A Survey on Ontology Evaluation Methods

Joe Raad and Christophe Cruz
CheckSem Team, Le2i, University of Burgundy, Dijon, France

Keywords: Semantic Web, Ontology, Evaluation.

Abstract: Ontologies nowadays have become widely used for knowledge representation, and are considered as foundation for Semantic Web. However with their wide spread usage, a question of their evaluation increased even more. This paper addresses the issue of finding an efficient ontology evaluation method by presenting the existing ontology evaluation techniques, while discussing their advantages and drawbacks. The presented ontology evaluation techniques can be grouped into four categories: gold standard-based, corpus-based, task-based and criteria based approaches.

1 INTRODUCTION

For most people, the World Wide Web has become quite a long time ago an indispensable means of providing and searching for information. However, searching the web in its current form usually provides a large number of irrelevant answers, and leaves behind some other interesting ones. The main reason of these unwanted results is that existing Web resources are mostly only human understandable. Therefore, we can clearly see the necessity of extending this web and transform it into a web of data that can be processed and analysed also by machines.

This extension of the web through defined standards is called the Semantic Web, or could also be known by the term Web 3.0. This extended web will make sure that machines and human users will have a common communicating language, by annotating web pages with information on their contents. Such annotations will be given in some standardized, expressive language and make use of certain terms. Therefore one needs the use of ontologies to provide a description of such terms.

Ontologies are fundamental Semantic Web technologies, and are considered as its backbone. Ontologies define the formal semantics of the terms used for describing data, and the relations between these terms. They provide an "explicit specification of a conceptualization" (Gruber, 1993). The use of ontologies is rapidly growing nowadays, as they are now considered as the main knowledge base for several semantic services like information retrieval, recommendation, question answering, and decision making services. A knowledge base is a technology used to store complex information in order to be used by a computer system. A knowledge base for machines is equivalent to the level of knowledge for humans. A human's decision is not only affected by how every person thinks (which is the reasoning for machines), it is significantly affected by the level of knowledge he has (knowledge base for machines). For instance, the relationship of the two terms "Titanic" and "Avatar" does not exist at all for a given person. But, another person identifies them as related since these terms are both movie titles. Furthermore, a movie addict strongly relates these two terms, as they are not only movie titles, but these movies also share the same writer and director. We can see the influence and the importance of the knowledge base (level of knowledge for humans) in every resulting decision. Therefore we can state that having a "good" ontology can massively contribute to the success of several semantic services and various knowledge management applications. In this paper, we investigate what makes a "good" ontology by studying different ontology evaluation methods and discuss their advantages. These methods are mostly used to evaluate the quality of automatically constructed ontologies.

The remainder of this paper is organized as follows. The next section presents an introduction on ontologies and the criteria that need to be evaluated. Section three presents different types of ontology evaluation methods. Finally, before concluding, the last section presents the advantages of each type of

Raad, J. and Cruz, C..
A Survey on Ontology Evaluation Methods.
In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 2: KEOD*, pages 179-186
ISBN: 978-989-758-158-8

evaluation method and proposes an evaluation method based on the previous existing ones.

2 ONTOLOGY EVALUATION CRITERIA

The word ontology is frequently used to mean different things, (e.g. glossaries and data dictionaries, thesauri and taxonomies, schemas and data models, and formal ontologies and inference).

Despite having different functionalities, these different knowledge sources are very similar and connected in their main purpose to provide information on the meaning of elements. Therefore, due to the similarity of these knowledge sources, and in order to simplify the issue, we use the term ontology in the rest of this paper even though some of the papers are considering taxonomies in their approaches.

An example of one of the most used knowledge sources is the large English lexical database WordNet. In WordNet, there are four commonly used semantic relations for nouns, which are hyponym/hypernym (is-a), part meronym/part holonym (part-of), member meronym/member holonym (member-of) and substance meronym/substance holonym (substance-of). A fragment of (is-a) relation between concepts in WordNet is shown in Figure 1. We can also find many other popular general purpose ontologies like YAGO and SENSUS, and some domain specific ontologies like UMLS and MeSH (for biomedical and health related concepts), SNOMED (for clinical healthcare concepts), GO (for gene proteins and all concerns of organisms) and STDS (for earth-referenced spatial data).

