

An Evolutional Based Data-Driven Quality Model for Ontologies

Achim Reiz¹ [0000-0003-1446-9670]

¹ Rostock University, 18051 Rostock, Germany
achim.reiz@uni-rostock.de

Abstract. The use of ontologies today is still unabated, with upcoming usage scenarios through the rise of artificial intelligence. That increases the importance and the need for automatic, reliable evaluation techniques.

Even though numerous ontology metrics have been proposed in the past years, the interpretation of these metrics remains arbitrary. The specific influence of these simple metrics on quality attributes is not validated in a scientifically sound approach. The goal of this doctorate is to establish and validate a link between comprehensive quality attributes like “understandability” or “completeness” and the metrics proposed in the literature. Using a data-centric research design, the objective is the identification of quality grades and improvement recommendations through the application of a novel, data-driven quality framework. This has the potential to support especially inexperienced ontology engineers in assessing their work and the creation of better ontologies.

The novelty of this research lies in the data-centricity of its design. Using a collection of large amounts of evolutional ontology metric data, statistical relevant correlations between these metrics over time are to be found. This enables the validation of already proposed quality attributes and the identification of new ones.

Keywords: Ontology evaluation, ontology quality, ontology metrics

1 Introduction / Problem Statement

Even though the use of computational ontologies dates back until the 90s [1], their usage today is still unabated with application scenarios not only in the sharing of information but also in the inferring of new knowledge through reasoning or the enablement of natural language processing. Trends like explainable A.I [2] autonomous driving [3] or graph driven voice assistants like amazon alexa [4] further increase the use of ontology technologies. But as ontological representations are on the rise – how can one ensure the quality and correctness of such artifacts?

The usage of objective, quantifiable metrics provides reliable ontological measurements. Moreover, in the past, an enormous amount of metrics has been proposed, assessing – among others - the graph and schema attributes, annotations, relations, or instances. However, for most of the proposed measurements, their impact on the actual

quality is not researched systematically. For example, how, and in what kind of composition do the metrics proposed by Tartir et al. in the OntoQA-Framework [5] influence the understandability of an ontology? How does Gangemi et al.’s graph metrics [6] influence the reusability of an ontology? The impact of specific metrics on particular quality attributes is often not described and if so, not validated in an empirically sound approach. Even though the missing validation of the proposed metrics is criticized in many papers [7–9], it has not yet experienced great attention from the research community. Further, most of the metrics stay isolated due to their heterogeneous structure, degree of formality, and different objectives [10].

These shortcomings hinder the explanatory power and comprehensibility of ontology metrics. Especially inexperienced modelers are facing challenges selecting the right metrics for the right goals. Even though the ontology metrics are calculated objectively, their interpretation remains subjective [11].

Validated measurements of ontology quality attributes can help modelers to develop high-quality ontologies based on their aimed usage scenario. The envisioned purpose of this Ph.D. project is a translation of the abstract metrics into measurements for high-level quality dimensions like, among others, “completeness”, “clarity”, or “adaptability”. As a side effect, based on the calculated quality score, improvement recommendations can be derived, highlighting the artifacts that are the most influential factors for each quality dimension. In effect, it is expected that this can not only lead to better ontologies but in the long term also to better-trained modeling workforce.

2 Related Work

Related work for this research endeavor originates from the field of ontology evaluation. Over the past years, various evaluation methods have been proposed. The following section discusses the most influential and relevant publications and motivates this doctoral research through the shortcomings of the earlier approaches.

Often referenced is the categorization by Burton-Jones et al. into syntactic, semantic, pragmatic, and social quality within a total of 10 associated metrics [7]. This paper already assigns metrics to quality constructs like lawfulness or clarity. Nevertheless, the framework heavily depends on user inputs, one has to provide weights for the metrics and their aggregation. The overall 14 user-assigned weights make the application of this quality framework arbitrary, the practical implications for this framework regarding attributes like ontology reuse, understandability, or computational efficiency are not examined.

