Semiotic-based Ontology Evaluation Tool
S-OntoEval

Renata Dividino, Massimo Romanelli and Daniel Sonntag

Fraunhofer Institute for Computer Graphics (IGD), German Research Center for AI (DFKI)
Fraunhoferstrasse 5 D-64283 Darmstadt, Stuhlsatzenhausweg 3 D-66123 Saarbrücken
renata.dividino@igd.fraunhofer.de, massimo.romanelli@dfki.de, daniel.sonntag@dfki.de

Abstract

The objective of the Semiotic-based Ontology Evaluation Tool (S-OntoEval) is to evaluate and propose improvements to a given ontological model. The evaluation aims at assessing the quality of the ontology by drawing upon semiotic theory (Stamper et al., 2000), taking several metrics into consideration for assessing the syntactic, semantic, and pragmatic aspects of ontology quality. We consider an ontology to be a semiotic object and we identify three main types of semiotic ontology evaluation levels: the structural level, assessing the ontology syntax and formal semantics; the functional level, assessing the ontology cognitive semantics and; the usability-related level, assessing the ontology pragmatics. The Ontology Evaluation Tool implements metrics for each semiotic ontology level: on the structural level by making use of reasoner such as the RACER System (Haarselv and Möller, 2001) and Pellet (Parsia and Sirin, 2004) to check the logical consistency of our ontological model (TBoxes and ABoxes) and graph-theory measures such as Depth; on the functional level by making use of a task-based evaluation approach which measures the quality of the ontology based on the adequacy of the ontological model for a specific task; and on the usability-profiling level by applying a quantitative analysis of the amount of annotation. Other metrics can be easily integrated and added to the respective evaluation level. In this work, the Ontology Evaluation Tool is used to test and evaluate the SWIntO Ontology of the SmartWeb project.

1. Introduction

Ontology evaluation is a basis for defining what a good ontology is. By analyzing the heterogeneous nature of works with respect to the assessment of the ontology, we could note three groups of evaluation methods: those assessing the graph structure and formal semantics of an ontology (Guarino and Welty, 2002; Yao et al., 2005; Huang and Diao, 2006); the second group assessing an ontology's intended use (Maedche and Staab, 2002; Porzel and Malaka, 2004; Daelemans and Reinberger, 2004; Lozano-Tello and Gomez-Perez, 2004); and the last group addressing to the quality level of the ontology's annotations (Noy, 2004).

These three assessment approaches are directly analogous to a semiotic assessment (Stamper et al., 2000).[1] Following Saussures, semiotics embraces the traditional branches of linguistics: syntactics (deals with the *structure* of signs and sign systems), semantics (deals with the *meanings* of signs and sign systems; that is, the meanings of words, sentences, gestures, paintings, mathematical symbols, etc.), and pragmatics (deals with *inferential meaning*, not merely logical inference, but the subtler aspects of communication expressed through indirection, e.g., "It's drafty in here" = "Close the door", and through social context).

In order to evaluate an ontology at its semiotic dimensions and we first identify the ontology's semiotic characteristics. Fundamentally, an ontology is a special kind of information object, structured in graph-like format. This infor-

mation object structure is used to represent the ontology intended conceptualization, i.e., its meaning or formal semantics. An ontology's intended meaning is established within a communication setting related to specific social-cultural definitions. Considering an ontology is a semiotic object (see figure 1), the quality of an ontology may be assessed with respect to its structure (syntax and formal semantics), its intended conceptualization (cognitive semantics), and its communication setting (pragmatics).
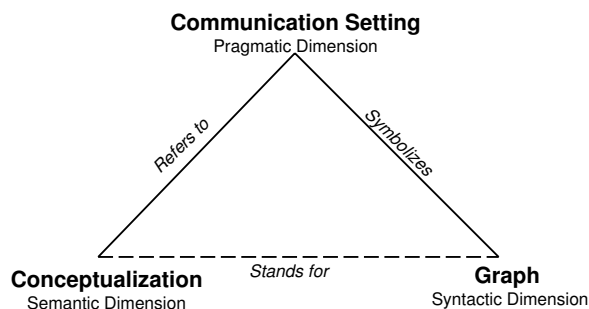

