ontology that have been made during those years. The two counts are weighted depending on the user's task and the result normalized between 1, indicating only one submission that was never updated, to a score of 100, indicating this ontology is the one in the library with the most total revisions over the longest number of years.

Definition 3: The Social Quality History (SQH) of an ontology is defined as the weighted average of the number of years it has been in the library (Y) and the number of submissions (including revisions) that have been uploaded (S) where w_y represents the percentage assigned by the user to the number of years and w_s represents the weight assigned to the total number of submissions. SQH $=w_y * Y + w_s * S$

9.6 Implementation

A prototype has been developed to assess community recognition of an ontology by applying the revised Social Quality metrics. This system can be employed by any community containing an ontology repository, and aids in the selection of an ontology when multiple options are available. By entering relevant keywords and desired metric weights into the system, a user retrieves a set of potential ontologies containing the keywords. The system then assesses the Authority of each of those ontologies by searching all the other ontologies in the repository counting the number of ontologies that link to each of the potential ontologies as well as the total number of links. Each ontology in the list of potential ontologies then has its History assessment computed by counting the number of years each ontology has been stored in the library and the number of revisions made to the ontology during that time. The Authority and History metrics are then weighted according to the metric weights entered by the user and the list of potential ontologies is sorted in decreasing order of the overall Social Quality score. At this time the user receives a list of recommended ontologies that contain the desired keywords and that rank high in social recognition from the community.

9.7 Case Studies

To obtain an understanding of how information about an ontology's acceptance within its community could help a user choose an appropriate ontology from a list of options, two case studies were carried out using the social quality metrics. The BioPortal ontology library was chosen for both studies as an example of a large, well maintained ontology repository that has been deemed useful to the biomedical community (Noy et al., 2009). The BioPortal website was also selected because of the availability of additional information included in the library that could be used to examine the results of the case studies with other information on its ontology profile pages. The BioPortal web-site allows members of the community to contribute reviews to its ontologies, list projects, and make suggestions. BioPortal also keeps track of the number of site visits for each of the ontologies, and provides annotation and term mappings services for its ontologies (Noy, Griffith, and Musen, 2008).

The first case study applied the social quality metrics to all 383 of the ontologies currently in the library, ranking them from highest Social Quality score to lowest. This case study was carried out to assess whether the highest-ranking ontologies in the library were in actuality the ones that were most endorsed by the biomedical community. The second case study searched the BioPortal library for ontologies matching key terms and determining a list of recommended ontologies ranked by their Social Quality assessments as well as using our SQ metric. The ontology list for each term was then examined to ascertain whether the highest-ranking ontologies on each list was actually more likely to be frequently accessed than the ontologies that showed up later on the list. In both case studies, all metrics were weighted equally in the over-all determination of Social Quality. It is possible to weight the individual metrics differently, depending on the particular task requirements. However, for the purposes of the case studies, all metrics were considered equally.

The two case studies showed that useful information could be obtained from assessing the ontologies on their level of endorsement within the BioPortal community. By examining the difference between ontologies high on the list to the ontologies that ranked lower, a pattern can be easily observed about whether the ontologies are well-supported by the BioPortal membership.

9.7.1 Case Study 1

The Authority metrics were first applied to all 383 of the ontologies currently part of the BioPortal library assigning equal weights to the number of links and the number of linking ontologies. For each ontology, all other ontologies were scanned for references to that ontology and counts made of the number of ontologies that included at least one reference to the ontology, as well as the total number of links to the ontology.

The History metric was then computed for all of the BioPortal ontologies using equal weighting for the number of years that the ontology has been in the library and the number of revisions that have been done to each of them, including the original submission.

The scores for Authority and History were individually normalized between 1 and 100 and then the two scores averaged to generate the overall Social Quality metric for each of them. The list of ontologies was ordered from 100 to 1 in order to identify the most highly ranked ontologies at the top of the list.

Figure 9.2 shows the highest ranked ontologies from the library on the combined social metrics. Examining other information available on the BioPortal website, it is clear

Table 8.1. Highest ranked ontologies from BioPortal using History and Authority metrics

Name of Ontology	*Authority*	*History*	*Combined Metrics*
Gene Ontology (GO)	69	100	85
Human Phenotype Ontology (HP)	51	100	75
Mosquito Insecticide Resistance Ontology (MIRO)	17	100	58
Mass Spectrometry Ontology (MS)	85	27	56
Systems Biology Ontology (SB)	10	100	55
Minimal Anatomical Ontology (MAT)	100	2	51
Sequence Types and Features Ontology (SO)	60	42	51
Human Disease Ontology (HD)	1	100	51
Mammalian Phenotype Ontology (MP)	38	61	50
Plant Trait Ontology (PTO)	7	79	43

Figure (9.2) Social rankings of BioPortal ontologies

that the ontologies that scored high on the Social Quality metric were the ones involved in many biomedical projects and with good reviews from members who have used them. On the other hand, 70 of the ontologies tested scored only 1 out of 100 on the combined social quality metrics. Exploring the BioPortal website revealed that these 70 had no other ontologies linking to them and no revisions after the initial submission, which was often several years prior, and were not currently involved in any listed projects.

