A Resemblance Study of Test Collections for World-altering Semantic Web Services

Hadi Saboohi, Sameem Abdul Kareem

Abstract—Approaches to Web services discovery, matchmaking, composition, and execution monitoring need a collection of descriptions of test services. Semantic Web services help attain better accuracy in these approaches by clarifying service descriptions. In the semantic Web service research area, there are some test collections. These test collections have some information-providing services. However, there is not a proper test collection that contains a reasonable number of both information-providing and world-altering services. This paper intends to investigate current available test collections, as well as other sources of service description. Their features (such as their description and expression languages) and a number of world-altering services, preconditions, and effects are compared, and their usage in current contests and challenges are discussed. Finally, a road map for standard test collections for world-altering semantic Web services is provided.

Index Terms—Semantic Web Service, Test Collection, World-altering Services, Precondition, Effect

I. INTRODUCTION

LIKE other scientific fields, specifically computer science, any evaluation needs test data. Test data are data that have been explicitly generated or collected to be used in evaluation exercises. This is typical for any computer program. Web services as remote applications in a service-oriented architecture (SOA) are not exempt from this issue.

The test data for Web services are along two axes. First, the set of inputs to be checked by Web services (as remote macro functions) if they are capable of generating the desired outputs. Second, is the description of services to be used in the discovery and the matchmaking of intended service with the composition of atomic services to fulfill a user's goal. These descriptions can be used in execution monitoring and failure recovery of atomic and composite services as well. This paper discusses the second type of test data for Web services.

"Semantics" as an add-on for Web service description can be used to disambiguate the various definitions of functional descriptions of services. These functional descriptions of semantic Web services include input, output, precondition, and effect.

Inputs and outputs describe information transformation done by the service. Preconditions and effects describe the state change of the world. Precondition is the state of the knowledge-world or real-world before the execution of the described service. Effect is the description of the world change made by the service invocation.

H. Saboohi is a PhD student in the Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya (UM), 50603 Kuala Lumpur, Malaysia.
e-mail: saboohi@siswa.um.edu.my

Dr. S. Abdul Kareem is an Associate Professor in the Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya (UM), 50603 Kuala Lumpur, Malaysia.
e-mail: sameem@um.edu.my

Web services are categorized as *world-altering* services and *information-providing* services [1]. This categorization is determined by the presence or lack of a kind of effect by the execution of the service. Any test data for semantic Web services surely needs to contain both information-providing services and world-altering services.

Gathering a reasonable number of services generates a collection of test data, which is called a "*Test Collection.*" As it is common, these test collections may be produced either manually, focusing on a relevant domain, or in an automatic manner, in which a large-volume of randomized data is generated programmatically. In the semantic Web service research field, both approaches have been used.

For a practical Web service testing approach to be applicable in the real-world, it needs realistic test data. This testing approach is for *mediation* techniques, including discovery, matchmaking, composition, and execution monitoring of Web services. Therefore, there is a critical need for actual standard test collections of semantic Web services. These standard test collections, like in any software development environment, are intended to be utilized to test mediation approaches to show that it has the claimed ability.

Another critical aspect in a test collection of semantic Web services is the ontology of the concepts used in the description of semantic Web services. Functional specification of semantic Web services, particularly input and output of a service, are described using taxonomies in an ontology. This is in contrast with the description of input and output of traditional Web services. Traditional Web services are commonly described using the Web Service Description Language (WSDL)[1]. The input and output of traditional Web services are described by the XML Schema[2], which does not convey any meaning. Therefore, bundling a required ontology (or some ontologies) is an important requirement for a real, applicable test collection.

In recent years, there have been a few efforts in making test collections for semantic Web services. These collections are used in testing mediation approaches of semantic services. Some of these test collections are specifically created for various contests and challenges in the field of semantic services (cf. to Section V).

A test collection of semantic Web services definitely requires a combination number of information-providing services and world-altering services. Currently, test collections lack the fair number of each of these categories. This paper intends to analyze existing test collections in terms of their diverse features, and to provide a road map for standard test collections of world-altering semantic Web services.