Figure 1: Ontology as a Semiotic Object

This paper presents a semiotic-based ontology evaluation tool (S-OntoEval) that makes use of an unique evaluation framework which allows to assess the quality of ontologies by drawing upon semiotic. The S-OntoEval tool enables the integration of the three different evaluation approaches into an unique evaluation suite. Basically, the semiotic-based evaluation framework is divided into three steps: firstly, the quality assessment of ontology syntax by making use of methods assessing the ontology's topological dimension and formal semantics (logical dimension); secondly, the semantic dimension by measuring the accuracy of the ontol-

---

[1] Semiotic, from the greek word *seemeiootikee*, denotes the study of signs, what they represent and signify, and how we act and think in their universe. Semiotic Theory was developed independently by the logician and philosopher Charler Sanders Peirce and the linguist Ferndinand de Saussure. Saussure's approach was a generalization of formal, structuralist linguistics; Peirce's was an extension of reasoning and logic in the natural sciences.

ogy with respect to its conceptualization (or intended use); and finally, its usability dimension by adopting approaches addressing the quality level of the set of annotations about the ontology and its elements.

The next section presents the implementation of the ontology evaluation tool according to the theoretical considerations. Subsequently, the evaluation results for the SmartWeb Integrated Ontology (SWIntO) (Oberle et al., 2007) use case is described.

2. Evaluation in Semiotic Levels

The Structural Evaluation employs structural measures focussed on the *syntax* of ontology graphs, thus assessing the ontology's topological dimension. Measures related to topological dimensions are *Depth, Breath, Modularity, Connectivity* (Gangemi et al., 2006; Huang and Diao, 2006). Measures related to logical adequacy (formal semantic) measures are, for instance, *Consistency, Complexity, Concept Satisfiability, Concept Subsumption* (Gangemi et al., 2006). The Functional Evaluation and corresponding measures assess the accuracy of the ontology with respect to its conceptualization. A good ontology should be a close approximation to the conceptualization that is supposed to be described. The more similar the ontologys intended meaning is to the conceptualization, the better the ontology is. The applicability of functional measurement is based on approximation measures, i.e., on a process of matching. Matching accuracy is evaluated by three measures: Precision, Recall, and F-measure (Baeza-Yates and Ribeiro-Neto, 1999). An appropriate evaluation strategy involve ways to capture the ontology conceptualization given expertise (Steels, 1990), task (Porzel and Malaka, 2004), or competency (Uschold and Grüninger, 1996). By applying this method, we are actually capturing domain functionality instead of capturing domain conceptualization itself. Thus, the functional dimension is related to the intended use of the ontology. The Pragmatic Evaluation is seen as an evaluation of a usability profile: Usability profiling measures focus on the ontology profile. An ontology profile is a set of annotations about the ontology and its elements. Measures of an ontology profile address the communication context (i.e., its *pragmatics*) of an ontology with the user.

From a semiotic point of view, an ontology profile should consist of set of structural, functional and user-oriented annotations containing all relevant information about the ontology from the point of view of an specific user (ontology engineer and ontology consumer) or group. Our S-OntoEval tool evaluates an ontology with respect to its three semiotic dimensions. The tool consists of the *structural module*, the *functional module*, and the *usability module* responsible for the assessment of the ontology in the correspondly dimension (see figure 2).[2]

S-OntoEval is a Java stand-alone application. It requires a configuration file at the start with settings such as the ontology file path, ontology repositories, reasoners, etc. The results of the evaluation is written in a report that can later

---

[2]Some other auxiliaries modules (*rating module*, *searching module*, and *publishing module*) enable the concurrent evaluation of more than only a single ontology and allow the use of multiple metrics at each semiotic evaluation level.
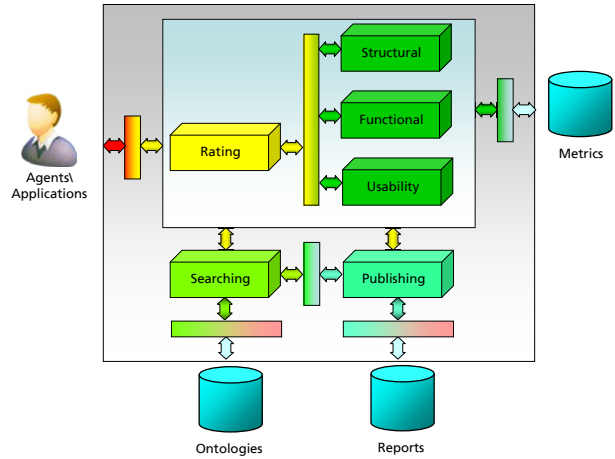

