A Framework for Ontology Evaluation

Jinie Pak and Lina Zhou

Department of Information Systems
University of Maryland, Baltimore County
Baltimore, MD 21250
{jpak3,zhoul}@umbc.edu

Abstract. The rapid growth in the number of ontologies has not met with the wide adoption of ontology in practice. Ontology evaluation can promote ontology use by facilitating the selection of a good ontology. Despite that a host of ontology evaluation methodologies are available, many of them are fragmentary and strongly tied to ontology development methodologies. Based on a review of extant ontology evaluation methods, we propose a framework for ontology evaluation. The framework provides a holistic view of ontology evaluation, suggesting both fundamental ontology dimensions and concrete criteria.

Keywords: Ontology, Ontology evaluation, Ontology quality.

1 Introduction

Ontology has witnessed significant growth and widespread application in many fields such as knowledge management, information retrieval, and semantic web. For example, ontology technologies are popular in inference engines, annotation tools, ontology-based crawlers, and mining tools [25]. Ontology is a formal, explicit specification of a shared conceptualization [12], which can capture semantic information at both concept and instance levels. Ontology can represent a set of objects and their associated relationships in a declarative language. The effectiveness of the ontological knowledge relies on the quality of the ontology.

Ontology evaluation is a technical judgment of the content of the ontology with respect to a frame of reference during every phase and between phases of its lifecycle [9]. Ontology evaluation is critical to the adoption and improvement of ontologies both in academics and industry. Nonetheless, it remains a challenging task to evaluate ontologies due in part to the fact that ontology is semantic oriented. By nature semantics involves ambiguity. Unlike software evaluation, the evaluation of ontology does not have access to concrete specifications for inputs and outputs. Therefore, ontology evaluation deserves separate attention.

Although there is extensive discussion of ontology evaluation in the literature [1], [7], [14], [25], systematic guidelines for performing ontology evaluation are still rare. Therefore, this research aims to improve our theoretical understanding of ontology evaluation and to provide guidance for ontology evaluation by proposing a framework. As a secondary objective, this research identifies future research issues in ontology evaluation, which call for new effort to advance the field.

R. Sharman, T.S. Raghu, and H.R. Rao (Eds.): WEB 2009, LNBIP 52, pp. 10–18, 2010.

2 Existing Methods for Ontology Evaluation

In the literature, various ontology evaluation approaches have been proposed with respect to domains, scopes, and purposes of ontologies Although Hartmann et al. [14] and Yu et al. [30] attempted to provide systematic guidelines for ontology evaluation, most ontology evaluation approaches are rather fragmentary. Some approaches may address certain evaluation dimensions to some extent and often involve more than one quality evaluation criterion. Other approaches evaluate ontologies from the application perspective by comparing ontologies against pre-defined gold standards [23]. There is a need to take a holistic look at ontology evaluation to gain a better understanding of the problem. This section presents a summarization of current ontology evaluation approaches and methods in two major categories: quality attributes based and task-oriented methods.

2.1 Quality Attribute-Based Approaches

Quality attributes-based approaches rely on a set of pre-define quality criteria in evaluation. Metrics are usually defined to measure individual quality attribute of an ontology (e.g. a score and a weight). This type of approach is particularly useful for checking for errors in ontologies such as redundant terms, inconsistencies between definitions and missing definitions. However, some quality attributes such as clarity and expandability can be difficult to evaluate as they may not be easily quantifiable. Thus, manual inspection by human experts is required.

OntoClean [28] is based on philosophical notions of formal evaluation of taxonomical structures. There are four core ontological notions: rigidity, unity, identity, and dependence. OntoClean was the first attempt to formalize notions of ontological analysis for information systems. The idea is to detect both formal and semantic inconsistencies in the properties defined in an ontology.