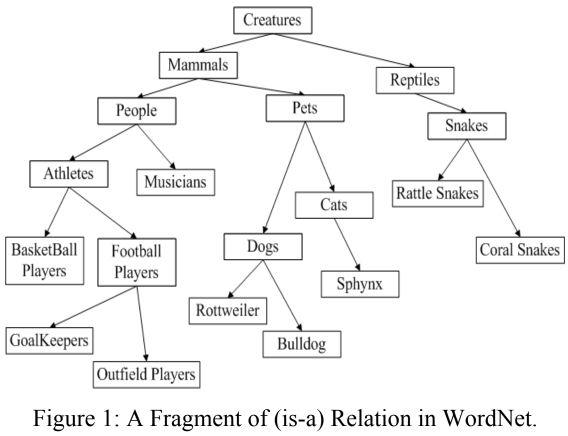

Figure 1: A Fragment of (is-a) Relation in WordNet.

However, the provided information by ontologies could be very subjective. This is mainly due to the fact that ontologies heavily depend on the level of knowledge (e.g. the case of an ontology constructed by human experts) or depend on its information sources (e.g. the case of an automatically constructed ontology).

In addition, while being useful for many applications, the size of ontologies can cause new problems that affect different steps of the ontology life cycle (d'Aquin et al., 2009). For instance, real world domain ontologies, and especially complex domain ontologies such as medicine, can contain thousands of concepts. Therefore these ontologies can be very difficult to create and normally require a team of experts to be maintained and reused. Another problem caused by large ontologies, is their processing. Very large ontologies usually cause serious scalability problems and increase the complexity of reasoning. Finally, the most important problem of large ontologies is their validation. Since ontologies are considered as reference models, one must insure their evaluation in the view of two important perspectives (Hlomani and Stacey, 2014): *quality* and *correctness*. These two perspectives address several criteria (Vrandečić, 2009; Obrst et al., 2007; Gruber, 1995; Gómez-Pérez, 2004; Gangemi et al., 2005):

- **Accuracy** is a criterion that states if the definitions, descriptions of classes, properties, and individuals in an ontology are correct.
- **Completeness** measures if the domain of interest is appropriately covered in this ontology.
- **Conciseness** is the criteria that states if the ontology includes irrelevant elements with regards to the domain to be covered.
- **Adaptability** measures how far the ontology anticipates its uses. An ontology should offer the conceptual foundation for a range of anticipated tasks.
- **Clarity** measures how effectively the ontology communicates the intended meaning of the defined terms. Definitions should be objective and independent of the context.
- **Computational Efficiency** measures the ability of the used tools to work with the ontology, in particular the speed that reasoners need to fulfil the required tasks.
- **Consistency** describes that the ontology does not include or allow for any contradictions.

In summary, we can state that ontology evaluation is the problem of assessing a given ontology from the point of view of these previously mentioned criteria, typically in order to determine which of several ontologies would better suit a particular purpose. In fact, an ontology contains both taxonomic and factual information that need to be evaluated. Taxonomic information includes information about concepts and their association usually organized into a hierarchical structure. Some approaches evaluate taxonomies by comparing them with a reference taxonomy or a reference corpus. This comparison is based on comparing the concepts of the two taxonomies according to one or several semantic measures. However, semantic measure is a generic term covering several concepts (Raad et al., 2015):

- **Semantic Relatedness**, which is the most general semantic link between two concepts. Two concepts do not have to share a common meaning to be considered semantically related or close, they can be linked by a functional relationship or frequent association relationship like meronym or antonym concepts (e.g. Pilot "is related to" Airplane).
- **Semantic Similarity**, which is a specific case of semantic relatedness. Two concepts are considered similar if they share common meanings and characteristics, like synonym, hyponym and hypernym concepts (e.g. Old "is similar to" Ancient).
- **Semantic Distance**, is the inverse of the semantic relatedness, as it indicates how much two concepts are unrelated to one another.

The following section presents the different existing types of ontology evaluation methods.

3 ONTOLOGY EVALUATION APPROACHES

Ontology evaluation is based on measures and methods to examine a set of criteria. The ontology evaluation approaches basically differ on how many of these criteria are targeted, and their main motivation behind evaluating the taxonomy. These existing approaches can be grouped into four categories: gold standard, corpus-based, task-based, and finally criteria based approaches.