9.7.2 Case Study 2

The BioPortal repository was searched for ontologies containing each of ten preselected keywords. A list of applicable ontologies was generated for each of the key-words, and each ontology's Social Quality score determined by applying the method outlined in Case Study 1. Each potential ontology list was sorted in descending order to identify the highest results for each of the terms based upon social quality. The keyword searches each retrieved at least 30 potential ontologies. The results of ranking these ontology lists in reverse order of Social Quality was used to identify the best candidates for a possible

Table 8.2. Top three ontologies recommended for each term with corresponding SQ scores.

Keyword	*Number of Potential Ontologies*	*Highest Scoring*	*Second Highest*	*Third Highest*
cell	47	CPT (53)	AURA (35)	FLU (29)
temperature	39	MAT (52)	ACGT-MO (51)	MA (40)
disease	46	ACGT-MO (51)	NMR (43)	AURA (35)
blood	37	MAT (52)	ACGT-MO (51)	AURA (35)
cortisol	30	CPT (53)	MESH (9)	PMA (21)
sucrose	33	AURA (35)	DDO (34)	PR (26)
water	45	AURA (35)	MESH (5)	CCONT (25)
patient	45	AERO (51)	ACGT-MO (51)	CTX (27)
oxygen	31	DCM (37)	AURA (35)	MESH (33)
cerebrum	30	MA (40)	CTX (37)	MESH (33)

Figure (9.3) Top three ontologies recommended with corresponding SQ scores

task requiring each of the keywords. The top three ontologies recommended for each of the keywords are shown in 9.3.

Additional information provided by the BioPortal website showed that these listed ontologies are favorably reviewed and frequently accessed. In comparison, ontologies retrieved by the keyword search but scoring low on the Social Quality metric, were accessed infrequently, indicating little use within the community. For example, the Current Procedural Terminology ontology (CPT), which ranked highest for two of the keywords and obtained an SQ score of 53, received over 35,000 site visits in the last two years. In contrast, the Bone Dysplasia Ontology (BDO) containing the same two keywords, received an SQ score of 2 and only received 894 site visits.

9.8 Social quality assessment summary

This chapter has introduced two metrics for assessing the authority and history of an ontology within a community and illustrated their effectiveness by applying them to approximately four hundred of the ontologies in the BioPortal library. Results from that case study showed that application of the metrics was feasible and provided useful information regarding ontology recognition within its community.

Future work will consider other factors such as the number of times an ontology has been viewed or downloaded; user comments and rankings of ontologies; and the usability of ontologies to gain a more comprehensive view of the social quality metric. In addition, the social quality metrics need to be incorporated into broad metric suites that assess various attributes of ontologies. When users select an ontology from a number of options, a broad overview is required that considers syntax, semantics, pragmatics, as well as social acceptance, to make an appropriate recommendation to a user. Therefore, the Social metrics will be included in the DoORS web application.

CHAPTER 10

CONCLUSIONS

Motivated by the growing number of domain ontology applications, this research has attempted to find a way to easily assess ontology quality in order to improve the overall quality of ontologies available and to make the ontology selection process easier for potential users. Unfortunately, determining exactly what makes an ontology good or bad is not an easy problem to solve, as it is dependent on the task for which the ontology is to be used. Although I was unable to completely solve the ontology evaluation and selection problem, I have achieved success in several areas. This chapter summarizes the contributions of this thesis in the next section and identifies a few of the remaining issues and challenges pertaining to ontology evaluation in the subsequent section. The final section wraps up the thesis by addressing future steps.