Content and technology evaluation [11] focuses on the definition of a set of criteria useful in the evaluation process and empirically applies some of these criteria in the evaluation of bibliographic-data ontology. This approach proposes a global technical evaluation that ensures well-defined properties in: 1) the definition of the ontology, 2) the software environment used to build, reuse and share definition, and 3) documentation. Each evaluation step involves three notions: consistency, completeness, and conciseness. In addition, this approach draws a distinction between two main evaluation dimensions: content evaluation and technology evaluation. Content evaluation is related to the Knowledge Representation (KR) paradigm that underlies the language in which ontology is implemented. Ontology technology evaluation is to ensure smooth and correct integration with industrial software environment. This evaluation is directed at the expressiveness of the KR model underlying the ontology editor, the tool's interoperability, quality of import and export functions, scalability, navigability, usability, and content evaluation functions.

Yao et al. [29] proposes a number of cohesion metrics that are specific to ontologies. Grounded on a number of mathematical theories such as Graph theory and Metric theory, cohesion metrics are used to describe and measure structures. There are three main cohesion metrics or functions for evaluating ontologies: 1) number of root classes, 2) average depth of inheritance, and 3) tree of leaf nodes. However, this metric is not suitable for direct application to ontology evaluation. As a result, cohesion metrics is used to measure specific items such as modularity (e.g. relatedness of elements in ontologies). The metrics can also be used to measure the degree of the relatedness of OWL classes and relations by properties.

Gold standard based approach [23] aims to improving the evaluation procedures by proposing criteria for good evaluation measures. The criteria allow for a multi-dimensional evaluation and ensure that errors are weighted differently based on their position in a concept hierarchy. This approach can evaluate the lexical term layer of an ontology as well as the concept hierarchy and non-taxonomic relations contained in an ontology. On the lexical term layer, precision and recall measures are used to compare the terms from the reference and the learned ontology based on string match. For the comparison of concept hierarchies, measures are divided into two categories: local and global. The local measure compares the similarity of the positions of two concepts in the learned and the reference hierarchies. The global measure is computed by averaging the results of the local measure for concept compares from the reference and the learned ontology.

2.2 Task-Oriented Approaches

Task-oriented approaches focus on the practical use of ontology in applications. Such evaluation focuses on user types, usefulness, usability, and use cases. Although this type of approaches can directly benefit practical applications by putting ontology to use, it does not provide generic viewpoints about the quality of an ontology outside the application context.

The task-based approach [19] is to evaluate ontologies at three basic levels: vocabulary, taxonomy and (non-taxonomic) semantic relations. It presents an evaluation scheme that allows ontology engineers to employ a number of different ontologies and to measure their performance on specific tasks. This approach provides performance measures as follows; quantify the respective gains and losses of insertion, deletion, and substitutions errors; populate and improve the ontology as derived from the results of individual specific error types; and re-evaluate the respective performance increases resulting from the improvements.

Structure and function based evaluation [19] views ontology evaluation from customers' perspectives. It emphasizes practical ways (functions) for ontology consumers to discover and evaluate ontologies. This approach suggests ontology summarization, e-pinions for ontologies, views and customization.

OntoMetric [17] is a method that helps choose the appropriate ontologies for a new project or a new system among existing candidate ontologies. OntoMetric provides a set of processes that the user should follow to obtain the measures of suitability of existing ontologies, regarding the requirements of a particular system. The main drawback of OntoMetric is related to its usability: specifying the characteristics of ontologies is a complicated and time-consuming process; assessing its characteristics is quite subjective. Moreover, the number of use cases is limited, which is an important obstacle to specifying parameters based on a large number of comparable cases.

Hartmann et al. [14] provides a classification of various ontology evaluation methods and tools for industrial practice. The method and tools are classified by the following tasks: 1) selecting existing ontologies, 2) measuring the similarities between textual source and its corresponding ontologies, 3) evaluating the impacts of ontology on IS applications, 4) checking the quality and consistency of ontologies, and 5) monitoring ontology in use.

3 A Framework for Ontology Evaluation

Based on the review of ontology evaluation approaches, we propose a framework for ontology evaluation. The framework provides a set of dimensions for classifying ontology evaluation methods, and for each dimension, it proposes measurement criteria for ontology evaluation. Additionally, we also propose guidelines for choosing ontology evaluation methods based on the objectives of an ontology.