Figure 2: S-OntoEval Tool Architecture

serve as ontology-internal annotation or external HTML report. Basically, the GUI is composed of three tab windows, one for each semiotic level. Each tab presents the visual components of the corresponding semiotic level evaluation. In what follows, we provide a detailed description of the GUI elements and user interaction possibilities at each evaluation level.

2.1. Structural Evaluation

The structural evaluation tab is composed of five panels. As figure 3 shows, the left panel consists of a list of the structural metrics available for the evaluation organized in a tree structure. The tree structure presents an overview of the hierarchy among the available metrics. The tree tab also allows the selection of multiple metrics for a single evaluation step, i.e., metrics can be combined to generate a final evaluation score.
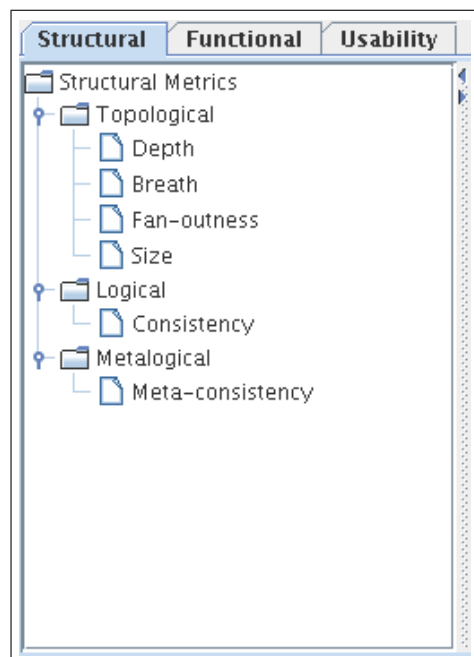

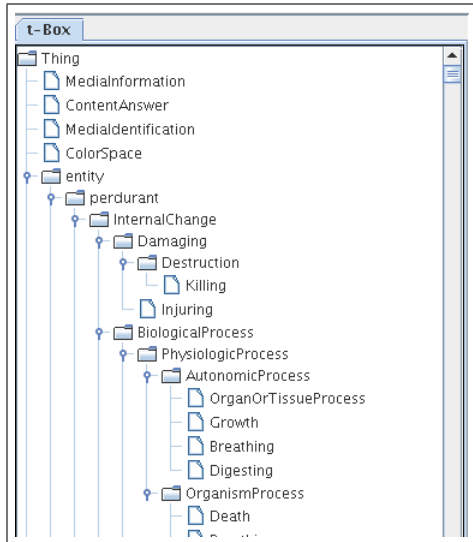Figure 3: Semiotic Tab and Structural Measures

Figure 4: Ontology Concept Hierarchy

The right panel of the structural tab presented in figure 4 shows the ontology to be evaluated in form of a tree structure (according to the ontology taxonomy). The tree structure facilitates the navigation through the ontology concept hierarchy, that is, the user is able to navigate through the concept tree by expanding and collapsing the tree nodes. Configuration settings for the evaluation at the structural level can be defined in addition. At this evaluation level the user should choose (from a list of reasoners plugged-in to the system) the reasoner to be used during the evaluation process. In order to start the process, the user either clicks on a metric in the tree structure of the right panel (with a right mouse click to choose the *evaluation* option at the pop-menu), or chooses one or more metrics in the table of the top panel. Then the evaluation can be started by pressing the evaluation button. The center panel of the structural tab (figure 5) shows the evaluation results.

2.2. Functional Evaluation

The functional evaluation tab of the S-OntoEval GUI is basically composed of four panels. Similarly to the structural tab, the left panel consists of functional metrics organized in a tree structure (see figure 6). Likewise, the right-bottom panel shown in figure 4 presents the ontology to be evaluated in form of the tree structure (ontology taxonomy), and the bottom panel presents the configuration settings defined by the user. Because the fact that in the functional tab the functional visual elements strongly depend on the functional metric chosen by the user during the evaluation process, different visual elements are displayed for each available functional metric. The S-OntoEval implements the *task-based approach* (or *performance approach*) which measures the performance of the ontology against a gold standard. For the *performance* metric, there are basically four elements to be set up in a configuration setting panel: a gold standard, set of instances, a corpus of question, and the reasoner.

