OntoMetrics: Putting Metrics into Use for Ontology Evaluation

Birger Lantow
Institute of Computer Science, Rostock University, Albert-Einstein-Str. 22, 18051 Rostock, Germany

Keywords: Ontology, Ontology Evaluation, Ontology Metrics, Ontology Quality.

Abstract: Automatically calculated metrics are needed in order to evaluate ontology quality. Otherwise, major resources are required in order to manually assess certain Quality Criteria of an ontology. While there is rule based support for the detection modelling errors and the violation of ontology modelling guidelines, there is a lack of support for calculating ontology metrics. However, metrics can serve as indicators for possible quality problems that are not covered rule based ontology evaluation. Many metrics have been proposed that correlate for example with ontology characteristics like Readability, Adaptability, and Reusability. However, there is a lack of tool support. OntoMetrics provides free access to metric definition and calculation. Furthermore it fosters the development of knowledge regarding the application of Ontology Metrics. This paper provides theoretical background and usage scenarios for the OntoMetrics on-line platform.

1 INTRODUCTION

Ontology evaluation is a task that has been subject to research for years. A strong focus has been laid on the selection of appropriate ontologies for reuse in ontology engineering. Considering existing ontologies as sources for the construction of new ones is part of accepted ontology engineering methods. Frenandez et al. for example have an explicit integration-step in their METHONTOLOGY approach (Fernández-López et al., 1997) . With the increasing number of existing ontologies and thus an increasing number of candidates for reuse in a certain domain, automated evaluation of ontologies is required. Swoogle[1] as a search engine for ontologies calculates an ontology rank for the order of the presented search results (Ding et al., 2004). Additionally, swoogle calculates some basic ontology-metrics that become part of the available ontology meta-data. Besides the reuse aspect, ontology quality should be monitored throughout the ontology life-cycle. This includes the creation of ontologies but also their maintenance. Again, there is a need for automated evaluation due to the complexity of ontologies and knowledgebases. A majority of the approaches suggests metric calculation in order to assess ontology characteristics (examples in (Alani et al., 2006; Burton-Jones et al., 2005; Gangemi et al., 2005; Guarino and Welty, 2009; Lozano-Tello and Gómez-Pérez, 2004; Tartir et al., 2005)).

[1] http://swoogle.umbc.edu/

Existing Ontology Development Environments like Protégé provide only basic support for ontology evaluation. Plug-ins for ontology evaluation that have been developed are bound to a certain ontology editor. This comes with some disadvantages: (a) the user is forced to use the editor the plug-in is developed for, (b) plug-ins tend to be outdated if new editor versions evolve (see also (Poveda-Villalón et al., 2012)), and (c) discontinuation of editor development makes the plug-in unavailable for use. Furthermore, a number of approaches remained in the status of prototypes or just proposals (e.g. (Gavrilova et al., 2012)). As a consequence, approaches to metric-based automated ontology evaluation are rarely available for empirical evaluation and practical use. So far, web based solutions seem to be the best at hand because of their public availability and their independence from ontology development environments. The OntoMetrics[2] on-line platform has been developed as a consequence of this discussion by:

1. Providing a freely accessible web-based platform for ontology metric calculation.
2. Supporting a standardized ontology format (OWL 2).
3. Collecting the theory behind the metrics and their validation.
4. Providing machine-readable XML-output for further analysis.

[2] http://www.ontometrics.org

Lantow, B.
OntoMetrics: Putting Metrics into Use for Ontology Evaluation.
DOI: 10.5220/0006084601860191
In *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2016) - Volume 2: KEOD*, pages 186-191
ISBN: 978-989-758-203-5

In the following, section 2 provides the theoretical background for OntoMetrics. The discussion describes the context in which OntoMetrics is intended to be used, why the on-line platform is needed, and what limitations can be expected. Section 3 describes the current status of OntoMetrics and how it can be used. Finally, the future development roadmap is drawn in section 4.

