OSTIS-2013
(Open Semantic Technologies for Intelligent Systems)

УДК 004.822:514

SEMANTIC EVALUATION AT LARGE SCALE

Yatskevich M.I., Apanasovich V.V.

School of Business and Management of Technologies, Belarusian State University
Minsk, Republic of Belarus

yatskevichN@sbmt.by

apanasovich@sbmt.by

The Semantic Web technology often called Web 3.0 is aiming at making the information easily accessible for humans on the Web to be accessible for computers. The Semantic Web technologies are based on the notion of an ontology namely specification of conceptualization for a given domain. The growth and success of the Semantic Web is built upon a wide range of semantic technologies. The results obtained will facilitate adoption of the Semantic Web technologies within business environment and will be used for semantic technology adoption within School of Business and Management of Technologies.

Keywords: semantic web; evaluation; ontology; reasoning.

INTRODUCTION

Semantic technologies play a critical role in the recent advances in both the Web (the Semantic Web) and corporate knowledge management by providing ways to express knowledge and data in an automated way for different purposes such as information retrieval or data integration.

The evaluation of such technologies is crucial for their sustained improvement and adoption, allowing users to assess the suitability of current technologies to their needs. Some initiatives have already created a basis for semantic technology evaluation, such as those in the areas of ontology matching [Euzenat et. al, 2010], ontology engineering [Garcia-Castro et. al, 2009, Garcia-Castro et. al, 2010b], ontology reasoning [Horrocks et. al, 1998, Massacci et. al, 2000], semantic search [Kaufmann, 2007] or semantic web services [Klusch et. al, 2012, Petrie et. al, 2009].

At the heart of the EU-funded Framework 7 SEALS Project is the development of the SEALS Platform [Garcia-Castro et. al, 2010a]: an open infrastructure for the evaluation of semantic technologies that offers independent computational and data resources for the evaluation of those technologies. SEALS evaluation campaign included tools from five different semantic technology fields (ontology engineering, semantic search, semantic web services, ontology matching, storage and reasoning) were formally evaluated [Nixon et. al, 2011]. The evaluation results demonstrated high level of semantic technologies development highlighting critical elements of the semantic technologies ecosystem. However, some of the promises of the semantic technologies namely soundness and completeness of results had not been achieved by all the systems taking part in the evaluation.

The paper is structured as follows. Preliminaries of Semantic Web technologies are described in Section 2. Section 3 introduces technologies to be evaluated and their evaluation methodology while Section 4 briefly discusses the evaluation outputs.

1. Preliminaries

Semantic Web technologies are based on the several standardized languages (see Figure 1 for an updated stack). While URI and UNICODE provide a way for identification and dealing with text, eXtended Markup Language (XML) defines a way for introducing structured data. Resource Description Framework (RDF) introduces a way for defining graphs from XML data. RDF data is composed from triples corresponding to two nodes of the graph and labelled arc between them. Resource Description Framework Schema (RDFS) introduces ontological dimension allowing definition of classes, properties and is-a relationships among classes. Ontology Web Language (OWL) introduces more advanced constructs including cardinality restrictions, nominal, property restrictions, inheritance and composition. OWL language has a direct counterpart in description logics [Baader et. al, 2002] a family of reasoning formalisms restricting first-order logic to obtain sound and complete reasoning procedures. OWL introduces several profiles namely OWL Lite allowing for 0 or 1 cardinality constraints; OWL DL corresponding to decidable description logic and OWL Full allowing for maximum expressivity. The OWL 2.0 introduced OWL 2 EL allowing for reasoning in polynomial time; OWL 2 QL designed for maximum interoperability with relational databases; and OWL 2 RL designed for interoperability with rules-based systems.

