Ontology Evaluation for the Web
Extended Abstract – PhD proposal

Denny Vrandečić

Institute AIFB, University of Karlsruhe, Germany
denny@aifb.uni-karlsruhe.de

1 Introduction

Ontologies are a pillar of the emerging Semantic Web. They capture background knowledge by providing relevant concepts and relations. Their role is to provide intensional knowledge in a machine processable way, and thus enable automatic aggregation and the proactive use of distributed data sources. Ontologies on the Semantic Web will come from a vast variety of different sources, spanning institutions and persons aiming for different quality criteria.

Ontology evaluation is the task of assessing the quality of an ontology. It answers the question: *How to measure the quality of an ontology for the web?*

Ontology evaluation will become an important task that must be addressed if ontologies are to be widely adopted, both in the Semantic Web and in other semantically enabled technologies. The following scenarios exemplify this:

- Users and machines facing a multitude of ontologies need to have a way of assessing them and deciding which one best fits their requirements, if any.
- People constructing an ontology need a way to evaluate their results and possibly to guide the construction process and any refinement steps. This will make the ontology engineers feel more confident about their results, and thus encourage them to share their results with the community and reuse the work of others for their own purposes.
- Automatic or semi-automatic ontology learning techniques also require effective evaluation measures, which can be used to select the "best" ontology out of many candidates, to tune the parameters of the learning algorithms appropriately, or to direct the learning process itself (if the latter is formulated as a path through a search space).

2 Related Work

The work on ontology evaluation has increased rapidly during the last few years. A lot of work was presented on the recent EON 2006 workshop[1], which was the first time an event focused on the topic of evaluating ontologies (instead of evaluating ontology-based technologies). The results of the workshop are not

[1] http://km.aifb.uni-karlsruhe.de/ws/eon2006

integrated in this proposal. We exemplify some approaches instead of giving a complete list of work in this field.

There is a set of approaches that evaluate ontologies implicitly through their **usage**. Typically, ontologies will be used in some application or task. The result of this application, or its performance on the given task, might depend on the used ontology. Thus one might argue that a good ontology is one which helps to produce good results. [11] is such an example. Such approaches to ontology evaluation have several drawbacks [8]: (1) the ontology is evaluated when used in a particular way for a particular task, but its difficult to generalize this observation; (2) the effect of the ontology on the outcome may be relatively small; (3) if evaluating a large number of ontologies, they must be sufficiently compatible that the application can use them all.

Other approaches are based on defining several decision **criteria** or attributes; for each criterion, the ontology is evaluated and given a numerical score. Additionally a weight is also assigned to each criterion, and an overall score for the ontology is then computed as a weighted sum of its per-criterion scores. In effect, the general problem of ontology evaluation has been deferred or relegated to the question of how to evaluate the ontology with respect to the individual evaluation criteria. [2] proposes an approach of this type, with ten simple criteria such as syntactical correctness, clarity of vocabulary, etc. [4] proposes another set of criteria, which is however geared more towards manual assessment and evaluation of ontologies. Their criteria involve: functional completeness generality, efficiency, perspicuity, precision/granularity, minimality, and others. [9] defines an even more detailed set of 117 criteria, organized in a three-level framework. The criteria cover various aspects of the formal language used to describe the ontology, the contents of the ontology (concepts, relations, taxonomy, axioms), the methodology used to construct the ontology, the costs (hardware, software, licensing, etc.) of using the ontology, and the tools available for working with the ontology. Many of the criteria are simple enough that the score of an ontology with respect to these criteria could be computed automatically or at least without much human involvement.

The evaluation based on a **gold standard** builds on the idea of using similarity measures to compare an ontology with an existing ontology that serves as a reference. This approach is particularly useful for the task of automatic learning of ontologies to evaluate the performance of the learning algorithms. The similarity between ontologies can be calculated using similarity functions [3]. The problem though is to define the reference ontology.