[1] http://www.w3.org/TR/wsdl20
[2] http://www.w3.org/TR/xmlschema-2/

This paper is structured as follows. In Section II, first, we describe publicly available test collections, and second, elaborate on a test collection used in a specific research. Section III discusses an online portal for semantic Web services. Next, Section IV identifies Web crawlers that cumulate service descriptions, including our crawler. Finally, Section V distinguishes contests and challenges that use these collections. Related work that is presented in Section VII and Section VIII summarizes the paper.

II. AVAILABLE TEST COLLECTIONS

There are three existing major test collections, namely: SWS-TC, OWLS-TC, SAWSDL-TC. Additionally, there is another unpublished test collection, which we call "100 Services" throughout this paper.

There exist some other collections, which are mentioned in [2], but are not included in our paper for reasons of unavailability, having non-standard formalisms, or having formalisms other than the major ones available now.

The four investigated test collections are differentiated and compared in the following.

A. SWS-TC

SWS-TC[3], generated manually by Ganjisaffar and Saboohi in 2006, contains 241 semantic services, mostly real Web services. Services are described using OWL-S [3] description language. OWL-S version 1.1[4] is used in this test collection.

SWS-TC contains 30 world-altering services. These service descriptions with preconditions and effects are written in SWRL [4]. Nevertheless, there is no expression available in their precondition and effect specifications. Precondition names, effect names, and informal descriptions in natural language are available, but formal SWRL specifications are "NIL."

There are two strong points for services available in SWS-TC. First, all the services are described with a single, unified ontology derived from WordNet[5]. Therefore, the ontology distance of concepts can be calculated for a matchmaking algorithm. This is a unique feature in this test collection; others use multiple ontologies to describe the concepts. This is not a weak point for the others because, in the real world, it is somehow impossible to have a unified ontology for all the concepts. Second, most of the services in this collection have been made from real Web services (they have been found in Web service search engines), and they are not the so-called "toy" services.

The limitations of SWS-TC are the following. First, the number of services is unsuitable for a test scenario. Second, the test collection is no longer being maintained and updated.

B. OWLS-TC

OWLS-TC[6] is an OWL-S service retrieval test collection. Services are described in OWL-S version 1.1 and some in OWL-S version 1.0[7]. This test collection contains service advertisements, sample requests, and relevance judgments of the advertisements regarding to the requests. Klusch, Kapahnke et al. developed it in 2005 to support the evaluation of the performance of OWL-S semantic Web service matchmaking algorithms.

The developers are actively improving OWLS-TC. There was no service, including precondition and effect, in the first three versions of this test collection. Some informal conditions written in services' comments could not convey practical world-altering service definitions.

Recently, in September 2010, they have released the fourth version of this collection. The latest release has service condition and effect specifications available both in SWRL and PDDL [5], which are bundled separately.

This test collection contains 158 service descriptions providing precondition and/or effect specifications. Among all, 54 services are world-altering services. Furthermore, in this test collection, there are five queries requiring world-altering services and 17 queries for services in need of preconditions. Comparing the number of world-altering service descriptions with the total number of 1083 semantic services (5%) is aspirational, but not sufficient.

C. SAWSDL-TC

A test collection similar to OWLS-TC is called SAWSDL-TC[8]. Klusch and Kapahnke also developed this test collection. It was first converted from OWLS-TC. They made it to support the evaluation of the performance of SAWSDL [6] semantic Web service matchmaking algorithms.

There is no condition or effect described in the formalism of this test collection. Therefore, the number of world-altering services in this test collection is unknown.

D. 100 Services

In [7], researchers created their own test collection. This test collection seems to be the first in semantic Web services test collections, which contains the full description of *world-altering* services. Prior to that, creators of SWS-TC described some world-altering services in their test collection. However, as discussed in Section II-A, these specifications of preconditions and effects are not complete.