Consequently, in the performance evaluation process, the center-top panel shows the corpus of question, i.e., the list of question to be evaluated by our tool to test its perfor-

mance, and finally, the right-top panel shows the concept hierarchy of the gold standard ontology as reference ontology. To start the evaluation process at the functional level, the user has to click on a metric. Figure 7 shows the functional tab of the S-OntoEval GUI.
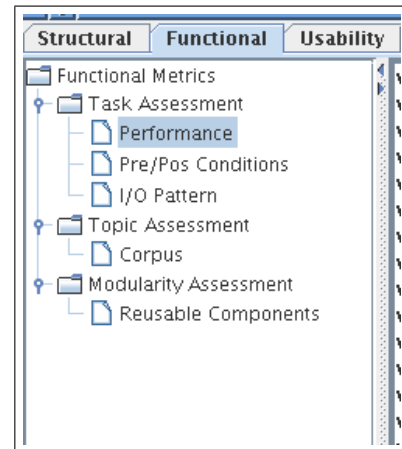

Figure 6: Functional Measures in Tree Structure

2.3. Usability Evaluation

In the S-OntoEval GUI, the usability evaluation is realized through the usability evaluation tab. From a semiotic point of view, usability metrics should consist of a set of structural, functional, and user-oriented metrics assessing the quality of the ontology annotation from a specific user's perspective (or the perspective of a group of ontology engineers or other ontology consumers). The three top panels in the usability tab show the metrics for structural, functional, and user-oriented evaluation at the usability assessing level in form of a tree structure. To start the usability evaluation, the user chooses the usability metric. Figure 8 shows the usability evaluation tab and its visual components. The bottom panel presents the evaluation results.

3. Evaluation of the SWIntO Use Case

In this section, we present the semiotic-based evaluation[3] of the SWIntO ontology (Oberle et al., 2007) developed in the SmartWeb project (Reithinger et al., 2006; Sonntag et al., 2007; Wahlster, 2007). SmartWeb aims to provide intuitive multimodal access to a rich selection of Web-based information services. The advanced ontology-based representation of facts and media structures (SWIntO) serves as central description for rich media content. We have made the choice of implementing simple evaluation metrics which on the one hand are relevant to the evaluation of the test object (e.g., the SWIntO ontology), and on the other hand we also believe to be representative in any ontology evaluation process. The main idea of conceptualizing the semiotic framework tool is not to implement at once all existent measures at each semiotic evaluation levels but to provide an extendable framework for attaching new more evaluation metrics (gradual enhancement). We describe the evaluation results of the SWIntO at each semiotic level.

---

[3]The tests were performed using an Intel Pentium-M processor 1.86 GHz and 1GB of main memory on Windows XP and Linux.

3.1. Structural Evaluation

We evaluated two metrics at the structural level: *maximum depth* and *consistency checking*. The *maximum depth* provides the number of nodes which lies on the longest path of the ontology's tree. The depth of a ontology's tree represents the length of a sequence of operations required to reach the terminal nodes. Figure 5 shows the SWIntO's evaluation with maximum path equal to 17, and all concepts that lie in that path. Since this context-independent information is not a significant measure of ontology quality, it can be combined with other measures to become more meaningful.

Consistency checking checks logical-adequacy of the ontology model. The consistency of an ontology model refers to its lack of contradiction, i.e., none of the facts deducible from the model contradict one another. Therefore, the ontology consistency can be considered as an agreement among ontology entities with the respect to the semantic of the underlying ontology language. An inconsistent ontology with respect to its theory may, for example, lead to erroneous conclusions in the reasoning process. Consistency checking is done with the help of two reasoners: RACER (Renamed ABox and Concept Expression Reasoner) (Haarselv and Möller, 2001) and Pellet (Parsia and Sirin, 2004). The evaluation output consists of a list of ontology errors and warnings which enable engineers to create an agenda for re-engineering the ontology. Similar to the *maximum depth* evaluation results shown in figure 5, consistency errors are presented in the GUI central panel and aware the user to inconsistency listing all names of inconsistent concepts, axioms and other ontology elements. For the SWIntO Use Case, the evaluation results in a "green" approval message to stress that the ontology is consistent with respect to its model.