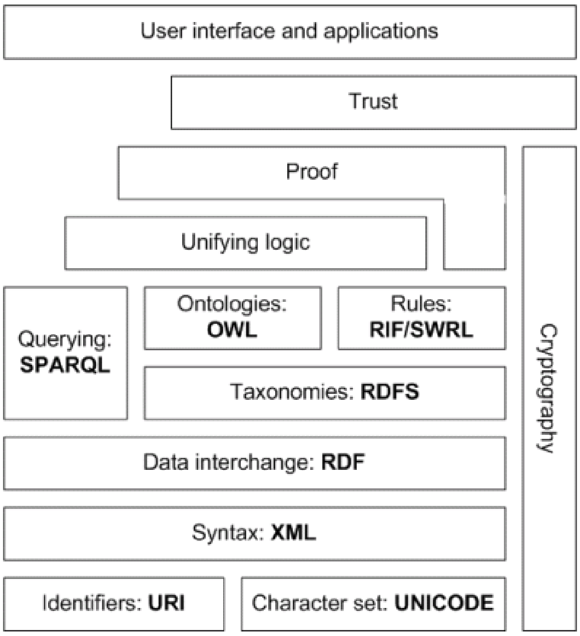

Figure 1. Semantic Web technologies stack.

SPARQL language allows for RDF data querying while Rule Interchange Format (RIF) and Semantic Web Rule Language (SWRL) allow adding ontological knowledge. The work on upper levels of the stack namely unifying logic, proof and trust is under way.

2. Target Technologies

The SEALS Project has identified five core technology areas which lie at the heart of the Semantic Web. The evaluations within these areas provide invaluable insights into the technologies themselves; insights which can be, and are being, used to improve performance of Semantic Web tools.

2.1. Ontology Engineering Tools

Two types of tools support ontology engineering tasks: ontology editors, which are user-oriented and allow creating and maintaining ontologies mainly through user interfaces, and ontology management programming interfaces, which are developer-oriented and allow the creation and maintenance of ontologies through programming interfaces. Conformance and interoperability evaluations [Garcia-Castro et. al, 2009, Garcia-Castro et. al, 2010b] use groups of ontologies defined in specific ontology languages as test data; these evaluations are performed by making tools process ontologies (coming either from test data or from other tools) and analysing the processed ontology (usually by comparing the processed ontology with that used as input).

2.2. Ontology Reasoning Tools

Description Logics (DLs) [Baader et. al, 2002] are a family of logic-based knowledge representation formalisms designed to represent and reason about the knowledge of an application domain in a structured and well-understood way. Besides their formal knowledge representation languages, DLs also provide inference services. The aim of such services is to extract new implied information out of the explicitly stated information. Every knowledge representation language usually offers a different set of inference services. The most widely used inference services include: class satisfiability, classification, logical entailment, ontology satisfiability, and instance retrieval. In order to interact with other systems an ontology reasoner must conform to standard input formats and must be able to provide standard inference services.

2.3. Ontology Matching Tools

Matching ontologies consists of finding a set of correspondences (alignment) between two different ontologies. A wide diversity of systems have been proposed, which can be classified according to the many features that can be found in ontologies (e.g., labels, structures, instances, semantics), or with regards to the techniques they use (e.g., statistics, combinatorics, semantics, linguistics, or machine learning) [Euzenat et. al, 2007b]. The most commonly used criterion for evaluating matching systems is the compliance of matcher alignments with respect to the expected reference alignments. Metrics such as precision and recall are largely adopted for quantitatively evaluating matching tools. Other evaluation criteria are efficiency, in terms of runtime and memory consumption, and scalability using large sets of tests; semantic measures, where the proximity between alignments is measured instead of their strict equality [Ehrig et. al, 2005, Euzenat, 2007a]; and task-specific evaluations, where alignments are evaluated according to their usage in some specific task.