This paper aims to distinguish between these categories of approaches and their characteristics while presenting some of the most popular works.

3.1 Gold Standard-based

Gold standard based approaches which are also known as ontology alignment or ontology mapping are the most straight-forward approach (Ulanov, 2010). This type of approach attempts to compare the learned ontology with a previously created reference ontology known as the gold standard. This gold standard represents an idealized outcome of the learning algorithm. However, having a suitable gold ontology can be challenging, since it should be one that was created under similar conditions with similar goals to the learned ontology. For this reason some approaches create specific taxonomies with the help of human experts to use it as the gold standard. While other approaches prefer to use reliable, popular taxonomies in a similar domain to consider it as their reference taxonomy, since it saves a considerable amount of work.

For instance, Maedche and Staab (2002) consider ontologies as two-layered systems, consisting of a lexical and a conceptual layer. Based on this core ontology model, this approach measures similarity between the learned ontology and a tourism domain ontology modelled by experts. It measures similarity based on the notion of lexicon, reference functions, and semantic cotopy which are described in details in (Maedche and Staab, 2002).

In addition, Ponzetto and Strube (2007) evaluate its derived taxonomy from Wikipedia by comparing it with two benchmark taxonomies. First, this approach maps the learned taxonomy with ResearchCyc using lexeme-to-concept denotational mapper. Then it computes semantic similarity with WordNet using different scenarios and measures: Rada et al., (1989), Wu and Palmer (1994), Leacock and Chodorow (1998), and Resnik's measure (1995).

Treeratpituk et al., (2013) evaluate the quality of its constructed taxonomy from a large text corpus by comparing it with six topic specific gold standard taxonomies. These six reference taxonomies are generated from Wikipedia using their proposed GraBTax algorithm.

Zavitsanos et al., (2011) also evaluate the learned ontology against a gold reference. This novel approach transforms the ontology concepts and their properties into a vector space representation, and calculates the similarity and dissimilarity of the two ontologies at the lexical and relational levels.

This type of approach is also used by Kashyap and Ramakrishnan (2005). They use the MEDLINE database as the document corpus, and the MeSH thesaurus as the gold standard to evaluate their constructed taxonomy. The evaluation process compares the generated taxonomy with the reference taxonomy using two classes of metrics: (1) Content Quality Metric: it measures the overlap in the labels between the two taxonomies in order to measure the precision and the recall. (2) Structural Quality Metric: it measures the structural validity of the labels. i.e. when two labels appear in a parent-child relationship in one taxonomy, they should appear in a consistent relationship (parent-child or ancestor-descendant) in the other taxonomy.

Gold standard-based approaches are efficient in evaluating the accuracy of an ontology. High accuracy comes from correct definitions and descriptions of classes, properties and individuals. Correctness in this case may mean compliance to defined gold standards. In addition, since a gold standard represents an ideal ontology of the specific domain, comparing the learned ontology with this gold reference can efficiently evaluate if the ontology covers well the domain and if it includes irrelevant elements with regards to the domain.

3.2 Corpus-based

Corpus-based approaches, also known as data-driven approaches, are used to evaluate how far an ontology sufficiently covers a given domain. The concept of this type of approach is to compare the learned ontology with the content of a text corpus that covers significantly a given domain. The advantage is to compare one or more ontologies with a corpus, rather than comparing one ontology with another existing one.

One basic approach is to perform an automated term extraction on the corpus and simply count the number of concepts that overlap between the ontology and the corpus. Another approach is to use a vector space representation of the concepts in both the corpus and the ontology under evaluation in order to measure the fit between them. In addition, Brewster et al., (2004) evaluate the learned ontology by firstly applying Latent Semantic Analysis and clustering methods to identify keywords in a corpus. Since every keyword can be represented in a different lexical way, this approach uses WordNet to expand queries. Finally, the ontology can be evaluated by mapping the set of concepts identified in the corpus to the learned ontology.

Similarly, Patel et al., (2003) evaluate the coverage of the ontology by extracting textual data from it, such as names of concepts and relations. The extracted textual data are used as input to a text classification model trained using standard machine learning algorithms.