OntoQa by Tartir et al. provides another set of metrics categorized in schema, class, and knowledge base metrics. Even though an interpretation is provided for most of the results, a holistic view on the ontology is missing and the metrics stay isolated [5].

Duque-Ramos et al. adapted the software quality framework SQuaRE towards ontologies, naming it OQuaRE. Here, ontologies are measured using 14 metrics. For these measurements, threshold values are provided for the grading into a numerical scale from 1 to 5. Further, these metrics are mapped to some quality characteristics like testability or modularity [12]. Using an expert evaluation, the relation between the quality

attributes and the calculated metrics were empirically validated [13]. The maturity of this framework regarding the validation of metrics and their linkage to quality attributes exceeds the other approaches. However, even though the relationship between metrics and quality dimensions were established, this relationship merely associates the calculated metrics to quality attributes, without proposing a composition.

A recent review on metrics by Lourdasamy and John came to the same conclusion and criticized the missing holistic view on ontologies as well as the lack of empirical validation [14].

Most of the current approaches just hypothesize the combination of the isolated metrics into intuitive, comprehensive quality aggregations. Even if a validation is performed, it is often done in a rather narrow quantitative study with limited significance. In conclusion, metrics for ontologies are already researched extensively, the room for novelties in that area is limited. To ensure a useful application though, these proposed metrics need to be validated and set into a useful context. This context can be provided by the development of a holistic ontology quality framework as proposed in this Ph.D. proposal.

3 Research Questions

In the previous sections, current challenges were outlined regarding the support of (especially inexperienced) knowledge engineers using empirically validated quality measurements and improvement recommendations. The following research questions are derived based on the identified shortcomings of the approaches currently available:

3.1 RQ 1: How do Ontology Metrics Develop over the Evolution of Ontologies?

Ontology metrics just provide a snapshot of an artifact that is itself dynamic. Taking into account historical data, the evolvement of the ontology can be made visible. As Malone et al. stated, ontologies develop in different stages, that all come with specific characteristics regarding the performed changes [15]. It is expected that these stages can be identified using the historical development of the calculated metrics. The maturity level of an ontology then indicates the kind of assessment and evaluation that is needed to rate and improve the model.

3.2 RQ 2: How do Ontology Metrics Correlate with Each Other, and how do these Correlations form Comprehensive Quality Dimensions?

The next research question is concerned with the correlations between metrics. The combination of multiple isolated metrics into comprehensive quality attributes is expected to raise the understandability of the evaluation. The validation based on a statistical analysis of a large metric repository ensures their significance. Further, the decrease in the amount of shown measurements can improve the comparability between ontologies.

3.3 RQ 3: Under what Condition is the Quality of an Ontology Sufficient for a Given Use-Case?

To be able to infer a quality indication out of measurements, an interpretation of these values is obligatory. Without guidance, their interpretation remains arbitrary, especially for inexperienced ontology engineers. Based on the analysis of large ontology repositories, their use-cases, and their respective threshold-values for other metrics, the goal is to research quality scores that are common for often used knowledge representations. These common threshold values can be taken as a threshold recommendation for future ontology developments.

3.4 RQ 4: Which Improvement Recommendation can be Derived from the Metric Calculations?

The awareness of the quality of an artifact is indispensable for its improvement. However, especially for inexperienced knowledge engineers, a pure numerical assessment is not sufficient to ensure the creation of better ontologies. It is argued that the inexperienced workforce needs more support in the form of recommendations. Based on the quality attribute calculations that are the output of RQ 2 and their suggested scores that are the output of RQ 3, the goal is to derive modeling recommendations out of the assessed characteristics. These scores enable the modeler to fix the weaknesses in his ontology design that have the most significant impact on the respective quality score.

4 Research Plan and Preliminary Results

This section highlights the current methodological plan, as well as steps that have been achieved so far. Especially the former is subjected to changes in the future, as the maturity of this research grows.