Furthermore, the tool provides a plug-in framework to support various existing reasoners which ensures the independence of the metrics' results with respect to an individual reasoner.

3.2. Functional Evaluation

The functional evaluation follows a *task-based* approach which measures the quality of the ontology with respect to its performance, given a particular task. The task-based evaluation approach deals with measuring the quality of an ontology based on the adequacy of the ontological model of a specific task. The main idea is to observe the improvement or degradation of the performance of the task which performance depends on the ontology used. The task-based evaluation approach is a gold standard-based method (Maedche and Staab, 2002) which involves the creation of a validated corpus of answers for a certain task (a gold standard corpus). The gold standard is used as a reference for checking the performance of an ontology driven system. Adopting the *task-based* approach we choose the question answering task from the SmartWeb project which enables the user to pose closed and open domain multimodal questions, and delivers answers based on multiple Semantic Web information resources. Here, we restrict the task of question answering to a specific domain which makes use of ontological knowledge resources to answers

queries issued by the user.

Basically, the task-based evaluation process consists of comparing the time-performance of the question-answering task to a ideal time-performance. By plugging the gold standard ontology (Dividino, 2007) to the application, and running the task, we get (correct) answers which we regard as the gold standard answers with the ideal time-performance for each query: Figure 7 shows the results for the first query "*Which matches took place in the semifinals in 1954?*". The figure shows that the system takes $175ms$ to answer the question when the task is supported by the SWIntO, and only $97ms$ using the gold standard. Furthermore, all ontology axioms of the SWIntO and the gold standard which are divergent are presented in the result list. These axiom divergences point out that semantic reasoning is done thought different paths in the ontology, which is the basis of the task's performance improvement or degradation.

Considering the overall result, the results the SWIntO ontology are comparable to the SWIntO gold standard ontology with respect to the task performance. Although the axioms provided by the SWIntO do not follow all optimization patterns found in the gold standard ontology, they, nevertheless, obviously do not have a strong impact on the complexity of reasoning, thus on the degradation of its performance.

3.3. Usability Evaluation

S-OntoEval implements the *annotation analysis* metric which consists of quantify the total amount of metadata linked to the tag `rdf:comments` returning the percentage value indicating the result of the analysis. The main idea is to emphasize the importance of any kind of metadata in the ontology, addressing the pragmatic dimension of an ontology, which refers to the ontologys usefulness for users or their agents. The choice to implement the *annotation analysis* metric was based on the fact that unfortunately, none of the existent annotation tags provided by ontology languages follows the semiotic approach. An extension the OWL syntax annotation following the semiotic approach is proposed in (Dividino, 2007). Figure 8 shows the results of the annotation analysis evaluation with a lower usability score of 7%, the quantity of the annotated elements to 1.528 elements, the total of ontology elements to 21.597, and, for each annotated element, its respective annotation linked to the tag `rdf:comments`.

4. Conclusion

We looked at existing ontology evaluation methods and implemented a semiotic based evaluation tool (S-OntoEval) from the perspective of their integration into one single framework. The tools realizes a complete evaluation combining several evaluation metrics categorized into semiotic levels. The main goal was to provide an ontology evaluation tool which support users to assess their ontology quality considering different evaluation methods and theories but based in a semiotic evaluation framework. By adopting the semiotic approach, the user is able to evaluate how well the ontology's vocabulary, taxonomy, and semantic relationships between concepts are. This is (1) achieved at the

structural level, (2) modeled in relation to its given intended task (its cognitive semantics); and (3) addressed the communication context of an ontology, its pragmatics. In the evaluation example, we choose four metrics among others (*depth*, *consistency checking*, *task-based*, and *annotation analysis*) which we believe to be representative in any ontology evaluation process. The main idea is to provide an extendable framework for attaching new evaluation metrics as gradual enhancement.