2 ONTOLOGY METRIC CALCULATION

As mentioned in the introduction, there are many approaches to automated ontology metric calculation. Basically three aspects need to be covered in order to provide valid metrics:

1. A theory must be developed that explains the correlation of a proposed metric with the degree of fulfilment with a relevant *Ontology Quality Criteria*.
2. A (automated) measurement procedure must be formulated for the proposed metric.(Gangemi et al., 2005)
3. The theory must be (empirically) validated. Hence, the proposed correlation between *Ontology Metric* and *Ontology Quality Criteria* fulfilment must be proven.

Sometimes the first step is omitted. However, the last step is still an issue because of the small number of cases or special domains used for validation. For example, Gavrilova et al. (Gavrilova et al., 2012) assess 10 ontologies with 3 experts and compare the experts ranks to calculated metrics. Alani et al. (Alani et al., 2006) had 12 ontologies and 4 experts/users. Fernandez at al. (Fernández et al., 2009) have 5 ontologies in their study. All studies are restricted to a certain application domain each. The work of Lantow and Sandkuhl (Lantow and Sandkuhl, 2015) considers only a special type of ontologies – Ontology Design Patterns (ODP). Tartir et al. (Tartir et al., 2005) only demonstrate that their metrics seem to be plausible applied to a set of 3 ontologies.

In order to have a better validation of the proposed metrics, the metrics calculation needs to be made available, reproducible, and analysable. A platform like OntoMetrics addresses these issues. However, as mentioned in the introduction, ontology evaluation is worthwhile throughout the ontology life-cycle. Thus, different evaluation situations cause distinct quality requirements and distinct base data for evaluation. Furthermore, the way of metric calculation can differ significantly. Ontology evaluation frameworks give orientation here.

Pak and Zhou (Pak and Zhou, 2011) define five dimensions in their framework for ontology evaluation characterization: *(I) Ontology Scope, (II) Ontology Layers, (III) Ontology Life-cycle, (IV) Ontology Quality Criteria (Principles),* and *(V) Ontology Evaluation Methods*. Gavrilova et al. define in (Gavrilova et al., 2012) 6 quite similar dimensions, which in addition include the nature of the analysis result (quantitative/qualitative) and the level of automation. However, in the case of OntoMetrics we only consider fully automated methods that have quantitative results. Still, there is a broad variety of suggested evaluation methods and of those that may be developed in future. Their suitability for integration in an on-line platform like OntoMetrics depends on their characteristics within the framework. The dimensions can not be seen independently. For example, an *Ontology Evaluation Method* is suitable for the assessment of a certain *Ontology Quality Criteria*, and the relevance of the criteria varies with the *Ontology Scope*.

In the following, these dependencies are described taking possible dimension *(I) Ontology Scope* values as a starting point. Typical general *(IV) Ontology Quality Criteria* and *(V) Evaluation Methods* for the scopes are named. Three aspects of ontologies can be evaluated and thus set the scope (cf. (Pak and Zhou, 2011)):

How Well Does the Ontology Represent the Real World? – *Domain Scope*. Pak and Zhou (Pak and Zhou, 2011) name this aspect *Domain Scope* while Hlomani and Stacey (Hlomani and Stacey, 2014) and other authors speak of ontology correctness. Generally, data driven evaluation methods can be applied here like presented by Alani et al. (Brewster et al., 2004). These methods either use text corpora of the represented domain or golden standard ontologies and compare the evaluated ontology to them. *Ontology Quality Criteria* that are evaluated in the *Domain Scope* (correctness of the ontology) are *Accuracy, Completeness, Conciseness,* and *Consistency* (Hlomani and Stacey, 2014).