This test collection is not publicly available. We could analyze it because it was graciously provided for us by the creators. This allowed us to include it in our investigation.

Services are described in OWL-S 1.1. They have used multiple ontologies for the concepts of inputs and outputs for Web services.

Among the 100 services available in this test collection, 39 services are world-altering services. There are two unique preconditions and two unique results (effects). These preconditions and effects are in SWRL and replicated in various combinations for the services with assorted numbers and combinations of inputs and outputs. This permutation generated 39 distinctive world-altering services. Consequently, there are 61 information-providing services in this test collection.

III. A Portal for Semantic Services - OPOSSum

OPOSSum[9] is an online portal for semantic services [2]. It assembles data from SWS-TC, OWLS-TC and some other sources to create an assemblage of semantic Web services with different description languages. It presently contains over 2800 descriptions for more than 1500 services.

Unlike the test collections mentioned, OPOSSum is not file-centered, and it is based on a relational database. Its web interface enables the users to add, update, and improve the data.

Unfortunately, despite the OPOSSum developers' hope that the services and their descriptions would be improved by semantic Web service community, the number of the services in this portal did not go further than 1600 in early 2011, which was almost equal to those in 2008.

IV. Web Crawlers to Cumulate

An effort called Seekda's Web Services portal[10] provides a Web service search engine. It helps users find Web Services based on a catalogue of more than 28,000 service descriptions. It utilizes a focused crawler that collects respective information about services available on the Web. The system monitors these services and allows users to edit certain data regarding providers or services. Unfortunately, this portal does not support semantic services. As we have investigated, there is no facility in their search system to find world-altering services.

There are some attempts to gather semantic service descriptions. These efforts try to use a Web crawler to look for public Web sites that publish semantic service descriptions in one or more description languages.

A. SouSuo

The meta-search engine SouSuo[11] 1.6 searches for semantic Web services in semantic service description languages like OWL-S, WSDL-S [8], WSMO [9], and SAWSDL, then evaluates the results in terms of their type, location, domain, and category [10].

B. Our Crawler

We have also configured a crawler (written in Java) to find service descriptions publicly available on the web. The objective was similar to the one in [10] with the distinguishing feature that we were specifically looking for world-altering services. Likewise, the total number of semantic services we could find was not satisfactory. Eventually, the number of world-altering services, or even services with some kind of conditions, was small. Unfortunately, services containing SWRL definitions were few and far between.

V. Contests and Challenges

There are four major contests and challenges for various research fields of semantic services. These include SWS-Challenge, S3 Contest, IEEE WSC, and SEALS SWS Discovery Evaluation.

A. SWSC

The SWS (Semantic Web Services) Challenge[12] intends to design and develop a standard methodology and testbed for the evaluation of Semantic Web Services technologies [11].

Currently, they focus on Web services described in WSDL 2.0. The SWS Challenge is interested in comparing the effectiveness of various formalisms for different problems. It has two tracks: data and process mediation, and service discovery.

B. S3

The S3 (Semantic Service Selection) Contest[13] is a competition that measures the speed and precision of retrieval algorithms in performing discovery tasks. Moreover, the contest requires commitment to a specific semantic formalism. The S3 contest has different kinds of problem sets.

The latest edition of the contest, which was in 2010, had two tracks for OWL-S and SAWSDL matchmaker evaluations. The contest was conducted to compare the retrieval performance for services in the OWL-S and the SAWSDL formats over test collections OWLS-TC and SAWSDL-TC respectively.

This contest uses SME^2[14] for evaluation purposes. SME^2 evaluates matchmakers for Semantic Web services over given test collections in terms of standard retrieval performance evaluation measures.

C. WSC

The IEEE WSC (Web Services Challenge)[15] encourages both industry and academic researchers to participate. These include the groups that develop software components or intelligent agents. These applications should have the ability to discover relevant web services and also generate composite services.