Since this type of evaluation approach can be considered similar in many aspects to the gold-standard based approach, the two types of approaches practically cover the same evaluation criteria: accuracy, completeness and conciseness. In addition, the main challenge in this type of approach is also similar to the challenge in the gold-standard based approaches. However, it is easier. Finding a corpus that covers the same domain of the learned ontology is notably easier than finding a well-represented domain specific ontology. For example, Jones and Alani (2006) use the Google search engine to find a corpus based on a user query. After extending the user query using WordNet, the first 100 pages from Google results are considered as the corpus for evaluation.

3.3 Task-based

Task-based approaches try to measure how far an ontology helps improving the results of a certain task. This type of evaluation considers that a given ontology is intended for a particular task, and is only evaluated according to its performance in this task, regardless of all structural characteristics.

For example, if one designs an ontology for improving the performance of a web search engine, one may collect several example queries and compare whether the search results contain more relevant documents if a certain ontology is used (Welty et al., 2003).

Haase and Sure (2005) evaluate the quality of an ontology by determining how efficiently it allows users to obtain relevant individuals in their search. In order to measure the efficiency, the authors introduce a cost model to quantify the necessary user’s effort to arrive at the desired information. This cost is determined by the complexity of the hierarchy in terms of its breadth and depth.

Task-based approaches are considered the most efficient in evaluating the adaptability of an ontology, by applying the ontology to several tasks and evaluating its performance for these tasks. In addition, task-based approaches are mostly used in evaluating the compatibility between the used tool and the ontology, and computing the speed to fulfil the intended task. Finally, this type of approach can also detect inconsistent concepts by studying the performance of an ontology in a specified task.

3.4 Criteria-based

Criteria-based approaches measures how far an ontology or taxonomy adheres to certain desirable criteria. One can distinguish between measures related to the structure of an ontology and more sophisticated measures.

3.4.1 Structure-based

Structure-based approaches compute various structure properties in order to evaluate a given taxonomy. For this type of measure, it is usually no problem to have a fully automatic evaluation since these measures are quite straightforward and easy to understand. For instance, one may measure the average taxonomic depth and relational density of nodes. Others might evaluate taxonomies according to the number of nodes, etc. For instance, Fernandez et al., (2009) study the effect of several structural ontology measures on the ontology quality. From these experiments, the authors conclude that richly populated ontologies with a high breadth and depth variance are more likely to be correct. On the other hand, Gangemi et al., (2006) evaluate ontologies based on whether there are cycles in the directed graph.

3.4.2 Complex and Expert based

There are a lot of complex ontology evaluation measures that try to incorporate many aspects of ontology quality. For example, Alani and Brewster (2006) include several measures of ontology evaluation in the prototype system AKTiveRank, like class match measure, density and betweenness which are described in details in (Alani and Brewster, 2006).

In addition, Guarino and Welty (2004) evaluate ontologies using the OntoClean system, which is based on philosophical notions like the essence, identity and unity. These notions are used to characterize relevant aspects of the intended meaning of the properties, classes, and relations that make up an ontology.

Lozano-Tello and Gomez-Perez (2004), evaluate taxonomies based on the notion of multilevel tree of characteristics with scores, which includes design qualities, cost, tools, and language characteristics.

Criteria based approaches are the most efficient in evaluating the clarity of an ontology. The clarity could be evaluated using simple structure-based measures, or more complex measures like OntoClean. In addition, this type of approach is capable on measuring the ability of the used tools to work with the ontology by evaluating the ontology properties such as the size and the complexity. Finally, criteria-based measures and especially the more complex ones are efficient in detecting the presence of contradictions by evaluating the axioms in an ontology.

4 DISCUSSION

4.1 Overview

In section two, we presented the criteria that need to be available in a "good" ontology. Then in section three, we presented several ontology evaluation methods that tackle some of these criteria. The relationship between these criteria and methods is more or less complex: criteria provide justifications for the methods, whereas the result of a method will provide an indicator for how well one or more criteria are met. Most methods provide indicators for more than one criteria. Table I presents an overview of the discussed ontology evaluation methods.

Table 1: An overview of ontology evaluation methods.