Table 1. Taxonomy of Ontology Evaluation Approaches

Ontology Evaluation Dimensions	Articles
Scope	(Gómez-Pérez 2003), (Hartmann et al. 2004), (Lozano-Tello et al. 2004), (Porzel et al. 2004), (Daelemans et al. 2004),(Noy et al. 2004), (Staab et al. 2006)
Layer	(Welty et al. 2001), (Porzel et al. 2004), (Daelemans et al. 2004), (Yao et al. 2005), (Spyns et al. 2005), (Staab et al. 2006)
Life Cycle	(Welty et al. 2001), (Gómez-Pérez 2003), (Hartmann et al. 2004), (Lozano-Tello et al.2004), (Brewster et al. 2004), (Spyns et al. 2005)
Quality Principles	(Welty et al. 2001), (Gómez-Pérez 2003), (Yao et al. 2005)
Methods	(Hartmann et al. 2004), (Brewster et al. 2004), (Yao et al. 2005), (Spyns et al. 2005)

3.1 Ontology Scopes

Determining the domain and scopes of an ontology is important to both evaluation and building of the ontology. In ontological engineering, ontology scope is equivalent to specification and design aspects for the representation of knowledge underlying things, concepts, and phenomena [4]. Also ontology scope can be useful to confirming whether the specification and design aspects in ontologies are appropriately implemented. The scope of ontologies can be classified in three aspects.

- **Domain Scope** is to evaluate whether the scopes of ontologies are relevant to the tasks that ontologies are meant to accomplish in the intended domain.
- **Conceptual Scope** is to evaluate whether ontologies well-represent hierarchical and taxonomical concepts.
- **Technical Scope** is to evaluate whether the specifications and requirements for ontologies are integrated smoothly and correctly in terms of ontology integration and application in practice.

3.2 Ontology Layers

Ontology has a complex structure and it is not practical to directly evaluate ontology as a whole. Accordingly, it would be more efficient to focus on the evaluation of different layers of ontology separately. In addition, the layer-based approach allows for the use of distinctively different automatic techniques for different layers [15].

- **Lexicon/vocabulary layer.** This includes criteria relevant to knowledge representation and conceptualization of ontologies such as naming criteria in concepts, instances, and facts. Naming criteria assesses how well terms and definitions are formulated. Evaluation on this layer tends to involve comparisons with various sources of data concerning the problem domain (e.g. domain-specific corpora), as well as techniques such as string similarity measures (e.g. edit distance).
- **Structure/ architecture layer.** The structural attributes of ontologies are of primary interest in this layer. This layer is more relevant to hierarchical and taxonomic elements of ontology such as hierarchical relations among concepts, reasoning, and modularization. Usually this layer can be evaluated with both pre-defined principles or criteria and the suitability for further development [8].
- **Representation/ semantic layer.** This layer contains criteria relevant to the semantic elements of ontologies. Accordingly, the evaluation focuses on how adequately the semantic elements conceptually describe the structural elements that have been defined [2].
- **Context/application layer.** An ontology may be part of a large collection of ontologies, and may reference or be referenced by various definitions in these other ontologies [1]. It is very important to take this context into account when evaluating ontology. Another relevant issue is the application context where the ontology is to be used in. Typically evaluation looks at how the outcomes of the application are affected by the use of ontology.

3.3 Ontology Life Cycle

A technical evaluation must be performed throughout the life cycle of an ontology. The purpose is to detect the absence of some well-defined properties in the definitions in specification, knowledge acquisition, conceptualization and integration [9]. Ontology life cycle is the set of stages through which the ontology moves during its time. Each stage involves different levels of activities with regard to management, development, and support of ontologies. Accordingly, specific techniques used in each activity and the outputs from each activity need to be evaluated [10].

- **Specification.** The purpose of this phase is to develop informal, semi-formal or formal ontology requirement specification, using a set of intermediate representations or using competence questions. Accordingly, the intended scope, domain, and purpose of the ontology should be identified. Well-characterized specification is important to the design, evaluation, and reuse of an ontology. The specification needs to be checked for problems such as incompleteness, inconsistency, and conciseness and for feasibility. Thus, the evaluation is performed in the following areas: the degree of required formality, the set of terms to be represented, their characteristics and the required granularity of contents.
- **Knowledge acquisition.** As an independent activity in ontology development process, most acquisition is done simultaneously with the requirements specification phase. Knowledge can be elucidated, inspecting potential relevant glossary and refining the list of terms and their meanings. Sources span the complete range of domain experts who hold knowledge. Motivating scenarios and informal competency questions are collected. The competency questions, informal and formal text analysis are used to check whether the ontology is fit for purpose. Thus, this phase

verifies the domain facts and the abstract inference rules by using collected competency questions.