The sixth competition, which was held in 2010, focused exclusively on semantic composition of web service chains, whereas in the early editions, it was a syntactic-based contest. Rather than XML Schema, it incorporates the use of OWL ontologies to define services and their relationships to each other. The participants were required to determine relations between different types during the process of service composition.

The IEEE WSC has its own test set generator. This tool generates an arbitrary number of services using any number of concepts that the user likes. These concepts are also randomly generated and saved in an OWL taxonomy file.

D. SEALS - SWS

The semantic Web services (SWS) track of the SEALS Campaign[16] is focused on the evaluation of algorithms and tools for semantic service discovery and matchmaking.

[12]Semantic Web Service Challenge: Evaluating Semantic Web Services Mediation, Choreography and Discovery, available at http://www.sws-challenge.org/

[13]Annual International Contest S3 on **Semantic Service Selection** - Retrieval Performance Evaluation of Matchmakers for Semantic Web Services, http://www-ags.dfki.uni-sb.de/~klusch/s3/

[14]The Semantic Web Service Matchmaker Evaluation Environment (SME^2), available at http://www.semwebcentral.org/projects/sme2/

[15]http://www.wschallenge.org/

[16]Semantic Web Service Tools Evaluation Campaign, http://www.seals-project.eu/seals-evaluation-campaigns/semantic-web-services

[9]http://fusion.cs.uni-jena.de/OPOSSum/

[10]http://webservices.seekda.com/

[11]http://www.semwebcentral.org/projects/sousuo/

This evaluation tests the retrieval performance of tools submitted by contestants using SEALS platform. Additionally, these matchmakers will be compared based on the datasets and language by the defined workflows, i.e. the evaluation description.

E. Comparison of the Contests

These contests and challenges are compared in terms of four criteria: the problem sets, the formalisms they allow to be used, test collections they use, and the relevance of the use of world-altering services.

First, regarding the problem sets, all the S3, the WSC, and the SWS track of SEALS are contests that have some specialized problem sets. They are limited to semantic service discovery, even though the WSC and the SWSC also includes semantic service composition. The general mediation is another goal of the SWSC. Furthermore, the SWSC and the S3 seek to make a common testbed for semantic services.

Second, the S3, the WSC, and the SWS track of SEALS have selected semantic formalisms, i.e., OWL-S. However, the SWSC is not biased towards a specific formalism. Moreover, SWSC allows participants to add any semantic annotation to solve the problems and evaluates the proposed formalisms (which are derived from the natural language descriptions).

Third, the S3 and the SWS track of SEALS use both OWLS-TC and SAWSDL-TC. The WSC has its own test set generator to make a randomized test set. The SWSC has no test collection, and it encourages the participants to use any formalism that they believe appropriate for their solution to annotate services.

Fourth, the contests that use OWLS-TC, either implicitly or explicitly, consider world-altering services in their scenarios. These include the S3, and the SWS track of SEALS. Furthermore, the SWSC, which is the only challenge with mediation of semantic services as one of its goals, uses world-altering services in its scenarios. These world-altering services are such as Hardware Purchasing, and Shipment of Products.

VI. Discussion

Figure 1 illustrates the fraction of the *world-altering* services compared to the *information-providing* services in the major test collections. SAWSDL-TC is not shown in this figure because the world-altering services in SAWSDL language have no distinctive feature to be considered[17].

The only test collection with more than a third of its total services consisting of the world-altering ones is the "100 Services". Nevertheless, the total number of services in this test collection is not sufficient for a standard test collection to be applicable for a practical test.