What Is the Quality of the Ontology in Analogy to Internal Software Quality Characteristics?– *Conceptual Scope*. The second aspect for evaluation is the internal quality of the ontology, it is called *Conceptual Scope* (Pak and Zhou, 2011) or ontology quality(Hlomani and Stacey, 2014). Generally, ontology structure based methods can be applied here. We divide between schema metrics suggested for example by Tartir et al. (Tartir et al., 2005) that consider the

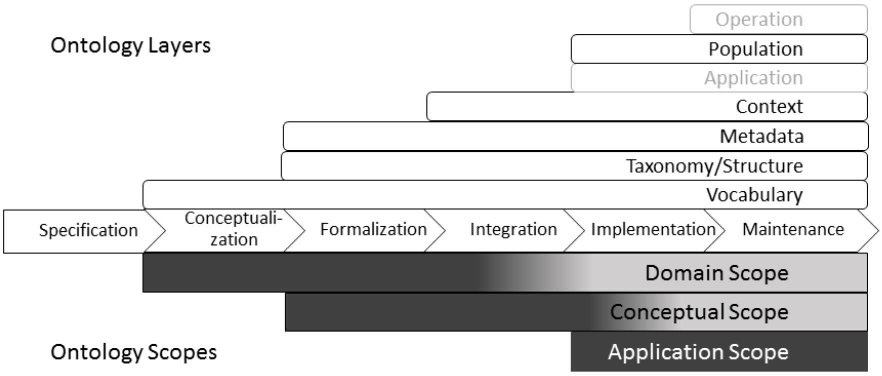

Figure 1: Ontology Layers and Scopes during Ontology Life-cycle.

special semantics of the ontology schema graph elements and graph-based metrics that calculate general graph characteristics like size and breadth for the taxonomical part of the ontology (for example Gangemi et al.(Gangemi et al., 2005)). Furthermore, Gangemi et al. suggest metrics based on annotations within the ontology. *Ontology Quality Criteria* that are evaluated in the *Conceptual Scope* are for example *Computational Efficiency, Adaptability, Clarity* (Hlomani and Stacey, 2014), *Reuseability* (Lantow and Sandkuhl, 2015), and *Readability* (Tartir et al., 2005).

How Well Does the Ontology in Use as a Component of an Ontology based Information System (External Software Quality)? – *Application Scope*. The third aspect of evaluation is the external quality of an ontology in conjunction with an information system – the *Application Scope*. Thus, the characteristics of the used information system have an influence on the on the measured quality. Task-based methods as described by Porzel and Malaka (Porzel and Malaka, 2004) can be used here. Another possibility is the assessment of usage statistics (Gavrilova et al., 2012). However, little effort has been spent on the development and validation of methods that evaluate ontologies within the *Application Scope*. A reason lies in the effort of evaluating different ontologies for the use within the same information system or vice versa to evaluate the same ontology in different information systems in order to exclude the information system's influence on the measurement. Nonetheless, monitoring changes in the metrics that are used to measure these criteria should provide information regarding ontology maintenance. *Ontology Quality Criteria* that can be evaluated in the *Application Scope* are *Efficiency, Effectivity, Accuracy* and general *Value* (measured by Popularity).

So far, the dimensions *(I) Ontology Scope, (IV) Quality Criteria,* and *(V) Evaluation Method* have been discussed in conjunction. In the following the dimensions *(II) Ontology Layer* and *(III) Ontology Life-Cycle* are added. Generally, there is an agreement on the phases of the ontology life-cycle: *(1) Specification, (2) Conceptualization, (3) Formalization, (4) Integration, (5) Implementation*, and *(6) Maintenance* (taken from the MethOntology ontology development method (Fernández-López et al., 1997)). There are only differences in the granularity (Gavrilova et al. name just 3 phases in (Gavrilova et al., 2012)) and in the order of the phases (Noy and McGuinnes put the integration/reuse at an earlier stage (Noy et al., 2001)).

Each step within the ontology-lifecycle adds some sub-artefact to the overall artefact 'ontology'. Such a sub-artefact is usually called an *Ontology Layer* (Pak and Zhou, 2011) or level (Hlomani and Stacey, 2014). Gavrilova et al. (Gavrilova et al., 2012) just call them objects to analyse. Figure 1 shows the common view on *Ontology Layer*s and their creation during the *Ontology Life-cycle* in conjunction with the *Ontology Scope* and their relevance during the *Ontology Life-cycle*. While the *Vocabulary* of an ontology is defined during *Conceptualization*, formal *Metadata* and ontology *Taxonomy/Structure* become available for analysis during *Formalization*. With the *Integration* of other ontologies, the *Context* of the ontology is set. *Implementation* then adds *Population* and the *Application* within an information system that uses the ontology. *Operation*al data of the information system becomes available when the system is in use and the ontology is in the *Maintenance* phase.

