Engineering Applications of Artificial Intelligence 26 (2013) 1413–1420

Contents lists available at SciVerse ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

A keyword-driven approach for generating OWL DL conformance test data

Raúl García-Castro*, Asunción Gómez-Pérez

Ontology Engineering Group, Facultad de Informática, Universidad Politécnica de Madrid, Spain

ARTICLE INFO

Article history:
Received 10 May 2012
Received in revised form
3 July 2012
Accepted 17 July 2012
Available online 9 August 2012

Keywords:
Semantic technology
Conformance
Keyword-driven test data generation
OWL DL

ABSTRACT

The conformance of semantic technologies has to be systematically evaluated to measure and verify the real adherence of these technologies to the Semantic Web standards. Current evaluations of semantic technology conformance are not exhaustive enough and do not directly cover user requirements and use scenarios, which raises the need for a simple, extensible and parameterizable method to generate test data for such evaluations. To address this need, this paper presents a keyword-driven approach for generating ontology language conformance test data that can be used to evaluate semantic technologies, details the definition of a test suite for evaluating OWL DL conformance using this approach, and describes the use and extension of this test suite during the evaluation of some tools.

1. Introduction

The W3C Semantic Web Activity has produced different standards[1] that enable technology interoperability in the open environment of the (Semantic) Web. However, a systematic evaluation of the conformance of semantic technologies is required to measure and verify their real adherence to these standards.

Conformance is a primary requirement for semantic technologies and its evaluation essentially covers two scenarios in terms of: (a) *tool validation*, which is mainly relevant to tool developers and involves checking whether the tool correctly meets the specifications and (b) *feature analysis*, which is mainly relevant to tool users and involves checking which parts of the specification the tool covers, either the whole specification or a subset of it.

Current evaluations of semantic technology conformance are not exhaustive enough, both in terms of technology coverage and of standard coverage. However, while other characteristics of semantic technologies (e.g., efficiency or usability) are non-critical in most use scenarios (i.e., users can come to terms with a variation in tool quality), users expect full conformance to the specifications included in standards or to the subset required by them.

Clearly, full conformance evaluation is impossible, since it is not possible to define every possible variation of the requirements included in a certain specification (e.g., to define every possible OWL ontology or SPARQL query), and we need to produce effective conformance evaluation methods and evaluation data. A similar issue is largely covered in the area of software testing where it is acknowledged that, besides expertise in the functionality to be tested and in the domain of the data to be used, effective testing requires understanding the different use scenarios of the software (Burnstein, 2003), in our case, the different semantic technology use scenarios.

The ideal approach would be to cover these use scenarios when evaluating semantic technology conformance, but currently semantic technology users are passive actors in conformance evaluations. First, current conformance evaluations are generic and do not directly cover user requirements and use scenarios and, second, it is difficult for users to evaluate technology conformance on their own, since this is a resource-consuming task and they do not have enough expertise in semantic technologies and their specifications.

This raises the need for a method to define conformance test data that is simple, to facilitate users the definition of test data suited to their use scenarios and the understanding of existing evaluations, and extensible and parameterizable, to allow defining test data as exhaustive as needed.

Previous work on the evaluation of semantic technology conformance with regards to the ontology language model has covered the definition of test data to be used in conformance evaluations and of methods for the execution of these evaluations, both manually and automatically. Once we have a way of automatically executing conformance evaluations, we can afford to increase the exhaustiveness of these evaluations and to involve users in them, that is, to generate larger quantities of test data and to allow users to define these data according to their needs.

This paper presents a keyword-driven approach for generating ontology language conformance test data that can be used to evaluate semantic technologies. The paper only covers the generation

* Corresponding author.
E-mail addresses: rgarcia@fi.upm.es (R. García-Castro), asun@fi.upm.es (A. Gómez-Pérez).

[1] http://www.w3.org/2001/sw/Specs.html.

0952-1976/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2012.07.009

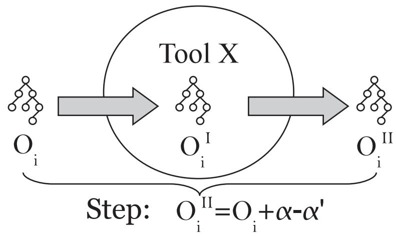

Fig. 1. Steps of a conformance test execution.

of test data and not the use of these data in evaluations, since this is already covered in previous work. Nevertheless, we also describe how we have defined, using this approach, a test suite for evaluating the OWL DL conformance of semantic technologies and present how we have used and extended this test suite during the evaluation of some tools.