Table I summarizes some common features of these test collections. The number of total services available in these collections is shown. Moreover, without considering inputs and outputs of services, the number of services with precondition specification, result (effect) specification, and services with both precondition and result (effect) are separately identified. Other characteristics, including last release year, last

---

[17] http://www.w3.org/2002/ws/sawsdl/spec/examples/#conditions

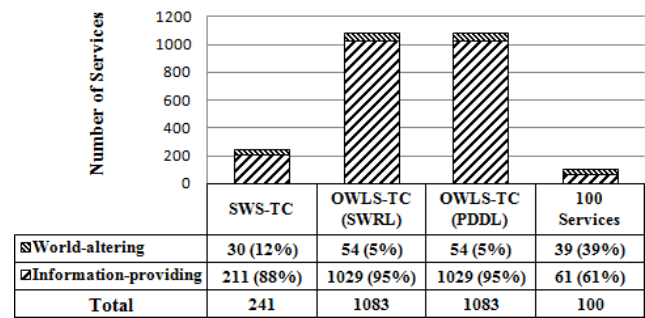|	SWS-TC	OWLS-TC (SWRL)	OWLS-TC (PDDL)	100 Services
World-altering	30 (12%)	54 (5%)	54 (5%)	39 (39%)
Information-providing	211 (88%)	1029 (95%)	1029 (95%)	61 (61%)
Total	241	1083	1083	100

Fig. 1. Fractions of the *world-altering* services in compare to the *information-providing* services

version, quantity of used ontologies, description language, and expression language are keyed in as well. In each row, the best value is emphasized if it is relevant.

As we have investigated, in current search facilities of OPOSSum, there is no specific way of finding either world-altering services or services with a particular precondition or effect. Searching the whole description of services, there were no precondition and effect definitions except for the ones from SWS-TC, which was discussed earlier. Another problem of this portal is that it is not updated, and current releases of the test collections are not imported into this portal.

There is one crucial problem in the actual using and testing of world-altering services available in these test collections. As the nature of world-altering services reveals, these services make a change in the knowledge- or real-world. Hence, any execution of these services has some kind of effects in the world, which sometimes even need payment to the provider. Some of these effects can be compensated and some not. Therefore, evaluations are usually done based on a simulation of execution, which is not naturalistic.

For a test collection to be standard and realistic, it needs to contain a reasonable number of service descriptions in different formalisms. Hence, there is a need for a comprehensive test collection with the following features:

- There is a strong need for OWL-S described services with SPARQL [12] RDF query languages as one of the expression languages for conditions and results (effects) in OWL-S. SPARQL is now a W3C Recommendation[18] for a query language for RDF.
- There is currently no test collection available for WSMO/WSML.
- WSMO-Lite[19], which is a restricted subset of WSMO, and the latest W3C submission for a semantic service description language, has no related test collection as well.

VII. Related Work

Researchers in [2] elaborate the needed features of test collections applicable for an evaluation of semantic Web service approaches. They listed desirable characteristics of a test collection as a large number of services, contribution by different groups of people, services from diverse domains, and both informal (natural language) descriptions and formal

---

[18] http://www.w3.org/TR/rdf-sparql-query/
[19] http://www.w3.org/Submission/WSMO-Lite/

TABLE I
SUMMARIZATIONS OF THE INVESTIGATED TEST COLLECTIONS

Test Collection Name		SWS-TC	OWLS-TC (*SWRL*)	OWLS-TC (*PDDL*)	SAWSDL-TC	100 Services
Total number of services		241	*1083*	*1083*	1080	100
Number of services with	Precondition	10	158	158	Unknown	41
	Result (Effect)	30	54	54	Unknown	39
	Both { Precondition Result (Effect)	7	46	46	Unknown	39
Last release year		2006	2010	2010	2010	Paper published in 2009
Last version		1.1	4	4	3	Not applicable
Ontology		A Single, Unified	Multiple	Multiple	Multiple	Multiple
Semantic Service Description Language		OWL-S 1.1	OWL-S 1.0, OWL-S 1.1	OWL-S 1.0, OWL-S 1.1	SAWSDL	OWL-S 1.1
Precondition and Result (Effect) Language		SWRL	SWRL	PDDL	Not applicable	SWRL

semantic descriptions. These features are categorized under expressivity, diverse scope, usability, scalability, and decoupling. We add a new test collection to this analysis. Moreover, the new versions of some of the test collections mentioned in their paper are investigated. Their paper lacks the required characteristics of world-altering services to be included in test collections.