Except for *Operation* and *Application* layer (grayed in figure 1), data of all ontology layers is part of the OWL 2 specification and thus can be evaluated by an on-line platform like OntoMetrics. For the *Ontology Scope* it is depicted, when they can be addressed in the ontology life-cycle and at which phases they are most important (dark areas in figure 1).

Table 1 shows the accessibility of *Ontology Layer*s to the *Evaluation Method*s. A + marks com-

Table 1: Evaluation Methods and Ontology Layers.

Ontology Scope		Ontology Layer	Evaluation Method		
			Data-driven	Golden Standard	Structure Analysis
	Application	Operation			
		Population	+	-	+
		Application			
	Conceptual	Context	-	-	+
		Metadata	-	-	+
Domain		Taxonomy/ Structure	-	+	+
		Vocabulary	+	+	-

binations if there are *Evaluation Method*s that calculate metrics for the respective *Ontology Layer*. In the case of a -, there is no support. *Operation* and *Application* layers are not considered. Data Driven and Golden Standard based methods rely on additional data – either provided by domain specific text corpora or by domain specific best practice ontologies (golden standard). In consequence, these methods could be applied by a domain-specific non-universal platform for ontology evaluation if there is an agreement on appropriate text corpora or golden standard ontologies. Vice versa, a platform like OntoMetrics provides little coverage for the *Domain Scope* unless the required additional data is accessible. *Application Scope* is also poorly covered as stated before with regards to the validation problem. Thus, the main focus of OntoMetrics is on the *Conceptual Scope* and hence on *Conceptualisation*, *Formalization* and *Integration* phases of the ontology life-cycle.

3 OntoMetrics: STATUS AND USAGE

At its current state, the OntoMetrics platform has the following functional areas:

- A web-interface to upload OWL ontologies in RDF-XML representation and to calculate a set of *Ontology Quality Metrics* for them.
- An XML-download of calculated *Ontology Quality Metrics*.
- A wiki that explains the semantics and the calculation of the *Ontology Quality Metrics*.

Table 2 shows the metrics that can currently be calculated using OntoMetrics in addition to the standard OWL-API counting metrics. There are four general types of metrics. *Schema Metrics* take the special meaning of the OWL-Schema-definition constructs into account for the calculation of metrics on the ontology structure. *Graph Metrics* are metrics that can be generally applied to graphs (esp. trees). In the case of ontology evaluation, they are calculated for the taxonomy tree of the ontology. *Knowledgebase Metrics* do not only assess the type structure of an ontology but also instances that populate the ontology. At last, *Class Metrics* narrow the focus to single class evaluation. The class metrics which have been adopted from Tartir et al. (Tartir et al., 2005) cause a higher computational effort. Thus, they can be excluded from calculation if required.

For an explanation of the calculated metrics, the reader can refer to the OntoMetrics-Wiki or to the given sources in the table. Although mainly two sources are given with Gangemi et al. (Aldo Gangemi, Carola Catenacci, Massimiliano Ciaramita, Jos Lehmann,) and Tartir et al. (Tartir et al., 2005), most of these metrics have also been proposed by other authors or are basic graph metrics.

The table-head contains quality criteria as mentioned in section 2. For more comprehensive semantics of the criteria, refer to (Pak and Zhou, 2011). Those quality criteria have been selected, where a correlation to the calculated metrics has been proposed in the literature. Thus, for each of the quality criteria at least one metric is available in conjunction with the proposed direction of correlation. '$+$' means positive correlation and '$-$' negative correlation. As expected from the previous discussion, quality criteria that are relevant within the *Domain* and the *Application Scope* are under-represented. Furthermore, the Metric-Criteria-Matrix provided by table 2 is sparsely filled. Some of the metrics have just been proposed as an indicator for good ontology quality without explicitly naming *Quality Criteria* (e.g. in (Tartir et al., 2005)). In consequence, there is a lot of room for further research on ontology metrics. The following usage scenarios in research and practice are seen for the OntoMetrics platform:

1. Empirical validation of proposed correlations between metrics and quality criteria regarding strength and significance.
2. Determination of influences like the ontology usage context on these correlations.
3. Determination of of best practise metric profiles and values for certain domains and usage contexts.
4. Analysis of Domain Specific Languages (DSL) and Models formulated in these languages if an OWL representation is available.[3]

[3]For example Archimate models from the Enterprise Architecture Domain can be transformed to OWL using the toolset provided by the Timbus project: https://opensourceprojects.eu/p/timbus/context-model/converters/wiki/Home/

5. Practical ontology quality assessment by calculating validated metrics.
6. Practical ontology quality assessment by monitoring anomalies in calculated metric values.
7. Proposal and validation of new metrics and their application by providing them on OntoMetrics.
8. Internal collection of evaluated ontologies for later calculation and validation of new metrics or theories regarding ontologies.

OntoMetrics supports these usage scenarios by providing easy, reliable and repeatable access to metric calculation. The XML-Representation of the calculation can be used for further automated processing of the results. Additionally, the wiki gives orientation regarding the application of already implemented metrics and also room for the discussion of new ideas. Researchers are invited to contribute to the platform with their own proposals of quality metrics. This can be done by presenting a metric proposal for discussion in the wiki or by providing an implementation that can be included in the platform and thus would be available for validation and use on a broad scale.

4 CONCLUSION AND OUTLOOK

At its present state, OntoMetrics is a lightweight, handy tool for comparable, metric based ontology evaluation. In combination with the Quality-Metrics-and-Criteria-Matrix (table 2) it can be used to assess ontologies using already accepted and suggested metrics. The main focus lies on the *Conceptual Scope*. However, *Domain Scope* and *Application Scope* are partially covered as described in section 2. The wiki provides information on the theoretical background of the calculated metrics. For the future development of the OntoMetrics platform three main directions are planned:

Enhance the Knowledgebase of OntoMetrics. The information in the wiki regarding the proposed metrics is planned to be extended by additional sources, use case descriptions and systematizations. For example, the Quality-Metrics-and-Criteria-Matrix (table 2) can be extended. Additionally, other systematizations of ontology metrics compared to Schema, Graph, Knowledgebase and Class Metrics can be used. Sources for this additional knowledge regarding ontology metrics are already existing and new literature as well as validation and experience reports based on metric calculation using OntoMetrics. Furthermore, suggestions of new metrics are possible within the wiki.

Table 2: Quality-Metrics-and-Criteria-Matrix.

Metric	Accuracy	Understandability	Cohesion	Comp. Efficiency	Conciseness
Schema Metrics[1]					
Att. Richness [2]	+ [2]				
Inh. Richness [2]	+ [2]				
Rel. Richness [2]	+ [2]				
Att. Class Ratio					
Equiv. Ratio					
Axiom/Class Ratio					
Inv. Rel. Ratio					
Class/Rel. Ratio					
Graph Metrics [1]					
Abs. root card.			+ [3]		
Abs. leaf card.		+[4]	+ [3]		
Abs. sibling card.					
Abs. depth		- [1]			
Av. depth	+ [5]	- [1]			
Max. depth	+ [5]	- [1]			
Abs. breadth		- [1]			
Av. breadth	+ [5]	- [1]			
Max. breadth	+ [5]	- [1]			
Leaf fan-out ratio					
Sibl. fan-out ratio					
Tangledness		- [1]		- [1]	
Total no. of paths					
Av. no. of paths					
ADIT-LN			+ [3]		
Knowledgebase Metrics [2]					
Average Population					+ [2]
Class Richness					+ [2]
Class Metrics [2]					
Cl. connectivity					
Cl. fulness					
Cl. importance					
Cl. inh. richness	+ [2]				
Cl. readability		+ [2]			
Cl. rel. richness	+ [2]				
Cl. children count					+ [2]
Cl. instances count					
Cl. properties count					

[1](Gangemi et al., 2005), [2](Tartir et al., 2005), [3](Yao et al., 2005), [4](Lantow and Sandkuhl, 2015), [5](Fernández et al., 2009).