4.1 Development of the Technical Prerequisites

As stated in section 2, most of the current approaches are based on argumentative quality aggregations and limited statistical validation. The approach proposed for this doctorate follows a more data-centric paradigm. Based on the extensive analysis of sizeable evolutionary data sets, metric correlations out of literature shall either be validated or new ones detected. The data will be collected through the tool “OntoMetrics” of Rostock University. First proposed in 2016 by Lantow [16], it offers 81 metrics, mostly based on the work of [17] and [5]. While it currently just supports the analysis of single ontology files, in the future, the tool will be extended towards support for git repositories and an easy to use GitLab and GitHub¹ interface. All metrics of the analyzed data

¹ gitlab.com, github.com

sets will be stored for analysis. It is expected that the further development of the OntoMetrics-tool increases its usage and enables a growing database of analyzed ontologies. Recently, a large-sized company approached Rostock University for collaboration in assessing their ontologies. The collaboration with this industry partner can also lead to a growing database of collected ontology metrics.

4.2 Methodical Blueprint for Answering the Research Questions

As soon as enough data is collected to infer valid correlations, the answering of the research questions can begin. The planned methodological process is shown in figure 1 below. As the research is still in an early stage, the outlined process may be subjected to changes in the future.

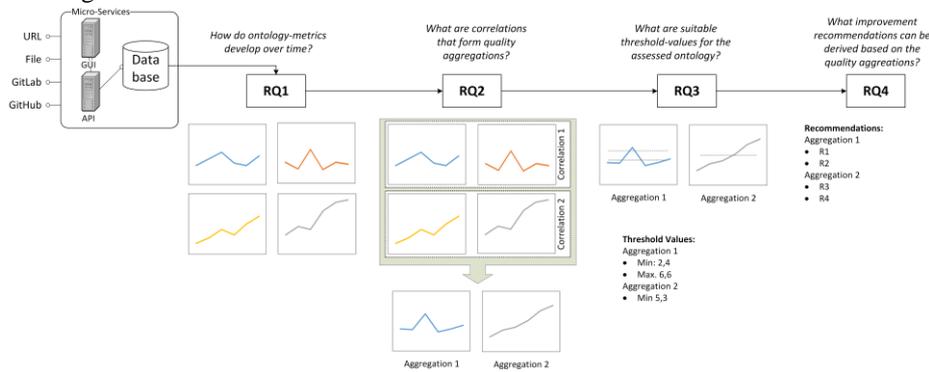


Fig. 1. Research Questions, the planned methodological proceedings, and their outcomes

The research question 1 is at first concerned with the description of the gathered data. An initial overview of how the various ontologies differ can provide an insight into the variance of metrics as well as their historical development. It guides the deriving of the metric aggregations for the next research question, RQ2. These aggregations shall include compositions of metrics into comprehensible quality dimensions like “understandability”, “reusability”, “learnability” and more. The creation of the metric compositions can be based on mere data analysis using statistical data mining approaches, from already proposed metric aggregations in the literature or a combination of both.

The distinctive feature between this and previous research is the intense focus on the data-centric validation of the measurements. Using the various analyzed ontologies, it is expected to see that identified commonalities are to be true not only for a limited set of ontologies but for a statistically relevant proportion of the ontologies that are available for analysis.

As the measurements are identified and validated, the next step for RQ3 is now to find common, recommended threshold values for these metrics. Most likely, a prior classification of the ontology is necessary – an upper-level ontology will look different than a task-dependent domain ontology. With the metrics repository, it is expected to derive threshold values for evaluation scores for the quality framework. An example of

this threshold could be that most mature, often used domain ontologies have an “understandability” of at least 4.3. This enables the ontology engineer to directly compare his work against other ontologies that he knows or that are used in his organization, while at the same time having an idea when a desirable quality level is reached.

The metrics compositions that are the output of RQ2 are in the latter also the input for RQ4. As ontologies get analyzed using these measurements, there are metrics within these compositions that may have a significant impact on the calculated score. They can be, therefore, identified as the most essential items to improve.