	Gold	**Corpus**	**Task**	**Criteria**
Accuracy				
Completeness				
Conciseness				
Adaptability				
Clarity				
Computational Efficiency				
Consistency				

It is difficult to construct a comparative table that compares the ontology evaluation methods based on their addressed criteria. This is mainly due to the diversity of every evaluation approach, even the ones that are grouped under the same category. In Table I we present a comparison of the evaluation methods, based on the previously presented criteria.

A darker colour in the table represents a better coverage for the corresponding criterion.

Accuracy is a criterion that shows if the axioms of an ontology comply with the domain knowledge. A higher accuracy comes from correct definitions and descriptions of classes, properties and individuals. Evaluating if an ontology has a high accuracy can typically be achieved by comparing the ontology to a gold reference taxonomy or to a text corpus that covers the domain.

Completeness measures if the domain of interest is appropriately covered. An obvious method is to compare the ontology with a text corpus that covers significantly the domain, or with a gold reference ontology if available.

Conciseness is the criteria that states if the ontology includes irrelevant elements with regards to the domain to be covered or redundant representations of the semantics. Comparing the ontology to a text corpus or a reference ontology that only contain relevant elements is an efficient method to evaluate the conciseness of a given ontology. One basic approach is to check if every concept in the ontology (and its synonym) is available in the text corpus or the gold ontology.

Adaptability measures how far the ontology anticipates its use. In order to evaluate how efficient new tools and unexpected situations are able to use the ontology, it is recommended to use the ontology in these new situations and evaluate its performance depending on the task.

Clarity measures how effectively the ontology communicates the intended meaning of the defined terms. Clarity depends on several criteria: definitions should be objective and independent, ontologies should use definitions instead of description for classes, entities should be documented sufficiently and be fully labelled in all necessary languages, etc. Most of these criteria can ideally be evaluated using criteria based approaches like OntoClean (Guarino and Welty, 2004).

Computational Efficiency measures the ability of the used tools to work with the ontology, in particular the speed that reasoners need to fulfil the required tasks. Some types of axioms, in addition to the size of the ontology may cause problems for certain reasoners. Therefore evaluating the computational efficiency of an ontology could be done by checking its performance in different tasks. This will allow us to compute the compatibility between the tool and the ontology, and the speed to fulfil the task. Furthermore, structure based approaches that evaluate the ontology size, in addition to more sophisticated criteria based approaches that evaluate the axioms of the ontology can also prove to be a solution to evaluate the computational efficiency in a given ontology.

Consistency describes that the ontology does not include or allow for any contradictions. An example for an inconsistency is the description of the element Lion being “A lion is a large tawny-coloured cat that lives in prides”, but having a logical axiom ClassAssertion (ex: Type_of_chocolate ex: Lion). Consistency can ideally be evaluated using criteria based approaches that focus on axioms, or also can be detected and evaluated according to the performance of the ontology in a certain task.

As figured in Table 1, all type of approaches provide indicators for more than one criteria. However, still none of the mentioned approaches can evaluate an ontology according to all the mentioned criteria. In order to target as many criteria as possible, one can evaluate an ontology by combining two or more type of approaches. According to Table I, we clearly see the resemblance of the gold standard and corpus based approaches. We also see the resemblance of the criteria evaluated by the task based and criteria based approaches, despite having completely different evaluation principles. Therefore, evaluating an ontology using a gold standard based or a corpus based approach, in addition of evaluating the ontology based on a task based or criteria based approach can target at least six out of seven evaluation criteria. However, the challenging part is to find the most efficient and compatible measures in every type of approach in order to succeed in combining two (or more) approaches.

4.2 Proposition

Now after we studied different ontology evaluation methods, which approach is the most efficient one? Unfortunately, we cannot conclude from this survey which approach is the “best” to evaluate an ontology in general. We believe that the motivation behind evaluating an ontology can give one approach the upper hand on the others. In this context, and according to Dellschaft and Staab (2008), we should distinguish between two scenarios. The first scenario is choosing the best approach to evaluate the learned ontology, and the second scenario is choosing the best approach to evaluate the ontology learning algorithm itself. According to (Dellschaft and Staab, 2008) task-based, corpus-based and structure based approaches are identified to be more efficient in evaluating the learned ontology. While gold standard based and complex and expert based approaches are identified to be more efficient in evaluating the ontology learning algorithm.

