

Chapter 1

Standard for Evaluation

We derive a set of evaluation principles from [DGFE16; KKRK10; BBL76] and consider them in designing our quality indicators. These works summarize as:

- The work in [DGFE16] introduced four standards to provide a framework for determining a good evaluation for people and organizations involved in the evaluation society. The standards claim to substantiate the idea of a professional evaluation. The evaluation principles are (1) usefulness, (2) feasibility, (3) fairness, and (4) accuracy. Each principle is presented with a group of sub-statements. We consider all the mentioned principles and adapt the respective sub-statements. Note that some sub-statements cannot be aligned in our context, as they are related to human evaluation, such as human error checking.
- The work in [KKRK10] presented a criteria model for Semantic Web Services (SWS) evaluation. The model comprises five dimensions of evaluation: (1) performance/scalability, (2) usability/effort, (3) correctness, (4) coupling, and (5) supported scope and automation. We classify them as sub-statements within the introduced principles in [DGFE16] and adjust them in our context. We consider them as desirable sub-statements rather than as obligatory ones because they are introduced as criteria (not principle).
- The work in [BBL76] established a conceptual framework for analyzing the characteristics of software quality. We adapt the respective characteristics as sub-statements within the introduced principles in [DGFE16].

Accordingly, we address the following principles and sub-statements (given by P_i) for the quality evaluation of the merged ontology:

- **Usefulness principle:**

P1. Detecting the involved and affected variables: Issues affecting evaluation and those affected by the evaluation should be identified beforehand to take them into account as far as possible when creating the evaluation.

- In our context, analyzing the source ontologies should be considered, since they are affected by the evaluation and the merge results. The application that wants to use the evaluated ontology should take into account, too, because it affects the evaluation.

- P2. Transparency of the impact of influenced variables:* The impact of those involved in and those affected by the evaluation should be transparent and clear. These values can be used for the classification and interpretation of the results.
- The importance degree of the source ontologies and the application that wants to use the merged ontology should be clear to use these degrees in concluding the result and interpreting the output.
- P3. Clarification of evaluation purposes:* The purpose of the evaluation is varied with quality dimensions. Thus, the purpose of each aspect of quality dimensions should be clarified separately for the users.
- The purpose of each evaluation aspect should be clarified and defined for the user.
- P4. Competence and credibility of the underlying system:* The system (environment or framework) that conducts evaluations should be competent. Thus, the result of the evaluation can be acceptable.
- The reliability of the evaluation tool and framework should be approved.
- P5. Selection and scope of indicators:* The selection and scope of indicators should be considered based on evaluation goals.
- Users should be able to adjust the scope of the evaluation. Thus, they can adjust desired quality indicators based on the purpose of the given task.
- P6. Supported scope and automation:* The feasibility of the automated evaluation is admired. However, it depends on the supported aspects of the framework. An automated ontology evaluation is a necessary precondition [BGM05] for the healthy development of automated ontology processing techniques.
- Evaluation is desired to be automatic as possible for the given quality dimension.
- P7. Completeness of the output:* The result of the evaluation should be complete and comprehensible.
- The output of evaluation should provide all essential quality aspects by covering different quality indicators in order to be comprehensive and complete.
- P8. Clarity of the result:* The result of the evaluation should be clear and understandable for the user.
- Presenting the result of the evaluation should be understandable for the user. The understandability of the result can be augmented with a user-friendly GUI with ease of use feature.
- P9. Use and benefits of evaluation:* The results should be useful, and the users should get benefit from the evaluation.
- The evaluation criteria are intended to emphasize existing gaps and weaknesses to provide better insight for the user and suggest a possible

solution when it is applicable. Accordingly, the user could get benefits from evaluation results.

- **Feasibility principle:**

P10. Appropriate methods: Evaluation procedures should be chosen in such a way that, on the one hand, the evaluation is carried out professionally in accordance with the requirements. On the other hand, the effort for those involved and those affected is kept in an appropriate ratio to the intended benefit of the evaluation.

→ The evaluation methods should consider the requirements and keep in balance the effort for analyzing source ontologies (those involved in) and the application that wants to use the merged ontology (those affected by) with an adequate ratio to the desired evaluation output.

P11. Efficiency of evaluation: The effort for evaluation should be in reasonable proportion to the benefits of the evaluation.

→ The amount of effort on producing the desired results should be feasible in practice (with a reasonable complexity), and the output of the evaluation should be in appropriate with this complexity.