[1] suggests using a **data-driven** approach to evaluate the structural fit of an ontology to a corpus. (1) Given a corpus, a clustering algorithm is used to determine a probabilistic mixture model of hidden "topics" such that each document can be modeled as having been generated by a mixture of topics. (2) Each concept C of the ontology is represented by a set of terms including its name in the ontology and its hypernyms, taken from WordNet. (3) The probabilistic models obtained during clustering can be used to measure how well the concept C fits to each topic. (4) At this point, we require that concepts associated with

the same topic should be closely related in the ontology. This would indicate that the ontology is reasonably well aligned with the hidden structure of topics in the domain-specific corpus. A drawback of this method as an approach for evaluating relations is that it is difficult to take the directionality and type of relations into account.

Other approaches deal with the syntactic correctness or **logical consistency** of an ontology. They tackle well-defined properties like the satisfiability of ontologies [10] or a few well-described classes of conceptual mistakes, for example circularities in the concept hierarchies or partition problems [5]. Such questions are often covered by specific tools, like SWOOP[2] or ODEval[3].

A somewhat different aspect of **formal ontology evaluation** has been discussed with the Ontoclean methodology [6]. They point out several philosophical notions (essence, rigidity, unity, etc.) that can be used to better understand the nature of taxonomic relations, and to discover possible problems in the ontology's subsumption hierarchy. A downside of this approach is that it requires manual intervention by a trained human expert familiar with the above-mentioned notions such as rigidity; at the very least, the expert should annotate the concepts of the ontology with appropriate metadata tags, whereupon checks for certain kinds of errors can be made automatically. As pointed out e.g. in [7], applications where evaluation of this sort justifies the costs are probably relatively rare. However, [13] recently proposed an approach to automate the assignment of the metaproperty tags.

3 Proposed Approach

Although all the presented approaches have their merits and use-cases, most of them do not take the context of the Semantic Web into appropriate consideration. What does happen to ontologies, if tools gather them, mash them together, mix them, rip them apart, and push them around? How do ontology evaluation approaches deal with partial, changing, connected, or merged ontologies? To take an example: simple, and often suggested metrics like the depth of the class subsumption hierarchy can become victim to "intruding" axioms that simply add another class into a subsumption path, and thus change the metric without necessarily changing the semantics of the ontology for a given application.

The evaluation approaches presented in the previous section need to be examined with regards to their applicability to the Semantic Web. Thus a first step will be a synthesis of previous and current approaches in ontology evaluation, and related fields like software engineering (e.g. metrics like cohesion) or database schema development (e.g. normal forms). Experience may be reused from these related fields, but one needs to be careful when adapting these approaches to take the peculiarities of ontologies into account.

Based on this analysis new approaches need to be developed to extend the existing repertoire. Just to take one example: the Open World Assumption is

[2] `http://www.mindswap.org/2004/SWOOP/`

[3] `http://minsky.dia.fi.upm.es/odeval`

often ignored when evaluating ontologies, and thus it probably will be fruitful to examine the implications of the OWA with regards to the evaluation.

Ontologies will be shared and distributed and will have to fulfill numerous tasks within a multitude of different contexts – e.g. an ontology developed to express social connections could also be used for identity management. Thus ontology evaluation will have to be adaptable to the context of the given task. Capturing context and allowing the evaluation to include contextual information will be necessary to provide a meaningful evaluation. This also allows to take into account dynamic properties of the context that change over time, like the size of the domain, and thus, e.g., efficiency considerations with regards to reasoning.

This work will provide both a theoretical framework and a practical implementation for the context-based evaluation of ontologies for the Semantic Web.

4 Research Methodology

In order to assess the quality of the proposed ontology evaluation framework itself, and the approaches used within the implementation, we will regard four aspects: (1) the evaluation measures have to be either so obvious that they are generally accepted without much question; or (2) the results of the evaluation must correspond to the quality of measured ontologies; or (3) the adoption of the evaluation framework and its implementation within the Semantic Web community; or (4) the quality measures are derived from extensive experience reports of ontology based systems.

The first aspect deals with a limited number of evaluation approaches, most of them already covered within existing work. This includes aspects like documentation, logical satisfiability, or standard conformance. A collection, critical assessment and implementation must be an early step within the proposed work.