Although a lot of the research has been successfully performed in ontology evaluation field, there are still open issues that should be resolved such as improvements on the pragmatic dimension evaluation. The S-OntoEval tool allows to combine different evaluation scores which we plan to extend to more user-specific personalized combinations. After using S-OntoEval on the SWIntO ontology, an optimized version could be deployed for usage in the project THESEUS.[4]

Acknowledgment

5.   References

R. Baeza-Yates and B. Ribeiro-Neto. 1999. Modern information retrieval. page 75, NY, USA. ACM Press, Addison-Wesley.

W. Daelemans and M. Reinberger. 2004. Shallow text understanding for ontology content evaluation. *IEEE Intelligent Systems*, 19(4):1541–1672.

R. Dividino. 2007. Semiotic-based ontology evaluation tool. Master's thesis, Universität des Saarlandes.

A. Gangemi, C. Catenacci, M. Ciaramita, and J. Lehmann. 2006. Qood grid: A metaontology-based framework for ontology evaluation and selection. In *Proceedings of the 4th International Workshop on Evaluation of Ontologies for the Web (EON2006)*, Edinburgh, United Kingdom.

N. Guarino and C. Welty. 2002. Evaluating ontological decisions with Ontoclean. *Commun. ACM*, 45(2):61–65.

V. Haarselv and R. Möller. 2001. Racer system description. In *Proc. of the International Joint Conference on Automatic Reasoning (IJCAR 2001)*, volume 2083, pages 701–706.

N. Huang and S. Diao. 2006. Structure-based ontology evaluation. In *ICEBE '06: Proceedings of the IEEE International Conference on e-Business Engineering*, pages 132–137, Washington, DC, USA.

A. Lozano-Tello and A. Gomez-Perez. 2004. Ontometric: A method to choose the appropriate ontology. *Journal of Database Management*, 15:1–18.

A. Maedche and S. Staab, 2002. *Measuring similarity between ontologies*. Springer, Madrid, Spain.

N. Noy. 2004. Evaluation by ontology consumers. *IEEE Intelligent Systems*, 19(4):74–81.

D. Oberle, A. Ankolekar, P. Hitzler, P. Cimiano, C. Schmidt, M. Weiten, B. Loos, R. Porzel, H. Zorn, V. Micelli, M. Sintek, M. Kiesel, B. Mougouie, S. Vembu, S. Baumann, M. Romanelli, P. Buitelaar, R. Engel, D. Sonntag, N. Reithinger, F. Burkhardt, and J. Zhou. 2007. DOLCE ergo SUMO: On Foundational and Domain Models in the Smartweb Integrated Ontology (SWIntO). *Web Semant.*, 5(3):156–174.

B. Parsia and E. Sirin. 2004. Pellet: An OWL DL Reasoner. In *3rd International Semantic Web Conference (ISWC2004)*, Hiroshima, Japan, November.

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

N. Reithinger, S. Bergweiler, A. Blocher, R. Engel, G. Herzog, A. Pfalzgraf, N. Pfleger, M. Romanelli, G. Sonnenberg, D. Sonntag, and W. Wahlster. 2006. SmartWeb: Multimodal Interaction with Web Services. In *Proceedings of the 29th annual German Conference on Artificial Intelligence (KI2006)*, Bremen, Germany, June.

D. Sonntag, R. Engel, G. Herzog, A. Pfalzgraf, N. Pfleger, M. Romanelli, and N. Reithinger. 2007. SmartWeb Handheld - Multimodal Interaction with Ontological Knowledge Bases and Semantic Web Services. In *Artifical Intelligence for Human Computing*, pages 272–295.

R. Stamper, M. H. K. Liu, and Y. Ades. 2000. Understanding the roles of signs and norms in organizations: a semiotic approach to information systems design. In *Behaviour and Information Technology - BIT*, volume 19, pages 15–27.

L. Steels. 1990. Components of expertise. *AI Mag.*, 11(2):30–49.

M. Uschold and M. Grüninger. 1996. Ontologies: Principles, methods, and applications. *Knowledge Engineering Review*.

Wolfgang Wahlster, 2007. *40 Jahre Informatikforschung in Deutschland*, chapter SmartWeb: Ein multimodales Dialogsystem fr das semantische Web. Springer, Heidelberg, Berlin.

H. Yao, A. M. Orme, and L. Etzkorn. 2005. Cohesion metrics for ontology design and application. *Journal of Computer Science*.