We propose, based on Porzel and Malaka’s approach (2004), to evaluate the learned ontology using a task-based approach that also require the use of a gold standard. For instance, let’s consider that the learned ontology is intended to be used in a system that classifies a large number of documents. This system will classify documents based on several criteria like their themes and authors, and will use the learned ontology as its knowledge base. Therefore, the classification process is influenced by two main factors: the classification algorithm and the ontology being used as a knowledge base.

We propose to evaluate the ontology by comparing the classification results obtained using the automatically learned ontology with the classification results obtained using a gold standard ontology. We should mention that all the classification factors, and mainly the classification algorithm should kept unchanged between the two experiences. Figure 2 illustrates the evaluation process.

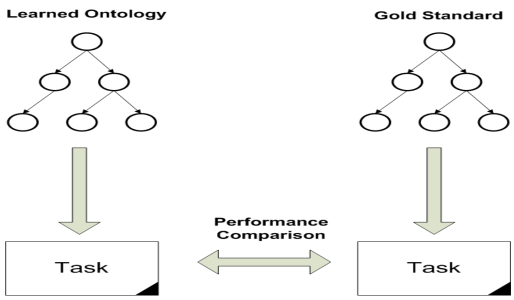

Figure 2: Ontology Evaluation Proposition.

We manage in this proposition to cover the two mentioned scenarios and to cover the maximum number of criteria by combining the task-based approach with the gold standard approach. This approach benefits from the simplicity of the task-based measures compared to the complexity of the similarity measures used in the gold-standard based approaches. It also benefits from the importance of having an ideal reference ontology for comparison. However it carries the main drawback of the gold-standard based approaches, which is finding or constructing a matching reference ontology to compare the performance.

5 FUTURE WORK

This survey can be considered as an introduction to a large topic. Finding an efficient approach to evaluate any ontology is still an unresolved issue, despite the large number of researches targeting this issue for many years.

After presenting several evaluation methods and discussing their drawbacks and advantages, our next objective is to directly compare its efficiency with the other evaluation methods. Our aim is to finally have a unified (semi-)automatic approach to evaluate an ontology with the minimum involvement of the human experts.

6 CONCLUSIONS

In the last years, the development of ontology-based applications has increased considerably. This growth increases the necessity of finding an efficient approach to evaluate these ontologies. Finding efficient evaluation schemes contributed heavily to the overwhelming success of disciplines like information retrieval, machine learning or speech recognition. Therefore having a sound and systematic approach to ontology evaluation is required to transform ontology engineering into a true scientific and engineering discipline.

In this paper, we presented the importance of ontologies, and the criteria expected to be available in these ontologies. Then we presented different approaches that aim to guarantee the maintenance of some of these criteria in automatic constructed ontologies. These approaches can be grouped into four categories: golden-standard, corpus-based, task-based, and finally criteria based approaches. Finally we proposed an approach to evaluate ontologies by combining the task-based and the gold-standard approaches in order to cover the maximum number of criteria.

ACKNOWLEDGEMENTS

The authors would like to thank the “Conseil Régional de Bourgogne” for their valuable support.

REFERENCES

Alani, H., & Brewster, C. (2006). *Metrics for ranking ontologies.*

d’Aquin, M., Schlicht, A., Stuckenschmidt, H., & Sabou, M. (2009). Criteria and evaluation for ontology modularization techniques. In *Modular ontologies* (pp. 67-89). Springer Berlin Heidelberg.

Brewster, C., Alani, H., Dasmahapatra, S., & Wilks, Y. (2004). *Data driven ontology evaluation.*

Dellschaft, K., & Staab, S. (2008, June). Strategies for the evaluation of ontology learning. In *Proceedings of the 2008 Conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge, Frontiers in Artificial Intelligence and Applications* (Vol. 167, pp. 253-272).

Elias Zavitsanos, George Paliouras, and George A. Vouros. (2011, November) *Gold standard evaluation of ontology learning methods through ontology transformation and alignment. IEEE Trans. on Knowl. and Data Eng., 23(11):1635–1648.*

Fernández, M., Overbeeke, C., Sabou, M., & Motta, E. (2009). What makes a good ontology? A case-study in fine-grained knowledge reuse. In *The semantic web* (pp. 61-75). Springer Berlin Heidelberg.