The second aspect is hard to grasp. Taken for granted that experienced ontology engineers develop some intuitive gut feeling about ontologies, their assessment of ontologies must correspond to the result of the evaluation provided by the evaluation framework. If this is the case, users of the evaluation framework will start to trust and appreciate the framework, and point other users to it, who trust the experts' opinion.

This leads to the third aspect: a wide adoption of the evaluation framework and its implementation by the Semantic Web community may, because of the above reasons, be regarded as an indicator for the quality of the framework.

The fourth aspect is the most promising one in the long run, but also out of scope for this thesis. In software engineering the similar problem of defining quality measures for code was tackled by looking for correspondences between code metrics and project success. This was possible due to the availability of data on many software projects. It is doubtful that during the course of this thesis enough data will be available to follow a similar approach, i.e. to regard a high number of projects, their metrics and their results, and then be able to identify the most interesting quality metrics.

5 Expected Impact

A viable and easy to use web ontology evaluation framework will help the Semantic Web become reality. Both, expert ontology engineers and non-expert users will benefit from a tool that is able to assess the quality of ontologies. It is crucial to implement evaluation measures that are important within the web context: a high quality of ontologies must support the overall development and evolution of the Semantic Web.

It is further envisaged that providing a set of agreed on evaluation measures will lead to an evolution in ontology engineering environments. The environments will encompass functionality to make the creation of high quality ontologies easier. Tools can automatically check if certain quality criteria are fulfilled during the engineering phase, and, if not, issue appropriate warnings.

If ontologies are a pillar of the emerging Semantic Web, they better be good in order to provide a stable foundation.

References

1. C. Brewster, H. Alani, S. Dasmahapatra, and Y. Wilks. Data-driven ontology evaluation. In *Proc. of the 4th International Conference on Language Resources and Evaluation*, Lisbon, 2004. European Language Resources Association.
2. A. Burton-Jones, V. Storey, V. Sugumaran, and P. Ahluwalia. A semiotic metrics suite for assessing the quality of ontologies. In *Data and Knowl. Engineering*, 2004.
3. M. Ehrig, P. Haase, N. Stojanovic, and M. Hefke. Similarity for ontologies - a comprehensive framework. In *13th European Conf. on Information Systems*, 2005.
4. M. Fox, M. Barbuceanu, M. Gruninger, and J. Lin. An organization ontology for enterprise modelling. *Simulating organizations: Computational models of institutions and groups*, 1998.
5. A. Gómez-Pérez. Ontology evaluation. In Staab and Studer [12], chapter 13.
6. Guarino and Welty. Overview of OntoClean. In Staab and Studer [12], chapter 8.
7. J. Hartmann, Y. Sure, A. Giboin, D. Maynard, M. del Carmen Surez-Figueroa, and R. Cuel. Methods for ontology evaluation, DEC 2004. KWeb Deliverable.
8. D. M. J. Brank, M. Grobelnik. A survey of ontology evaluation techniques. In *Proc. of the 8th Int. multi-conference Information Society IS-2005*, 2005.
9. A. Lozano-Tello and A. Gómez-Pérez. OntoMetric: A method to choose the appropriate ontology. *J of DB Mgmt. Spec. Issue on Ontological analysis, Evaluation, and Engineering of Business Systems Analysis Methods*, 15(2), April-June 2004.
10. B. Parsia, E. Sirin, and A. Kalyanpur. Debugging OWL ontologies. In *Proc. of the 14th World Wide Web Conference (WWW2005)*, Chiba, Japan, May 2005.
11. R. Porzel and R. Malaka. A task-based approach for ontology evaluation. In R. L. de Mántaras and L. Saitta, editors, *Proc. of the 16th European Conf. on Artificial Intelligence (ECAI 2004)*, Valencia, Spain, August 2004.
12. S. Staab and R. Studer, editors. *Handbook on Ontologies in Information Systems.* International Handbooks on Information Systems. Springer, 2004.
13. J. Völker, D. Vrandečić, and Y. Sure. Automatic evaluation of ontologies (AEON). In *Proc. of the 4th International Semantic Web Conference (ISWC'05)*, Nov 2005.