Provide Additional Quality Metrics. New metrics of the known types can be added using the provided API. Furthermore, a support of data driven ontology metrics calculation could be implemented by allowing the upload of text corpora or supposedly golden standard ontologies on the platform. Furthermore, a WordNet integration in order to assess the clarity of labels used in the ontology as suggested by Burton Jones et al.(Burton-Jones et al., 2005) could be implemented

Extend the Base Functionality of OntoMetrics. In order to allow tool integration of OntoMetrics, a Web-Service that provides metric calculation functionality is planned. Furthermore, the presentation and analysis of the resulting metric values can be improved. For example, classes can be filtered or sorted using criteria like class importance. Also, a special treatment of imports and ontology modules would be possible. Thus, a tool for the analysis of complex ontologies would become available that might not be used for ontology evaluation only, but also for exploring ontology structures and deriving new theories regarding the evolution of ontologies.

OntoMetrics is open for contributions of new metric suggestions and implementations as well as discussions of existing metrics.

ACKNOWLEDGEMENTS

Special thanks to Martin Lichtwark, Michael Poppe, and Julian Heckmann for their efforts to make OntoMetrics available as an on-line platform.

REFERENCES

Alani, H., Brewster, C., and Shadbolt, N. (2006). Ranking ontologies with AKTiveRank. In *International Semantic Web Conference*, pages 1–15.

Aldo Gangemi, Carola Catenacci, Massimiliano Ciaramita, Jos Lehmann. Ontology Evaluation and Validation: An integrated formal nmodel for the quality diagnostc task.

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

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

Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Peng, Y., Reddivari, P., Doshi, V. C., and Sachs, J. (2004). Swoogle: A semantic web search and metadata engine. In *Proc. 13th ACM Conf. on Information and Knowledge Management*, volume 304.

Fernández, M., Overbeeke, C., Sabou, M., and Motta, E. (2009). What makes a good ontology? A case-study in fine-grained knowledge reuse. In *Asian Semantic Web Conference*, pages 61–75.

Fernández-López, M., Gómez-Pérez, A., and Juristo, N. (1997). Methontology: from ontological art towards ontological engineering.

Gangemi, A., Catenacci, C., Ciaramita, M., and Lehmann, J. (2005). A theoretical framework for ontology evaluation and validation. In *SWAP*, volume 166.

Gavrilova, T., Gorovoy, V., and Bolotnikova, E. (2012). New Ergonomic Metrics for Educational Ontology Design and Evaluation. In *SoMeT*, pages 361–378.

Guarino, N. and Welty, C. A. (2009). An overview of OntoClean. In *Handbook on ontologies*, pages 201–220. Springer.

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

Lantow, B. and Sandkuhl, K. (2015). An Analysis of Applicability using Quality Metrics for Ontologies on Ontology Design Patterns. *Intelligent Systems in Accounting, Finance and Management*, 22(1):81–99.

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

Noy, N. F., McGuinness, D. L., et al. (2001). Ontology development 101: A guide to creating your first ontology.

Pak, J. and Zhou, L. (2011). A Framework for Ontology Evaluation. In Sharman, R., Rao, H., and Raghu, T., editors, *Exploring the Grand Challenges for Next Generation E-Business*, volume 52 of *Lecture Notes in Business Information Processing*, pages 10–18. Springer Berlin Heidelberg.

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

Poveda-Villalón, M., Suárez-Figueroa, M. C., and Gómez-Pérez, A. (2012). Validating ontologies with oops! In *International Conference on Knowledge Engineering and Knowledge Management*, pages 267–281.

Tartir, S., Arpinar, I. B., Moore, M., Sheth, A. P., and Aleman-Meza, B. (2005). OntoQA: Metric-based ontology quality analysis. In *IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources*, volume 9.

Yao, H., Orme, A. M., and Etzkorn, L. (2005). Cohesion metrics for ontology design and application. *Journal of Computer science*, 1(1):107.