5 Evaluation of the Research

The created metric compositions for the measurement of quality attributes of RQ2 and the created threshold values of RQ3 are based on statistical analysis. Using further collected test-data, the mathematical validity of the metric aggregations can be proven. But a mere mathematical construct does not guarantee the usefulness to the consumer of such metrics. To perform this kind of user-centric evaluation and therefore ensure that the research questions are answered to a sufficient degree, a last evaluation stage is required. The embodiment of this affirmative study is not yet decided. There are, however, various ideas for confirmative studies.

An often-used approach for the validation of usefulness to the user is a quantitative survey. Here, it is possible to question the perceived level of experience, as well as the perception of the usefulness of the various quality measurements.

In the beginning, we set up the hypothesis that especially inexperienced users tend to need automatic supported guiding using an empirically grounded evaluation tool. To connect the perceived impact of the tool with reality, the survey could be aligned with metrics and metadata out of the git repositories. For some metrics, the metadata can give a clear indication for its validity: Reusability, for example, should be visible in downloads or forks. If the questionnaire before also captured the user name of a given repository service, then a direct link between the perceived quality of individual users and their performance can be established. Higher perceived productivity should affect the number and size of commits. An improved perceived personal performance should be observable through more elaborate modeling techniques, thus resulting in an improved metric rating for their contributions. Using this approach, the actual effect of the proposed metrics on the performance of the modeler can be observed.

6 Discussion and Further Outlook

At a first glimpse, the field of ontology evaluation seems to be mature. The technology exists since the early 90s, and much work has been published. However, as outlined in section 2, today’s ontology evaluation approaches lack a sound empirical validation and the translation into human understandable, reliable metrics. This shortcoming is getting addressed with this doctorate. This research can help especially inexperienced knowledge engineers to create better ontologies by providing comprehensible, reliable feedback, including improvement recommendations.

This research is focused on the structural attributes of ontologies. Measurement methodologies that require additional input like golden standard, data-driven, or task/application based [18] are not in the scope of this research. It is therefore not assessed whether an ontology that might be not sufficient from a structural perspective excels in a specific usage scenario. It is likewise not planned to include user ratings or domain terminology coverage into the quality measurements. Nonetheless, further research might consider the combination of different methodologies with the result of this scientific work.

The collection of data is the main requirement for the planned research activities. As stated in section 4, the tool “OntoMetrics” will be the foundation for further data collection activities and is getting developed towards the integration of historical data and repository services. As soon as these services are integrated, public available ontologies like the gene ontology² can be analyzed towards a first outlook on RQ1, while the tool builds up a more extensive user base and therefore collects more data.

The acceptance of the tool OntoMetrics and the associated data gathering is an essential activity for the success of this research. The size and quality of the database will strongly affect the quality and validity of the research, as well as the next steps. Recently, the respective tool has gained some interest not only by researchers but also by industry partners, further collaboration is planned. This supports our hypothesis for a need for better evaluation of the growing amount of ontologies and motivates further the optimism for the success of this data-driven ontology evaluation research.

Acknowledgment. I would like to give special thanks to Prof. Kurt Sandkuhl for his mentoring and support.

References

1. Ginsberg, M.L.: Knowledge interchange format. The KIF of death. *AI Magazine* **12**, 57–63 (1991)
2. Futia, G., Vetrò, A.: On the Integration of Knowledge Graphs into Deep Learning Models for a More Comprehensible AI—Three Challenges for Future Research. *Information* (2020). <https://doi.org/10.3390/info11020122>
3. Kaleeswaran, A.P., Nordmann, A., Mehdi, A.u.: Towards Integrating Ontologies into Verification for Autonomous Driving. In: Suárez-Figueroa, M.C., Cheng, G., Gentile, A.L., Guéret, C., Keet, C.M., Bernstein, A. (eds.) *Proceedings of the ISWC 2019 Satellite Tracks, Auckland, New Zealand, 10/26/2019 - 10/30/2019* (2019)
4. Kollar, T., Berry, D., Stuart, L., Owczarzak, K., Chung, T., Mathias, L., Kayser, M., Snow, B., Matsoukas, S.: The Alexa Meaning Representation Language. In: Bangalore, S., Chu-Carroll, J., Li, Y. (eds.) *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics. Human Language Technologies, New Orleans, U.S., 06/01 - 06/06/2018*, pp. 177–184. Association for Computational Linguistics (2018). <https://doi.org/10.18653/v1/N18-3022>