10.1 Contributions

This thesis has proposed and developed a semiotic based ontology suite that can be used to assess the quality of domain ontologies. The suite of metrics has been formalized and implemented in a ranking system composed of modules to retrieve the ontologies from an online repository, assess them by employing the ontology suite, and rank them based on user quality input selections. The system has been testing in a multi-level empirical evaluation to prove its usefulness. The evaluation of the system that has been done has compared DoORS to existing ontology evaluation systems, compared DoORS to human domain experts, and compared the DoORS results when processing ontologies of varying quality. Advantages of the newly developed ranking system are that is scalable for large ontologies, that it is easily customized by users for specific needs, that it is automated to provide interoperability between software applications, and that it is available as a web application to make it easy to use. The motivation for this research, has been to facilitate ontology reuse rather than unnecessary new development, thus furthering domain ontology use overall.

In addition to the domain ontology ranking system, this research has also led to the development of a Taxonomy of Domain Ontology Assessment research directions that categorizes the many directions this area of research has taken. Additionally, this research has presented a framework for domain ontology assessment that takes a pipeline approach to solving the domain ontology selection problem by employing examples from all of the six directions that current domain ontology assessment research has taken.

The contribution of this research is to analyze and quantify quality assessment metrics for domain ontologies and to show the feasibility of automated domain ontology evaluation. The Domain Ontology Ranking System is able to retrieve ontologies from online repositories and evaluate their quality by applying a suite of assessment metrics. Testing showed that the results obtained from doing so provide similar results in terms of rankings to that of domain-specific assessment systems, but is broad enough to apply to almost all domains. The availability of the DoORS web application for knowledge engineers, software developers and other ontology users can further the use and reuse of domain ontologies. The selection of an ontology to use for a specific task is difficult and time-consuming, therefore the automated system should be useful in assisting in this task as well as assisting with ontology development and application.

10.2 Future Research

As access to information and computer resources continues to expand, the expectation of computer systems to seamlessly communicate is essential. 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 from the real world. These stocks of knowledge, defined as accumulated knowledge as-sets (DeCarolis and Deeds, 1999), may be represented in a variety of forms. The entire field of knowledge representation recognizes such difficulties (Borgo et al. 1996; Guarino, 1995; Guarino, 1997; Sowa 2000).

Other difficulties 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; J. Park et al. 2007). Furthermore, many ontologies are used successfully even though they may lack in consistency or coverage which should be the ultimate evaluation of an ontology, but which is difficult to assess (Obrst et al. 2007, page 18).

A few challenges related to the characteristics of a domain ontology that should be assessed were unable to be solved by this, or other, research. Some ontology of a domain ontology, for example, accuracy, although important, are not easily assessed without human domain experts. Other attributes, such as Ease of Use, are task specific therefore a broad metrics evaluation system, particularly one that is automated, cannot easily incorporate metrics to assess them.

Other challenges 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), understanding the importance of evaluation (Neuhaus et al., 2013). Improving semantic interoperability needs to be recognized by other types of engineers when developing domain specific systems such as components of the Internet of Things (Da Xu, He, and Li, 2014). A few ontology repositories exist, but many more are needed, and the support of the community of ontology users should strive to ensure that high quality ontologies are made available.

10.3 Concluding Remarks

Although significant work has been carried out on creating ontologies, it is important to be able to assess the quality of ontologies in a systematic way. Doing so can help developers to select an ontology from among available choices or to create their own, to obtain ontology-task 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 thesis has reviewed and classified research on ontology evaluation, identifying challenges that still to be overcome as ontologies continue to be integrated into information system applications that depend upon domain knowledge. A framework has been proposed that illustrates how combining evaluation methods could assist in solving one of the problems related to domain ontology evaluation. The approach of metrics development for domain ontology assessment has been pursued by implementing a domain ontology rating system that can assess the quality of domain ontologies automatically, and the software has been made available for other researchers to explore.

The web application that has been done as part of this thesis still needs further testing to ensure that it can be applied to even more tasks and used to represent other domains. Including natural language processing to the web interface in the form of a text box describing the task to be performed would improve the value of the system. I will continue to work on this and hope that the fact that the source code is made available, will encourage other researchers to test it with their ontologies.

A framework of how this rating system could be part of a larger framework for retrieving domain ontologies from repositories and assessing them to aid users in making an ontology selection has been proposed. Future work will consist of implementing the ontology ranking framework also as an open source web application with minimal human intervention in order to expand system interoperability among applications.

Bibliography

[1] Hassan Abolhassani, Babak Bagheri Hariri, and Seyed H Haeri. "On ontology alignment experiments". In: *Webology* 3.3 (2006), pp. 1–22.

[2] Manel Achichi et al. "Results of the ontology alignment evaluation initiative 2016". In: *CEUR workshop proceedings*. Vol. 1766. RWTH. 2016, pp. 73–129.

[3] Maha Al-Yahya et al. "An ontological model for representing semantic lexicons: an application on time nouns in the holy Quran". In: *Arabian Journal for Science and Engineering* 35.2 (2010), p. 21.