P12. Usability/Effort: The usability of the framework is desired to be evaluated in terms of the amount of effort that is required to use the framework, set it up, and maintain it.

→ The modeling effort for the evaluation framework is desired to be in balance with its usage and maintenance.

P13. Performance/scalability: The runtime performance and scalability of the evaluation are desired to scale well.

→ Scalability can be evaluated explicitly by exploring the degradation of the runtime performance of the evaluation functions with the increasing size of the merged ontology.

- **Fairness principle:**

P14. Formal disposal of indicators: The quality indicators, process, and goal of the evaluation should be written and available for users.

→ Each quality indicator is expected to have an exact and systematic definition, certified goal, and precise implementation of the process in practice. They should be available for users.

P15. Disclosure of results: The results of the evaluation should be made available.

→ The output of the evaluation should be available for users.

P16. Comprehensive and fair examination: Evaluations should examine and present the strengths and weaknesses of the object that wants to be evaluated as fairly and comprehensively as possible.

- In the quality evaluation of the merged ontology, both strengths and weaknesses should be reported to users. The evaluation should be comprehensive by covering various aspects.

P17. Impartial implementation and output: The evaluation process and result should be impartial and unbiased with respect to quality indicators or environments. Likewise, coupling criteria [KKRK10] evaluates the level of the framework's coupling with the effect on its modifiability.

- The merge evaluation techniques should be independent of the underlying merge method and environment.

- **Accuracy principle:**

P18. Description of the object to be evaluated: The theoretical aspect of the object that wants to be evaluated should be described and documented precisely and comprehensively.

- The theoretical aspect of the given merged ontology and the way that it is created upon should be comprehensively and accurately described.

P19. Description of purposes and procedures: The purposes and procedures of the evaluation itself, including the underlying evaluation methods, should be documented and described so precisely that they can be understood and assessed.

- For each aspect that the user wants to evaluate, there should be a well-documented description of its objective and implementation of the evaluation's function.

P20. Context analysis: The context of the object that wants to be evaluated should be analyzed in a sufficiently comprehensive and detailed manner and taken into account when interpreting results.

- The merged ontology should be evaluated via some criteria for a given context.

P21. Declaration of indicators: The quality indicators that are used for evaluation should be documented with sufficient accuracy so that the reliability and appropriateness of them can be assessed.

- On what basis the indicators are defined should be documented.

P22. Valid and reliable indicators: The quality indicators should be valid and reliable.

- On what basis the indicators are defined should be valid.

P23. Justified assessments and conclusions: The evaluative statements made in an evaluation should be based on explicit criteria and target values. Conclusions should be explicitly justified based on the given inputs so that they can be understood and assessed.

- Concluding the quality of the merged ontology should be based on the respective source ontologies, their mapping, and parameters.

P24. Meta-evaluation: Meta-evaluations should evaluate evaluations. In order to make this possible, evaluations should be documented, archived, and made accessible as far as possible in a suitable form.

→ A meta-evaluation on the output of the evaluation framework should be carried.

P25. Correctness: The evaluation framework is desired to be correct and acts precisely as the user acts on behalf of.

→ The output of the evaluation should be correct against an available benchmark or human evaluation.

We consider these sub-statements of the principles in designing our quality indicators.

Bibliography

- [BBL76] B. W. Boehm, J. R. Brown, and M. Lipow, "Quantitative evaluation of software quality", in *Proceedings of the 2nd international conference on Software engineering*, IEEE Computer Society Press, 1976.
- [BGM05] J. Brank, M. Grobelnik, and D. Mladenic, "A survey of ontology evaluation techniques", in *Proceedings of the conference on data mining and data warehouses (SiKDD 2005)*, Citeseer Ljubljana, Slovenia, 2005, pp. 166–170.
- [DGFE16] E. Degeval-Gesellschaft Für Evaluation, *Standards für evaluation*, M.I.-.-.-. Die Deutsche Bibliothek – CIP Einheitsaufnahme DeGEval – Gesellschaft für Evaluation e.V. Standards für Evaluation-Erste Revision 2016, Ed., 2016.
- [KKRK10] U. Küster, B. König-Ries, and M. Klusch, "Evaluating semantic web service technologies: Criteria, approaches and challenges", in *Progressive Concepts for Semantic Web Evolution: Applications and Developments*, IGI Global, 2010, pp. 1–24.