2.4. Semantic Search Tools

State-of-the-art semantic search approaches are characterized by their high level of diversity both in their features as well as their capabilities. Such approaches employ different styles for accepting the user query (e.g., forms, graphs, keywords) [Uren et. al, 2007] and apply a range of different strategies during processing and execution of the queries. They also differ in the format and content of the results presented to the user. All of these factors influence the user’s perception of performance and usability. Semantic search technologies can be evaluated on the basis of different criteria and metrics [Wrigley et. al, 2010, Kaufmann, 2007]. At the core of any search task is the retrieval of pertinent information; search evaluations employ several questions which are applied to a particular ontology and dataset. Since (for ontology-based search) the answer set for each question is finite and known a priori, the measures of precision and recall are used. We are also interested in how tools cope with increasingly large datasets (scalability). Since search is an inherently user oriented task, evaluation must also consider metrics such as how long it takes for a query to be executed.

2.5. Semantic Web Service Tools

Semantic Web Service (SWS) technologies enable the automation of discovery, selection, composition, mediation and execution of web services by means of semantic descriptions of their interfaces, capabilities and non-functional properties. SWS provide a layer of semantics for service interoperability by relying on a number of reference service ontologies and semantic annotation extension mechanisms. The evaluation of SWS technologies is currently being pursued by a number of initiatives using different evaluation methods (e.g., see [Klusch et. al, 2012, Petrie et. al, 2009]). Although these initiatives have succeeded in creating an initial evaluation community in this area, they have been hindered by the difficulties in creating large-scale test suites and by the complexity of manual testing to be done.

3. Project outputs

In this paper we will concentrate on the outputs of the most recent SEALS evaluation campaign.

3.1. Ontology Engineering Tools

The evaluation was focused on conformance, interoperability and scalability criteria. Three ontology management frameworks and three ontology editors have been evaluated in the evaluation of the first two criteria. In the RDFS conformance task only three out of six systems had outputted the same ontologies they had in input. In OWL Lite five out of six systems had outputted the same ontologies they had in input and one achieved more than 95% precision. In OWL DL conformance tasks five out of six system had achieved 95% of precision or more while one system demonstrated more than 75% of precision. In content pattern, expressive pattern and content pattern full conformance tasks five systems demonstrated precision of 90% and more. In OWL Full conformance tasks two tools had demonstrated 100% of precision while the rest were in 10-80% range. For what concerns interoperability RDF(S) results 3 out of 6 tools had achieved 99% of interoperability. In OWL Lite interoperability tests the same level of interoperability had been achieved for all 5 systems. In OWL DL interoperability the level of 76% or higher had been achieved for all 6 systems while in OWL Full interoperability 3 of the systems had demonstrated 74% or higher levels of interoperability while the rest enjoyed 19% or lower levels of interoperability. In scalability tests 9 ontology managing frameworks and ontology editors had competed on 4 real world and artificially generated datasets on the tasks ranging from 0.5 to 1500 Mb. Ontology managing frameworks had outperformed the editors up to several orders of magnitude.

3.2. Ontology Reasoning Tools

The evaluation was focused on performance, interoperability and scalability criteria. Four description logic reasoners had participated in the evaluation campaign in classification, class satisfiability, ontology satisfiability, entailment and instance retrieval tasks. For classification tasks the “golden standard” classifications were available for about half of the 102 ontologies used in the task. The OWL DL supporting systems demonstrated 80-85% of precision in the task. The average reasoning time for all ontologies was within 3 seconds range while the time cut off limit was set to 1 hour. Two ontologies for one of the reasoners and one for the other had not been decided within the given time frame. In OWL EL tests all the ontologies had been classified correctly and average execution time was within tenth of a second for all the systems. For OWL DL class satisfiability tasks both systems demonstrated 95% or more of precision while execution times were within 0.5 seconds. In OWL EL class satisfiability tests all the systems demonstrated correct results while the reasoning time not exceeded tenth of a second. For OWL DL ontology satisfiability tasks both systems demonstrated precision in 97%-100% range. The average reasoning time was about 0.5 seconds for all the systems. All OWL EL ontologies have been solved within 0.2 seconds in average. For entailment tests one of the tools was able to solve 98% of tasks while the others less than 25%. For non-entailment tests OWL DL systems were able to solve more than 85% of tasks while OWL EL systems had not solved any of the tasks. For instance retrieval tests all the tasks were solved correctly. In scalability evaluations one of the systems demonstrated exponential behaviour in class and ontology satisfiability depending on the class counts while in other cases the behaviour of the systems was in line with complexities of the corresponding logics.