[4] Harith Alani, Christopher Brewster, and Nigel Shadbolt. "Ranking ontologies with AKTiveRank". In: *International Semantic Web Conference*. Springer Berlin Heidelberg. 2006, pp. 1–15.

[5] Harith Alani et al. "Automatic ontology-based knowledge extraction from web documents". In: *IEEE Intelligent Systems* 18.1 (2003), pp. 14–21.

[6] Dorine C Andrews. "Audience-specific online community design". In: *Communications of the ACM* 45.4 (2002), pp. 64–68.

[7] Grigoris Antoniou and Frank Van Harmelen. "Web ontology language: Owl". In: *Handbook on ontologies*. Springer Berlin Heidelberg, 2004, pp. 67–92.

[8] Michael Ashburner et al. "Gene Ontology: tool for the unification of biology". In: *Nature genetics* 25.1 (2000), pp. 25–29.

[9] Tim Berners-Lee, James Hendler, Ora Lassila, et al. "The semantic web". In: *Scientific american* 284.5 (2001), pp. 28–37.

[10] Tim Berners-Lee and Lalana Kagal. "The fractal nature of the semantic web". In: *AI magazine* 29.3 (2008), p. 29.

[11] Diego Berrueta et al. "Best practice recipes for publishing RDF vocabularies". In: *Working draft, W3C* (2008).

[12] Leopoldo E Bertossi, Flavio Rizzolo, and Lei Jiang. "Data Quality Is Context Dependent." In: *BIRTE* 2010 (2010), pp. 52–67.

[13] Janez Brank, Marko Grobelnik, and Dunja Mladenić. "Automatic evaluation of ontologies". In: *Natural Language Processing and Text Mining*. Springer, 2007, pp. 193–219.

[14] Janez Brank, Dunja Mladenic, and Marko Grobelnik. "Gold standard based ontology evaluation using instance assignment". In: *Workshop on Evaluation of Ontologies for the Web, EON*. Edinburgh, UK. 2006.

[15] Christopher Brewster et al. "Data driven ontology evaluation". In: (2004).

[16] P Buitelaar et al. "A task-based framework for ontology learning, population and evaluation". In: *Ontology Learning from Text: Methods, Evaluation and Applications* 123 (2005), p. 107.

[17] Andrew Burton-Jones et al. "A semiotic metrics suite for assessing the quality of ontologies". In: *Data & Knowledge Engineering* 55.1 (2005), pp. 84–102.

[18] Pier Luigi Buttigieg et al. "The environment ontology: contextualising biological and biomedical entities". In: *Journal of biomedical semantics* 4.1 (2013), p. 43.

[19] Balakrishnan Chandrasekaran, John R Josephson, and V Richard Benjamins. "What are ontologies, and why do we need them?" In: *IEEE Intelligent Systems and their applications* 14.1 (1999), pp. 20–26.

[20] Michelle Cheatham et al. "Results of the ontology alignment evaluation initiative 2015". In: *10th ISWC workshop on ontology matching (OM)*. No commercial editor. 2015, pp. 60–115.

[21] Software Engineering Standards Committee et al. *IEEE Standard for a software quality metrics methodology, Std. 1061-1998*. Tech. rep. Technical Report, 1998.

[22] Li Da Xu, Wu He, and Shancang Li. "Internet of things in industries: A survey". In: *IEEE Transactions on Industrial Informatics* 10.4 (2014), pp. 2233–2243.

[23] Kathleen Dahlgren. "A linguistic ontology". In: *International journal of human-computer studies* 43.5-6 (1995), pp. 809–818.

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

[25] Jeremy Debattista, Sören Auer, and Christoph Lange. "Luzzu–A Framework for Linked Data Quality Assessment". In: *Semantic Computing (ICSC), 2016 IEEE Tenth International Conference on*. IEEE. 2016, pp. 124–131.

[26] Donna Marie DeCarolis and David L Deeds. "The impact of stocks and flows of organizational knowledge on firm performance: An empirical investigation of the biotechnology industry". In: *Strategic management journal* (1999), pp. 953–968.

[27] R Dividino. "Semiotic-based ontology evaluation tool". PhD thesis. MasterâĂŹs thesis, Universität des Saarlandes, 2007.

[28] Raimundo F Dos Santos et al. "The big data of violent events: algorithms for association analysis using spatio-temporal storytelling". In: *GeoInformatica* 20.4 (2016), pp. 879–921.