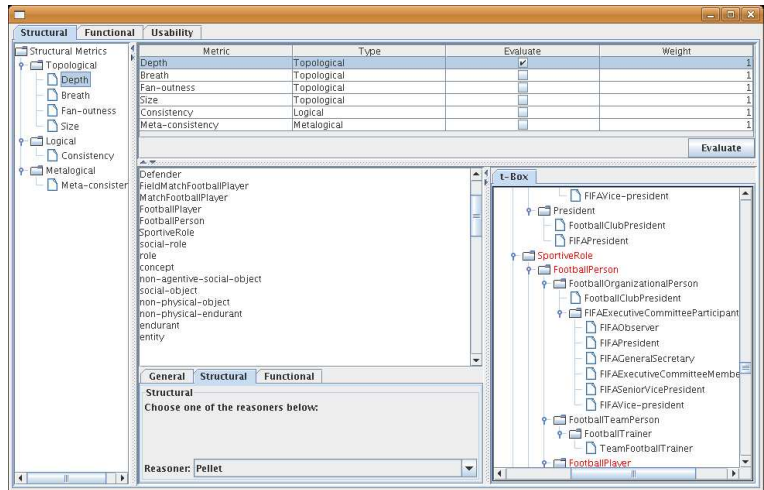[4]See *http://theseus-programm.de/front* for more information.

Figure 5: Structural Tab of the S-OntoEval GUI and the *Maximum Depth* Evaluation Results
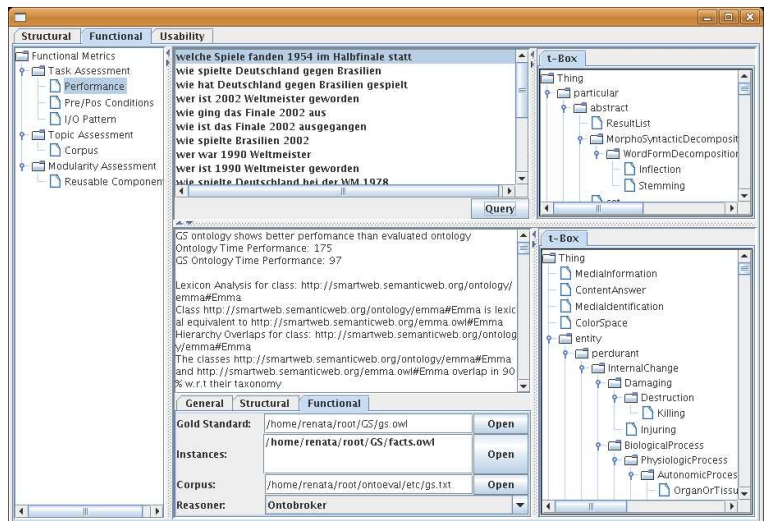

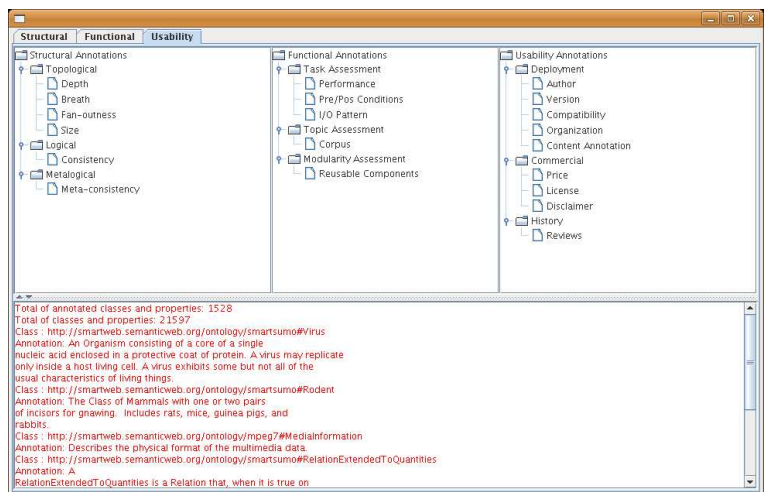Figure 7: Functional Tab and *Task-based* Evaluation Results

Figure 8: Usability Evaluation Tab and *Annotation Analysis* Evaluation Results