² <https://github.com/geneontology>

5. Tartir, S., Arpinar, B., Moore, M., Sheth, A.P., Aleman-Meza, B.: OntoQA: Metric-Based Ontology Quality Analysis. In: IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources, Houston, 11/27/2005 (2005)
6. Gangemi, A., Catena, C., Ciaramita, M., Lehmann, J.: A theoretical framework for ontology evaluation and validation. In: Bouquet, P., Tummarello, G. (eds.) *Semantic Web Applications and Perspectives*, Trento, Italy, 12/14/2005 - 12/16/2005. CEUR (2005)
7. Burton-Jones, A., Storey, V.C., Sugumaran, V., Ahluwalia, P.: A semiotic metrics suite for assessing the quality of ontologies. *Data & Knowledge Engineering* (2005). <https://doi.org/10.1016/j.datak.2004.11.010>
8. Alm, R., Kiehl, S., Lantow, B., Sandkuhl, K.: Applicability of Quality Metrics for Ontologies on Ontology Design Patterns. In: Filipe, J. (ed.) *Proceedings of the 5th International Conference on Knowledge Engineering and Ontology Development (IC3K)*, Vilamoura, Algarve, Portugal, 9/19/2013 - 9/22/2013, pp. 48–57. SciTePress, S.I. (2013). <https://doi.org/10.5220/0004541400480057>
9. Ashraf, J., Hussain, O.K., Hussain, F.K.: A Framework for Measuring Ontology Usage on the Web. *The Computer Journal* (2013). <https://doi.org/10.1093/comjnl/bxs134>
10. Sicilia, M.A., Rodríguez, D., García-Barriocanal, E., Sánchez-Alonso, S.: Empirical findings on ontology metrics. *Expert Systems with Applications* (2012). <https://doi.org/10.1016/j.eswa.2011.11.094>
11. Hammar, K.: *Content Ontology Design Patterns: Qualities, Methods, and Tools*. P.h.D., Linköping University (2017)
12. Duque-Ramos, A., Fernández-Breis, J.T., Stevens, R., Aussenac-Gilles, N.: OQuaRE: A square-based approach for evaluating the quality of ontologies. *Journal of Research and Practice in Information Technology* **43**, 159–176 (2011)
13. Duque-Ramos, A., Fernández-Breis, J.T., Iniesta, M., Dumontier, M., Egaña Aranguren, M., Schulz, S., Aussenac-Gilles, N., Stevens, R.: Evaluation of the OQuaRE framework for ontology quality. *Expert Systems with Applications* (2013). <https://doi.org/10.1016/j.eswa.2012.11.004>
14. Lourdasamy, R., John, A.: A review on metrics for ontology evaluation. In: *Proceedings of the 2nd International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, 01/19 - 01/20/2018, pp. 1415–1421. IEEE (2018). <https://doi.org/10.1109/ICISC.2018.8399041>
15. Malone, J., Stevens, R.: Measuring the level of activity in community built bio-ontologies. *Journal of biomedical informatics* (2013). <https://doi.org/10.1016/j.jbi.2012.04.002>
16. Lantow, B.: OntoMetrics: Application of on-line ontology metric calculation. In: Johansson, B., Vencovský, F. (eds.) *Joint Proceedings of the BIR - Workshops and Doctoral Consortium*, Prague, Czech Republic, 09/14 - 09/16/16 (2016)
17. Gangemi, A., Catenacci, C., Ciaramita, M., Lehmann, J., Gil, R., Bolici, F., Strignano Onofrio: Ontology evaluation and validation. An integrated formal model for the quality diagnostic task (2005)
18. Brank, J., Grobelnik, M., Mladenic, D.: A Survey of Ontology Evaluation Techniques. In: *Proceedings of the conference on data mining and data warehouses. SiKDD 2005*, Ljubljana, Slovenia, pp. 166–170 (2005)