[29] Astrid Duque-Ramos et al. "OQuaRE: A SQuaRE-based approach for evaluating the quality of ontologies". In: *Journal of Research and Practice in Information Technology* 43.2 (2011), p. 159.

[30] S Falconer. "OntoGraf". In: *Protégé Wiki* (2010).

[31] Miriam Fernández et al. "What makes a good ontology? A case-study in fine-grained knowledge reuse". In: *Asian Semantic Web Conference*. Springer Berlin Heidelberg. 2009, pp. 61–75.

[32] Donna Fritzsche et al. "Ontology Summit 2016 Communique: Ontologies within semantic interoperability ecosystems". In: *Applied Ontology* Preprint (2017), pp. 1–21.

[33] Aldo Gangemi et al. "Modelling ontology evaluation and validation". In: *European Semantic Web Conference*. Springer Berlin Heidelberg. 2006, pp. 140–154.

[34] Aldo Gangemi et al. "Qood grid: A metaontology-based framework for ontology evaluation and selection". In: *Proceedings of the EONâĂŹ2006 Workshop*. 2006.

[35] Aldo Gangemi et al. "Sweetening ontologies with DOLCE". In: *International Conference on Knowledge Engineering and Knowledge Management*. Springer Berlin Heidelberg. 2002, pp. 166–181.

[36] Juan García, Francisco JoseâĂŹGarcía-Peñalvo, and Roberto Therón. "A survey on ontology metrics". In: *World Summit on Knowledge Society*. Springer Berlin Heidelberg. 2010, pp. 22–27.

[37] Jennifer Golbeck et al. "The National Cancer Institute's thesaurus and ontology". In: *Web Semantics: Science, Services and Agents on the World Wide Web* 1.1 (2011).

[38] Asunción Gómez-Pérez. "Evaluation of ontologies". In: *International Journal of intelligent systems* 16.3 (2001), pp. 391–409.

[39] Asunción Gómez-Pérez. "Towards a framework to verify knowledge sharing technology". In: *Expert Systems with Applications* 11.4 (1996), pp. 519–529.

[40] Asuncion Gomez-Perez, Mariano Fernández-López, and Oscar Corcho. *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer Science & Business Media, 2006.

[41] Thomas R Gruber. "Toward principles for the design of ontologies used for knowledge sharing?" In: *International journal of human-computer studies* 43.5-6 (1995), pp. 907–928.

[42] Thomas R Gruber et al. "A translation approach to portable ontology specifications". In: *Knowledge acquisition* 5.2 (1993), pp. 199–220.

[43] Michael Grüninger. *COLORE: Common logic ontology repository*. 2009.

[44] Michael Grüninger. "Ontology repositories make a world of difference". In: *Proceedings of the 7th international conference on Theory, Practice, and Applications of Rules on the Web*. Springer-Verlag. 2013, pp. 12–12.

[45] Michael Grüninger et al. "Modular first-order ontologies via repositories". In: *Applied Ontology* 7.2 (2012), pp. 169–209.

[46] Nicola Guarino. "Semantic matching: Formal ontological distinctions for information organization, extraction, and integration". In: *Information Extraction A Multidisciplinary Approach to an Emerging Information Technology*. Springer, 1997, pp. 139–170.

[47] Nicola Guarino. "Understanding, building and using ontologies". In: *International Journal of Human-Computer Studies* 46.2-3 (1997), pp. 293–310.

[48] Nicola Guarino and Christopher Welty. "Evaluating ontological decisions with OntoClean". In: *Communications of the ACM* 45.2 (2002), pp. 61–65.

[49] Nicola Guarino and Christopher A Welty. "An overview of OntoClean". In: *Handbook on ontologies*. Springer Berlin Heidelberg, 2009, pp. 201–220.

[50] Nicola Guarino et al. "Formal ontology and information systems". In: *Proceedings of FOIS*. Vol. 98. 1998. 1998, pp. 81–97.

[51] Jon Atle Gulla, Darijus Strasunskas, and Stein L Tomassen. "Semantic Interoperability in Multi-Disciplinary Domain. Applications in Petroleum Industry". In: *Proceedings of the 2nd International Workshop on Contexts and Ontologies: Theory, Practice and Applications, Riva del Garda, Italy*. 2006.

[52] Jens Hartmann et al. "D1. 2.3 Methods for ontology evaluation". In: *EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB Deliverable D* 1 (2005).

[53] Jim Hendler and Tim Berners-Lee. "From the Semantic Web to social machines: A research challenge for AI on the World Wide Web". In: *Artificial Intelligence* 174.2 (2010), pp. 156–161.