To the best of our knowledge, there is not any other similar work on test collections of semantic Web services. This has been searched for both general semantic services and world-altering services.

VIII. SUMMARY

We have investigated some major test data collections in today's semantic service research field. Some of these test collections are publicly available, and there is also another test collection that was specifically used for a research experiment. We looked into all these test collections to find, in particular, their ability to be used for world-altering category of services. Furthermore, we presented the contests and the challenges of semantic services that use these test collections.

Overall, none of these test collections are suitable for a real standard test on world-altering services. A standard test collection for Web services needs to have a big number of test data, including a moderate number of information-providing services and world-altering services. The descriptions should be in various formalisms, along with the natural language descriptions of the services.

This lack of a proper test collection makes an empirical study on the world-altering semantic Web services an unsettled task.

ACKNOWLEDGMENT

REFERENCES

[1] S. A. McIlraith, T. C. Son, and H. Zeng, "Semantic web services," *IEEE Intelligent Systems*, vol. 16, no. 2, pp. 46–53, 2001.

[2] U. Küster and B. König-Ries, "Towards standard test collections for the empirical evaluation of semantic web service approaches," *International Journal of Semantic Computing*, vol. 2, no. 3, pp. 381–402, 2008.

[3] D. Martin, M. Paolucci, S. A. McIlraith, M. H. Burstein, D. V. McDermott, D. L. McGuinness, B. Parsia, T. R. Payne, M. Sabou, M. Solanki, N. Srinivasan, and K. P. Sycara, "Bringing semantics to web services: The OWL-S approach," in *First International Workshop on Semantic Web Services and Web Process Composition (SWSWPC)*, San Diego, CA, USA, July 2004, pp. 26–42.

[4] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean, "SWRL: A semantic web rule language combining OWL and RuleML," World Wide Web Consortium, W3C Member Submission, 2004. [Online]. Available: http://www.w3.org/Submission/SWRL

[5] M. Ghallab, C. K. Isi, S. Penberthy, D. E. Smith, Y. Sun, and D. Weld, "PDDL - The Planning Domain Definition Language," CVC TR-98-003/DCS TR-1165, Yale Center for Computational Vision and Control, Tech. Rep., 1998.

[6] SAWSDL Working Group, "Semantic Annotations for WSDL and XML Schema," Website, 2007, http://www.w3.org/TR/sawsdl/.

[7] A. B. Bener, V. Ozadali, and E. S. Ilhan, "Semantic matchmaker with precondition and effect matching using SWRL," *Expert Systems with Applications*, vol. 36, no. 5, pp. 9371–9377, 2009.

[8] R. Akkiraju, J. Farrell, J. Miller, M. Nagarajan, M.-T. Schmidt, A. Sheth, and K. Verma, "Web Service Semantics - WSDL-S," World Wide Web Consortium, W3C Member Submission, November 2005. [Online]. Available: http://www.w3.org/Submission/WSDL-S/

[9] J. d. Bruijn, C. Bussler, J. Domingue, D. Fensel, M. Hepp, U. Keller, M. Kifer, B. König-Ries, J. Kopecky, R. Lara, H. Lausen, E. Oren, A. Polleres, D. Roman, J. Scicluna, and M. Stollberg, "Web Service Modeling Ontology (WSMO)," Website, 2005, http://www.w3.org/Submission/WSMO/.

[10] M. Klusch and X. Zhing, "Deployed semantic services for the common user of the web: A reality check," in *International Conference on Semantic Computing (ICSC)*. Washington, DC, USA: IEEE Computer Society, August 2008, pp. 347–353.

[11] T. Margaria-Steffen, "The semantic web services challenge: Tackling complexity at the orchestration level," in *13th IEEE International Conference on Engineering of Complex Computer Systems*, Belfast, UK, April 2008, pp. 183–189.

[12] E. Prud'hommeaux and A. Seaborne, "SPARQL query language for RDF," Website, W3C, January 2008, http://www.w3.org/TR/rdf-sparql-query.