3.3. Ontology Matching Tools

The evaluation was focused on quality and scalability criteria. The rich set of ontologies has been collected including multilingual ontologies in nine languages. The size of ontologies ranged from tenth to tenth of thousands of classes and from tenth to hundreds of properties. Fourteen systems took part in the evaluation. Precision ranged from 1% to 99% in English and 1% to 97% in multilingual settings while recall was in 1% to 93% in English and 1% to 51% in multilingual settings. The tasks were solved in 7 to 66494 seconds depending on size.

3.4. Semantic Search Tools

The evaluation has been split into 2 phases user-in-the-loop and automated. The user-in-the-loop phase is concerned with asking a number of search questions from human subjects while in automated phase systems are assessed automatically. The datasets from geography and software engineering domains were used for the evaluation. Eight systems had participated in the evaluation. Twenty persons answered the questionnaire regarding usability of the tools. Perceived usability ranged from 32.5% to 63.75% while answer found rate was in 20% to 80% range and input time ranged from 19.9 to 102.52 seconds. The query times for automated phase ranged from 0.5 to 169.7 seconds.

3.5. Semantic Web Service Tools

Three datasets were used for retrieving web services that match to a given partial specification. Three systems participated in the task. Precision ranged from 13% to 98% while recall was in 4% to 92% range.

Conclusion

The evaluation results demonstrated importance of mutual assessments and perceived scalability for a wide range of semantic technologies. The results obtained will facilitate adoption of the Semantic Web technologies within business environment and will be used for semantic technology adoption within School of Business and Management of Technologies.

References

[Baader et. al, 2002] F. Baader, D. McGuinness, D. Nardi, and P. Patel-Schneider. The Description Logic Handbook: Theory, implementation and applications. Cambridge University Press, 2002.

[Ehrig et. al, 2005] M. Ehrig and J. Euzenat. Relaxed precision and recall for ontology matching. In B. Ashpole, M. Ehrig, J. Euzenat, and H. Stuckenschmidt, editors, Proc. of the Workshop on Integrating Ontologies, volume 156, page 8. CEUR-WS.org, August, 2005.

[Euzenat, 2007a] J. Euzenat. Semantic precision and recall for ontology alignment evaluation. In Proc. of the 20th International Joint Conference on Artificial Intelligence, pages 248–253, Hyderabad, India, 2007.

[Euzenat et. al, 2010] J. Euzenat, A. Ferrara, C. Meilicke, A. Nikolov, J. Pane, F. Scharffe, P. Shvaiko, H. Stuckenschmidt, O. Sv.ab-Zamazal, V. Svatek, and C. Trojahn dos Santos. Results of the Ontology Alignment Evaluation Initiative 2010. In Proc. of the 5th Workshop on Ontology Matching, pages 85–117, Shanghai, China, 2010.

[Euzenat et. al, 2007b] J. Euzenat and P. Shvaiko. Ontology matching. Springer, Heidelberg(DE), 2007.

[Garcia-Castro et. al, 2010a] R. Garcia-Castro, M. Esteban-Gutierrez, and A. Gomez-Perez. Towards an infrastructure for the evaluation of semantic technologies. In Proc. of the eChallenges 2010 Conference, October 27-29 2010.

[Garcia-Castro et. al, 2009] R. Garcia-Castro and A. Gomez-Perez. RDF(S) Interoperability Results for Semantic Web Technologies. International Journal of Software Engineering and Knowledge Engineering, 19(8):1083–1108, 2009.