Gangemi, A., Catenacci, C., Ciaramita, M., & Lehmann, J. (2005). Ontology evaluation and validation: an integrated formal model for the quality diagnostic task.

On-line: http://www.loa-cnr.it/Files/OntoEval4OntoDev_Final. pdf.

Gangemi, A., Catenacci, C., Ciaramita, M., & Lehmann, J. (2006). *Modelling ontology evaluation and validation* (pp. 140-154). Springer Berlin Heidelberg.

Gómez-Pérez, A. (2004). Ontology evaluation. In *Handbook on ontologies* (pp. 251-273). Springer Berlin Heidelberg.

Gruber, T. R. (1993). A translation approach to portable ontology specifications.*Knowledge acquisition*, *5*(2), 199-220.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International journal of human-computer studies*, *43*(5), 907-928.

Guarino, N., & Welty, C. (2004). An Overview of OntoClean in S. Staab, R. Studer (eds.), Handbook on Ontologies.

Haase, P., & Sure, Y. (2005). D3. 2.1 Usage Tracking for Ontology Evolution.

Hlomani, H., & Stacey, D. (2014). Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey. *Semantic Web Journal, na (na)*, 1-5.

Jones, M., & Alani, H. (2006). Content-based ontology ranking.

Kashyap, V., Ramakrishnan, C., Thomas, C., & Sheth, A. (2005). TaxaMiner: an experimentation framework for automated taxonomy bootstrapping. *International Journal of Web and Grid Services*, *1*(2), 240-266.

Leacock, C., & Chodorow, M. (1998). Combining local context and WordNet similarity for word sense identification. *WordNet: An electronic lexical database*, *49*(2), 265-283.

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

Maedche, A., & Staab, S. (2002). Measuring similarity between ontologies. In *Knowledge engineering and knowledge management: Ontologies and the semantic web* (pp. 251-263). Springer Berlin Heidelberg.

Obrst, L., Ceusters, W., Mani, I., Ray, S., & Smith, B. (2007). The evaluation of ontologies. In *Semantic Web* (pp. 139-158). Springer US.

Patel, C., Supekar, K., Lee, Y., & Park, E. K. (2003, November). OntoKhoj: a semantic web portal for ontology searching, ranking and classification. In*Proceedings of the 5th ACM international workshop on Web information and data management* (pp. 58-61). ACM.

Ponzetto, S. P., & Strube, M. (2007, July). Deriving a large scale taxonomy from Wikipedia. In *AAAI* (Vol. 7, pp. 1440-1445).

Porzel, R., & Malaka, R. (2004, August). A task-based approach for ontology evaluation. In *ECAI Workshop on Ontology Learning and Population, Valencia, Spain*.

Raad, J., Bertaux, A., & Cruz, C. (2015, July). A survey on how to cross-reference web information sources. In *Science and Information Conference (SAI), 2015* (pp. 609-618). IEEE.

Rada, R., Mili, H., Bicknell, E., & Blettner, M. (1989). Development and application of a metric on semantic nets. *Systems, Man and Cybernetics, IEEE Transactions on*, *19*(1), 17-30.

Resnik, Philip. "Using information content to evaluate semantic similarity in a taxonomy." *arXiv preprint cmp-lg/9511007* (1995).

Treeratpituk, P., Khabsa, M., & Giles, C. L. (2013). Graph-based Approach to Automatic Taxonomy Generation (GraBTax). *arXiv preprint arXiv:1307.1718*.

Ulanov, A., Shevlyakov, G., Lyubomishchenko, N., Mehra, P., & Polutin, V. (2010, August). Monte Carlo Study of Taxonomy Evaluation. In *Database and Expert Systems Applications (DEXA), 2010 Workshop on* (pp. 164-168). IEEE.

Vrandečić, D. (2009). *Ontology evaluation* (pp. 293-313). Springer Berlin Heidelberg.

Welty, C. A., Mahindru, R., & Chu-Carroll, J. (2003, October). Evaluating ontological analysis. In *Semantic Integration Workshop (SI-2003)* (p. 92).

Wu, Z., & Palmer, M. (1994, June). Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics* (pp. 133-138). Association for Computational Linguistics.