- **Conceptualization.** This phase is to identify the key concepts that exist in the domain, their properties and the relationships that hold between them, and is to structure the domain knowledge in a conceptual model that describes the problem and its solution in terms of the domain vocabulary identified in specification activity. As an outcome of the conceptualization, a complete Glossary of Terms is built, including concepts, instances, verbs and properties. Thus, this phase needs to checks consistency and completeness of identified domain terms such as concepts, instances, verbs relations or properties and their applicable informal representations**.**
- **Integration.** Sometimes an ontology uses or specializes existing ontologies to speed up the construction of the ontology. This phase is closely related to the reusability. When reusing the existing definitions, a task may be frequently hindered by the inadequate documentation of existing ontologies and their implicit assumptions. This phase assesses some uniformity in definitions across ontologies.

3.4 Ontology Quality Principles

Ontology is semantic oriented and may take a variety of forms. This leads to ambiguity in ontology evaluation. Therefore, ontology evaluation deserves more precise and clear principles. The following principles are most commonly identified in the ontology evaluation literature [8], [13].

- **Consistency.** This principle refers to the capability of getting no contradictory conclusions simultaneously from valid input data [9]. For example, ontology engineers check whether an ontology has semantic consistency in the definitions, the meanings of formal and informal definitions, and sentences that may be inferred by using other definitions and axiom that may or may not belong to the same ontology.
- **Completeness.** Sometimes some of important information about concepts in ontologies can be overlooked. Such incompleteness often leads to ambiguity and a lack of reasoning mechanisms. Precision and recall, set-based measures can be used for examining the incompleteness errors of ontologies. Precision is a measure of the ability of an ontology present only relevant item and recall is a measure of the ability of an ontology to present all relevant items. The completeness of a definition depends on the level of granularity agreed to in the whole ontology. To detect incompleteness of an ontology and its definition, the class hierarchy, the domains and scopes of the functions and relations, and classes need to be verified. The incompleteness errors in ontologies can be classified by incomplete concept classification, partition errors, and disjoint knowledge omission [11].
- **Conciseness.** This principle is related to whether all the information collected in an ontology is useful and precise. Conciseness doesn't imply absence of redundancies. Sometimes, some degree of controlled redundancy can be useful in identifying definitions. Explicit redundancies do not exist among definitions and redundancies cannot be derived using axioms attached to other definitions. Finally, the set of properties in the definitions of a class can be precisely and exactly defined.
- **Reusability.** Ontologies can import and export modules among ontologies. However, the effectiveness to interoperate modules among ontologies is often reduced

due to the use of different ontological scheme, that is, structural and semantic heterogeneity between ontologies. String matching metric can be used for measuring similarity between two strings.

3.5 Ontology Evaluation Methods

The choice of ontology evaluation methods can be contingent upon ontology development methodologies. Such a consideration is also related to other factors such as degree of formality and domain coverage or future extensions of ontology. Nonetheless, along this dimension, there are three main methods that cover all possible aspects of ontology evaluation.

- **Independent method.** They are used for various types of ontologies developed with different tools and different knowledge representation and formalization. These methods can be applied to ontology evaluation, no matter which methodology or approach is used for ontology development [16]. Precision, recall and string matching matrix are the examples of the independent methods.
- **Verification.** The verification of ontologies is to detect anomalies can occur due to the combination of ontology definitions and rules. Taxonomy evaluation and OntoClean are examples of this type of approach.
- **Validation.** This refers to a diagnostic task over ontology elements, process, and attributes. Structural measures, functional measures and usability -profiling measures are the examples of validation [7].