[54] Martin Hepp. “Possible ontologies: How reality constrains the development of relevant ontologies”. In: *IEEE Internet Computing* 11.1 (2007).

[55] Martin Hepp, Daniel Bachlechner, and Katharina Siorpaes. “OntoWiki: community-driven ontology engineering and ontology usage based on Wikis”. In: *Proceedings of the 2006 international symposium on Wikis*. ACM. 2006, pp. 143–144.

[56] Hlomani Hlomani and Deborah Stacey. “Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey”. In: *Semantic Web Journal* (2014), pp. 1–5.

[57] Robert Hoehndorf et al. “Interoperability between biomedical ontologies through relation expansion, upper-level ontologies and automatic reasoning”. In: *PloS one* 6.7 (2011), e22006.

[58] Matthew Horridge and Sean Bechhofer. “The OWL API: a Java API for working with OWL 2 ontologies”. In: *Proceedings of the 6th International Conference on OWL: Experiences and Directions-Volume 529*. CEUR-WS. org. 2009, pp. 49–58.

[59] Matthew Horridge et al. “A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools Edition 1.0”. In: *University of Manchester* (2004).

[60] Matthew Horridge et al. “The Manchester OWL Syntax.” In: *OWLed*. Vol. 216. 2006.

[61] Clement Jonquet, Mark A Musen, and Nigam H Shah. “Building a biomedical ontology recommender web service”. In: *Journal of biomedical semantics* 1.1 (2010), S1.

[62] Zubeida Casmod Khan. “Evaluation Metrics in Ontology Modules.” In: *Description Logics*. 2016.

[63] Zubeida Casmod Khan and C Maria Keet. “Dependencies between modularity metrics towards improved modules”. In: *Knowledge Engineering and Knowledge*

Management: 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings 20. Springer. 2016, pp. 400–415.

[64] Oliver Kutz and Joana Hois. “Modularity in ontologies”. In: *Applied Ontology* 7.2 (2012), pp. 109–112.

[65] Hannu Kuusela and Paul Pallab. “A comparison of concurrent and retrospective verbal protocol analysis”. In: *The American journal of psychology* 113.3 (2000), p. 387.

[66] Rensis Likert. “A technique for the measurement of attitudes.” In: *Archives of psychology* (1932).

[67] Adolfo Lozano-Tello and Asunción Gómez-Pérez. “Ontometric: A method to choose the appropriate ontology”. In: *Journal of database management* 2.15 (2004), pp. 1–18.

[68] Yinglong Ma et al. “A graph derivation based approach for measuring and comparing structural semantics of ontologies”. In: *IEEE Transactions on Knowledge and Data Engineering* 26.5 (2014), pp. 1039–1052.

[69] Alexander Maedche and Steffen Staab. “Ontology learning”. In: *Handbook on ontologies*. Springer, 2004, pp. 173–190.

[70] Cynthia Matuszek et al. “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*. 2006, pp. 44–49.

[71] Melinda McDaniel, Veda C Storey, and Vijayan Sugumaran. “The Role of Community Acceptance in Assessing Ontology Quality”. In: *International Conference on Applications of Natural Language to Information Systems*. Springer. 2016, pp. 24–36.

[72] George A Miller. “WordNet: a lexical database for English”. In: *Communications of the ACM* 38.11 (1995), pp. 39–41.

[73] Till Mossakowski, Oliver Kutz, and Mihai Codescu. “Ontohub: A semantic repository for heterogeneous ontologies”. In: *IN: PROC. OF DACSâĂŹ14 (2014*. 2014.

[74] Mark A Musen et al. "The national center for biomedical ontology". In: *Journal of the American Medical Informatics Association* 19.2 (2012), pp. 190–195.

[75] Fabian Neuhaus et al. "Towards ontology evaluation across the life cycle". In: *Applied Ontology* 8.3 (2013), pp. 179–194.

[76] Tuong Huan Nguyen et al. "KBRE: a framework for knowledge-based requirements engineering". In: *Software Quality Journal* 22.1 (2014), pp. 87–119.

[77] Ian Niles and Adam Pease. "Towards a standard upper ontology". In: *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001*. ACM. 2001, pp. 2–9.

[78] Immanuel Normann and Oliver Kutz. "Ontology Reuse and Exploration via Interactive Graph Manipulation". In: *Proc. of the ISWC Workshop on Ontology Repositories for the Web (SERES-2010)*. Citeseer. 2010.

[79] Natalya F Noy, Nicholas Griffith, and Mark A Musen. "Collecting community-based mappings in an ontology repository". In: *International Semantic Web Conference*. Springer Berlin Heidelberg. 2008, pp. 371–386.