This paper is structured as follows. Section 2 presents our understanding of conformance and previous work related to this topic. Section 3 gives an overview of the keyword-driven test suite generation process and, then, Sections 4–6 provide some insights into the definition of the Keyword Library used in this process, the structure of test suite definition scripts, and the implementation of the test suite generator, respectively. Section 7 shows how the OWL DL Import Test Suite was defined following the abovementioned process and Section 8 explains how we have used this test suite to evaluate some tools. Finally, Section 9 includes some discussion about the work presented and Section 10 presents some conclusions and future lines of work.

2. Related work

The conformance characteristic for semantic technologies is related to the ability of these technologies to adhere to existing specifications; in this case, the most relevant specifications are those of the existing ontology representation languages (i.e., RDF(S), OWL and OWL 2).

With respect to an ontology language specification there are several aspects of semantic technology conformance, since such conformance can be evaluated in terms of:

- *The ontology language model*. Since different tools have different internal knowledge representation models, it is important to know the similarities and differences between these internal models and the knowledge representation model of the ontology language.
- *The ontology language serialization*. Existing ontology languages have different serializations, both normative and non-normative (e.g., N3, RDF/XML, OWL/XML, and Turtle). A tool that supports an ontology language should also support at least one of such serializations, including their syntactic variants.
- *The ontology language semantics*. Ontology language specifications include one or more formal semantics that can be used with the ontology language. A tool implementing one of these formal semantics should be consistent with it.

The work presented in this paper only covers conformance regarding the ontology language model and does not cover other types of conformance or other characteristics of semantic technologies (e.g., robustness and scalability).

Up to now, semantic technology conformance evaluations have been performed in terms of tool validation, being the two main efforts to this end those of the W3C ontology language specifications and of the RDF(S) and OWL Interoperability Benchmarking activities.

The *W3C ontology language specifications* include definitions of test cases for RDF(S) (Grant and Beckett, 2004), OWL (Carroll and Roo, 2004) and OWL 2 (Smith et al., 2009), which illustrate the correct usage of the ontology languages and the resolution of issues considered by the Working Groups. These test cases mainly cover conformance with regards to the ontology language semantics but also cover ontology language model and serialization conformance, both with correct and incorrect ontologies. Besides, the test cases are described in terms of ontologies to support the automation of their execution; however, software support is only provided to execute the OWL 2 test cases.

The *RDF(S) and OWL Interoperability Benchmarking activities* (García-Castro and Gómez-Pérez, 2009, 2010) involved the evaluation of the interoperability of semantic technologies using an interchange language and included a conformance evaluation with the goal of evaluating the conformance of semantic technologies with regards to an ontology language model.

During a conformance evaluation, described in detail in García-Castro and Gómez-Pérez (2010), a common group of tests is executed and each test describes one input ontology that has to be imported by the tool and then exported.

Each test execution comprises two steps, shown in Fig. 1. Starting with a file containing an ontology (O_i), the execution consists in importing the file with the ontology into the origin tool and then exporting the ontology to another file (O_i^{II}).

In these steps there is not a common way of checking how good the importers (by comparing O_i with O_i^I) and exporters (by comparing O_i^I with O_i^{II}) are. We just have the results of combining the import and export operation (the file exported by the tools), so these two operations are viewed as an atomic operation. It must be noted, therefore, that if a problem arises in one of these steps, we cannot know whether it was originated when the ontology was being imported or exported because we do not know the state of the ontology inside each tool.

After a test execution, we have two ontologies in the ontology representation language, namely, the original ontology (O_i) and the final ontology exported by the tool (O_i^{II}). By comparing these ontologies we can know up to what extent the tool conforms to the ontology language using the following metrics:

- *Execution* (*OK*/*FAIL*/*P.E.*) informs of the correct test execution. Its value is *OK* if the test is carried out with no execution problem; *FAIL* if the test is carried out with some execution problem; and *P.E.* (Platform Error) if the evaluation infrastructure launches an exception when executing the test.
- *Information added or lost* shows the information added to or lost from the ontology. We can know this information by comparing the original ontology with the final one; this comparison is performed both at the structural level and at the semantic level.
- *Conformance* (*SAME*/*DIFFERENT*/*NO*) explains whether the ontology has been processed correctly with no addition or loss of information. From the previous basic metrics, we can define *Conformance* as a derived metric that is *SAME* if *Execution* is *OK* and *Information added* and *Information lost* are void; *DIFFERENT* if *Execution* is *OK* but *Information added* or *Information lost* are not void; and *NO* if *Execution* is *FAIL* or *P.E.*

Two test suites were used to evaluate conformance, including only correct ontologies and covering the RDF(S) and OWL Lite languages. As in the case of the W3C test cases, the test suites were described using ontologies and the IBSE[2] tool was provided to automatically evaluate tools.

[2] http://knowledgeweb.semanticweb.org/benchmarking_interoperability/ibse/.