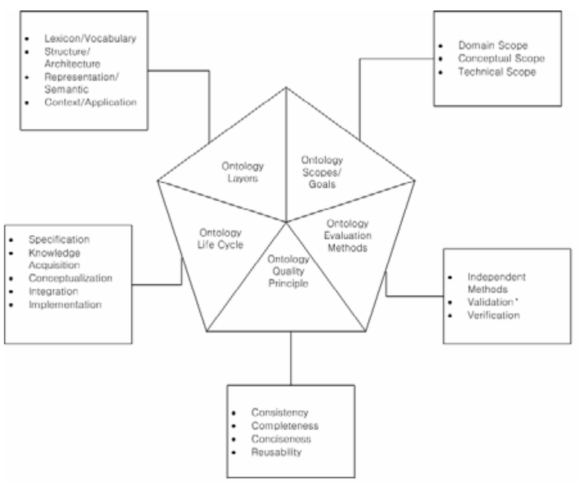

Fig. 1. Proposed Framework for Ontology Evaluation

4 Conclusion and Further Research

Ontology evaluation is a complex and time-consuming process. It is impossible for a single approach to ontology evaluation to work best for all application contexts.

Instead, the selection of an evaluation approach should consider a number of factors, including the phases where the ontology is evaluated, the application and domain in which the ontology is to be used, and the aspects of the ontology to be evaluated.

The proposed framework lays the groundwork for developing a comprehensive approach to ontology evaluation, addressing both basic ontology dimensions and concrete criteria. It allows conducting a multi-dimensional evaluation. It also supports the view that ontology evaluation should be considered throughout the entire life cycle of an ontology [8], [10]. This research can be continued in the following directions:

- Development of guidelines for choosing and prioritizing different dimensions of ontology evaluation in research and practice. High quality ontology measures and a set of agreed-on evaluation measures will support the general development and evolution of ontology engineering environment. Nonetheless, the importance of specific ontology evaluation criteria is likely to vary with ontology context.
- Identification of evaluation tools and methods that can be used to support each dimension of the framework. Tools can be used to automatically check if certain quality criteria are satisfied during the engineering phases, and if not, issue appropriate warnings.
- Implementation of the framework to demonstrate the validity and effectiveness of the framework.

The ontology evaluation framework proposed in this research is expected to help improve the use of ontologies by providing a set of dimensions for selecting good ontologies.

References

1. Brank, J., Grobelnik, M., Mladenic, D.: A survey of ontology evaluation techniques. In: The Conference on Data Mining and Data Warehouses, SiKDD (2005)
2. Brewster, C., Alani, H., Dasmahapatra, S., Wilks, Y.: Data Driven Ontology Evaluation. In: The Language Resources and Evaluation Conference (LREC), Lisbon, Portugal (2004)
3. Daelemans, W., Reinverger, M.L.: Shallow Text Understanding for Ontology Content Evaluation. IEEE Intelligent Systems 19(4), 74–81 (2004)
4. Devedzic, V.: Understanding Ontological Engineering. Communications of the ACM 45, 135–144 (2005)
5. Fernández, M., Gòmez-Pérez, A., Juristo, N.: METHONTOLOGY: From Ontological Art Towards Ontological Engineering. In: The AAAI Spring Symposium on Ontological Engineering, pp. 33–40. Stanford University, AAAI Press, Stanford, CA (1997)
6. Fernández-López, M.: Overview of Methodologies for Building Ontologies. In: The IJCAI 1999 Workshop on Ontologies and Problem-Solving Methods, Stockholm, Sweden (1999)
7. Gangemi, A., Catenacci, C., Ciaramita, M., Lehmann, J.: A theoretical framework for ontology evaluation and validation. In: The 2nd Italian Semantic Web Workshop (SWAP) (2005)
8. Gómez-Pérez, A.: Some ideas and examples to evaluate ontologies. In: The 11th Conference on Artificial Intelligence for Applications (CAIA 1995), Washington DC, p. 299 (1995)
9. Gómez-Pérez, A.: Towards a framework to verify knowledge sharing technology. Expert Systems with Applications 4, 519–529 (1996)
10. Gómez-Pérez, A.: Knowledge sharing and reuse. In: Liebowitz, J. (ed.) The Handbook of Applied Expert Systems, p. 10.1–10.35. CRC, Boca Raton (1998)