[Garcia-Castro et. al, 2010b] R. Garcia-Castro and A. Gomez-Perez. Interoperability results for Semantic Web technologies using OWL as the interchange language. Web Semantics: Science, Services and Agents on the World Wide Web, 8:278–291, November 2010.

[Horrocks, 1998] I. Horrocks and P. F. Patel-Schneider. DL systems comparison. In Proc. of the 1998 Description Logic Workshop, volume 11, pages 55–57, 1998.

[Kaufmann, 2007] E. Kaufmann. Talking to the Semantic Web — Natural Language Query Interfaces for Casual End-Users. PhD thesis, Faculty of Economics, Business Administration and Information Technology of the University of Zurich, September 2007.

[Klusch et. al, 2012] M. Klusch, A. Leger, D. Martin, M. Paolucci, A. Bernstein, and U. Kuester. Annual International Contest S3 on Semantic Service Selection. http://www-ags.dfki.uni-sb.de/~klusch/s3/

[Massacci et. al, 2000] F. Massacci and F. M. Donini. Design and results of TANCS-2000. In Proc. of the International Conference on Automated Reasoning with Analytic Tableaux and Related Methods, pages 52–56, London, UK, 2000. Springer-Verlag.

[Nixon et. al, 2011] L. Nixon, R. Garcia-Castro, S. N. Wrigley, M. Yatskevich, C. T. D. Santos, and L. Cabral. The state of semantic technology today – overview of the first seals evaluation campaigns. In Proc. of the 7th International Conference on Semantic Systems, 2011.

[Petrie et. al, 2009] C. Petrie, T. Margaria, H. Lausen, and M. Zaremba. Semantic Web Services Challenge: Results from the First Year. Springer, 2009.

[Uren et. al, 2007] V. Uren, Y. Lei, V. Lopez, H. Liu, E. Motta, and M. Giordanino. The usability of semantic search tools: a review. The Knowledge Engineering Review, 22(4):361–377, 2007.

[Wrigley et. al, 2010] S. N. Wrigley, D. Reinhard, K. Elbedweihy, A. Bernstein, and F. Ciravegna. Methodology and campaign design for the evaluation of semantic search tools. In Proc. of the International Workshop on Semantic Search, 2010.

КРУПНОМАСШТАБНАЯ ОЦЕНКА СЕМАНТИЧЕСКИХ ТЕХНОЛОГИЙ

Яцкевич Н.И., Апанасович В.В.

**Институт Бизнеса и Менеджмента Технологий, Белорусский ГосударственныйУниверситет, Минск, Республика Беларусь*

yatskevichN@sbmt.by

apanasovich@sbmt.by

Проект Оценка Семантической Масштабируемости создает открытую и устойчивую платформу, в которой все аспекты процесса оценки, его размещения и выполнения принимаются во внимание для большинства типов технологий.

ВВЕДЕНИЕ

Технологии семантического веба, известные как Веб 3.0, направлены на предоставление компьютерам того же уровня понимания информации что и людям. Они основаны на понятии онтологии или спецификации концептуализации данной предметной области.

ОСНОВНАЯ ЧАСТЬ

Развитие и успех семантического веба построены на широком спектре технологий, а именно, редакторов онтологий, систем автоматического принятия решений построенных на онтологических принципах, систем нахождения онтологических отображений, технологий семантического поиска информации и веб служб. Оценка качества и совместимости этих технологий критична для их устойчивого развития и распространения. Результаты оценки продемонстрировали высокий уровень развития семантических технологий и подчеркнули наиважнейшие их элементы.

ЗАКЛЮЧЕНИЕ

Полученные результаты продемонстрировали важность оценки взаимодействия и масштабируемость всего спектра семантических технологий и будут использованы для их внедрения в Институте Бизнеса и Менеджмента Технологий Белорусского Государственного Университета.