[80] Natalya F Noy, Deborah L McGuinness, et al. *Ontology development 101: A guide to creating your first ontology*. 2001.

[81] Natalya F Noy et al. "Bioportal: A web repository for biomedical ontologies and data resources". In: *Proceedings of the 2007 International Conference on Posters and Demonstrations-Volume 401*. CEUR-WS. org. 2008, pp. 112–113.

[82] Natalya F Noy et al. "BioPortal: ontologies and integrated data resources at the click of a mouse". In: *Nucleic acids research* 37.suppl 2 (2009), W170–W173.

[83] Natalya Fridman Noy, Ray W Fergerson, and Mark A Musen. "The knowledge model of Protege-2000: Combining interoperability and flexibility". In: *International Conference on Knowledge Engineering and Knowledge Management*. Springer Berlin Heidelberg. 2000, pp. 17–32.

[84] Natalya Fridman Noy and Mark A Musen. “SMART: Automated support for ontology merging and alignment”. In: *Proc. of the 12th Workshop on Knowledge Acquisition, Modelling, and Management (KAW'99), Banf, Canada*. Citeseer. 1999.

[85] Leo Obrst et al. “Semantic web and big data meets applied ontology”. In: *Applied Ontology* 9.2 (2014), pp. 155–170.

[86] Leo Obrst et al. “The evaluation of ontologies”. In: *Semantic web*. Springer US, 2007, pp. 139–158.

[87] Anthony M Orme, H Tao, and Letha H Etzkorn. “Coupling metrics for ontology-based system”. In: *IEEE software* 23.2 (2006), pp. 102–108.

[88] Anthony M Orme, Haining Yao, and Letha H Etzkorn. “Indicating ontology data quality, stability, and completeness throughout ontology evolution”. In: *Journal of Software Maintenance and Evolution: Research and Practice* 19.1 (2007), pp. 49–75.

[89] Heiko Paulheim, Sven Hertling, and Dominique Ritze. “Towards evaluating interactive ontology matching tools”. In: *Extended Semantic Web Conference*. Springer Berlin Heidelberg. 2013, pp. 31–45.

[90] Perrine Pittet and Jérôme Barthélémy. “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. 2015.

[91] Robert Porzel and Rainer Malaka. “A task-based approach for ontology evaluation”. In: *ECAI Workshop on Ontology Learning and Population, Valencia, Spain*. 2004.

[92] María Poveda-Villalón, Asunción Gómez-Pérez, and Mari Carmen Suárez-Figueroa. “Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation”. In: *International Journal on Semantic Web and Information Systems (IJSWIS)* 10.2 (2014), pp. 7–34.

[93] Andreas Radinger et al. "BauDataWeb: the Austrian building and construction materials market as linked data". In: *Proceedings of the 9th International Conference on Semantic Systems*. ACM. 2013, pp. 25–32.

[94] Arkalgud Ramaprasad and Thant Syn. "Ontological meta-analysis and synthesis". In: (2013).

[95] Alan Rector et al. "OWL pizzas: Practical experience of teaching OWL-DL: Common errors & common patterns". In: *International Conference on Knowledge Engineering and Knowledge Management*. Springer. 2004, pp. 63–81.

[96] Alan L Rector. "Normalisation of ontology implementations: Towards modularity, re-use, and maintainability". In: *Proceedings Workshop on Ontologies for Multiagent Systems (OMAS) in conjunction with European Knowledge Acquisition Workshop*. 2002, pp. 1–16.

[97] Stuart Robinson, Greg Aumann, and Steven Bird. "Managing fieldwork data with toolbox and the natural language toolkit". In: (2007).

[98] Roni Romano, Yair Horesh, and Jeremie Dreyfuss. *Automated Ontology Development*. US Patent App. 14/173,435. 2014.

[99] Rosetta Romano and Craig McDonald. "ASSESSING THE QUALITY OF ONTOLOGY". In: (2011).

[100] Cornelius Rosse and José LV Mejino. "A reference ontology for biomedical informatics: the Foundational Model of Anatomy". In: *Journal of biomedical informatics* 36.6 (2003), pp. 478–500.

[101] Daniel L Rubin. "Creating and curating a terminology for radiology: ontology modeling and analysis". In: *Journal of digital imaging* 21.4 (2008), pp. 355–362.

[102] Andreas Scheuermann and Joerg Leukel. "Supply chain management ontology from an ontology engineering perspective". In: *Computers in Industry* 65.6 (2014), pp. 913–923.