11. Gómez-Pérez, A.: Ontology Evaluation. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies, pp. 251–274. Springer, Heidelberg (2003)
12. Gruber, T.R.: A translation approach to portable ontology specifications. Knowledge Acquisition 5, 199–220 (1998)
13. Guarino, N., Welty, C.: Evaluating ontological decisions with OntoClean. Comm. of the ACM 45(2), 61–65 (2002)
14. Hartmann, J., Sure, Y., Giboin, A., Maynard, D., Suárez-Figueroa, M., Cuel, R.: Methods for Ontology Evaluation. KWeb Deliverable D1.2.3, University of Karlsruhe (2004)
15. Kehagias, D.D., Papadimitriou, I., Hois, J., Tzovaras, D., Bateman, J.: A methodological approach for ontology evaluation and refinement. In: 2nd International ASK-IT Final Conference, Nuremberg, Germany (2008)
16. Lovrencic, S., Cubrilo, M.: Ontology Evaluation-comprising verification and validation. In: The 19th Central European Conference on Information and Intelligent Systems, pp. 657–663 (2008)
17. Lozano-Tello, A., Gómez-Pérez, A.: ONTOMETRIC: A Method to Choose the Appropriate Ontology. Journal of Database Management. Special Issue on Ontological analysis, Evaluation, and Engineering of Business Systems Analysis Methods 15(2), 1–18 (2004)
18. Maedche, A., Staab, S.: Measuring Similarity between Ontologies. In: Gómez-Pérez, A., Benjamins, V.R. (eds.) EKAW 2002. LNCS (LNAI), vol. 2473, pp. 251–263. Springer, Heidelberg (2002)
19. Noy, N.F.: Semantic Integration: A Survey Of Ontology-Based Approaches. SIGMOD Record, Special Issue on Semantic Integration 33(4), 65–70 (2004)
20. Porzel, R., Malaka, R.: A Task-based Approach for Ontology Evaluation. In: ECAI 2004 Workshop on Ontology Learning and Population, Valencia, Spain (2004)
21. Spyns, P., Meersman, R., Jarrar, M.: Data modelling versus Ontology engineering. In: Sheth, A., Meersman, R. (eds.) SIGMOD Record Special Issue 31 (4), 12–17 (2002)
22. Spyns, P., Reinberger, M.-L.: Lexically evaluating ontology triples automatically generated from text. In: Gómez-Pérez, A., Euzenat, J. (eds.) ESWC 2005. LNCS, vol. 3532, pp. 563–577. Springer, Heidelberg (2005)
23. Dellschaft, K., Staab, S.: On How to Perform a Gold Standard Based Evaluation of Ontology Learning. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 228–241. Springer, Heidelberg (2006)
24. Stojanovic, N., Maedche, A., Staab, S., Studer, R., Sure, Y.: SEAL-A Framework for Developing Semantic PortALs. In: The 1st International Conference on Knowledge captures (ACM K-CAP), Vancouver, pp. 155–162 (2001)
25. Sure, Y., Gómez-Pérez, A., Daelemans, W., Reinberger, M., Guarino, N., Noy, N.F.: Why Evaluate Ontology Technologies? Because It Works! IEEE Intelligent Systems 19(4), 74–81 (2004)
26. Uschold, M., Gruninger, M.: Ontologies: Principles, methods, and applications. Knowledge Engineering Review 11(2), 93–155 (1996)
27. Ushold, M.: Creating, integrating and maintaining local and global ontologies. In: The First Workshop on Ontology Learning (OL-2000) in Conjunction with the 14th European Conference on Artificial Intelligence (ECAI). IOS Press, Berlin (2000)
28. Welty, C.A., Guarino, N.: Supporting ontological analysis of taxonomic relationships. Data & Knowledge Engineering 39, 51–74 (2004)
29. Yao, H., Orme, A.M., Etzkorn, L.: Cohesion metrics for ontology design and application. Journal of Computer Science 1(1), 107–113 (2005)
30. Yu, J., Thom, J., Tam, A.: Requirements-oriented methodology for evaluating ontologies. Information Systems, Sixteenth ACM 34(8), 686–711 (2009)