[103] Daniel Schober et al. “OntoCheck: verifying ontology naming conventions and metadata completeness in Protégé 4”. In: *Journal of biomedical semantics* 3.2 (2012), S4.

[104] Nigel Shadbolt, Tim Berners-Lee, and Wendy Hall. “The semantic web revisited”. In: *IEEE intelligent systems* 21.3 (2006), pp. 96–101.

[105] John F Sowa. “Ontology, metadata, and semiotics”. In: *International Conference on Conceptual Structures*. Springer Berlin Heidelberg. 2000, pp. 55–81.

[106] Ronald Stamper et al. “Understanding the roles of signs and norms in organizations-a semiotic approach to information systems design”. In: *Behaviour & Information Technology* 19.1 (2000), pp. 15–27.

[107] Darijus Strasunskas and Stein Tomassen. “On variety of semantic search systems and their evaluation methods”. In: *International Conference on Information Management and Evaluation*. Academic Conferences International Limited. 2010, p. 380.

[108] Besiki Stvilia et al. “A framework for information quality assessment”. In: *Journal of the Association for Information Science and Technology* 58.12 (2007), pp. 1720–1733.

[109] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Mariano Fernández-López. “The NeOn methodology for ontology engineering”. In: *Ontology engineering in a networked world*. Springer Berlin Heidelberg, 2012, pp. 9–34.

[110] Vijayan Sugumaran and Veda C Storey. “Ontologies for conceptual modeling: their creation, use, and management”. In: *Data & knowledge engineering* 42.3 (2002), pp. 251–271.

[111] Samir Tartir, I Budak Arpinar, and Amit P Sheth. “Ontological evaluation and validation”. In: *Theory and applications of ontology: Computer applications*. Springer Netherlands, 2010, pp. 115–130.

[112] Samir Tartir et al. “OntoQA: Metric-based ontology quality analysis”. In: (2005).

[113] Andreas Tolk and James A Muguira. "The levels of conceptual interoperability model". In: *Proceedings of the 2003 fall simulation interoperability workshop*. Vol. 7. Citeseer. 2003, pp. 1–11.

[114] Michael Uschold. "Making the case for ontology". In: *Applied Ontology* 6.4 (2011), pp. 377–385.

[115] Pierre-Yves Vandenbussche et al. "Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web". In: *Semantic Web* 8.3 (2017), pp. 437–452.

[116] Johanna Völker et al. "AEON–An approach to the automatic evaluation of ontologies". In: *Applied Ontology* 3.1-2 (2008), pp. 41–62.

[117] Denny Vrandečić. "Ontology evaluation". In: *Handbook on Ontologies*. Springer Berlin Heidelberg, 2009, pp. 293–313.

[118] Ron Weber. "Conceptual modelling and ontology: Possibilities and pitfalls". In: *Journal of Database management* 14.3 (2003), p. 1.

[119] Barry Wellman. "An electronic group is virtually a social network". In: *Culture of the Internet* 4 (1997), pp. 179–205.

[120] Chris Welty, Deborah L McGuinness, and Michael K Smith. "Owl web ontology language guide". In: *W3C recommendation, W3C (February 2004) http://www. w3. org/TR/2004/REC-owl-guide-20040210* (2004).

[121] Patricia L Whetzel et al. "BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications". In: *Nucleic acids research* 39.suppl 2 (2011), W541–W545.

[122] Adam Wyner and Rinke Hoekstra. "A legal case OWL ontology with an instantiation of Popov v. Hayashi". In: *Artificial Intelligence and Law* 20.1 (2012), pp. 83–107.

[123] Adam Z Wyner. “Towards annotating and extracting textual legal case elements”. In: *Informatica e Diritto: special issue on legal ontologies and artificial intelligent techniques* 19.1-2 (2010), pp. 9–18.

[124] Guanghao Yan, Wee Keong Ng, and Ee-Peng Lim. “Product schema integration for electronic commerce-a synonym comparison approach”. In: *IEEE Transactions on Knowledge and Data Engineering* 14.3 (2002), pp. 583–598.

[125] Haining Yao, Anthony Mark Orme, and Letha Etzkorn. “Cohesion metrics for ontology design and application”. In: *Journal of Computer science* 1.1 (2005), pp. 107–113.

[126] Liyang Yu. *A developer’s guide to the semantic Web*. Springer Science & Business Media, 2011.

[127] Ying Zhao and George Karypis. *Comparison of agglomerative and partitional document clustering algorithms*. Tech. rep. Minnesota University Department of Computer Science, 2002.