Towards an Infrastructure for the Evaluation of Semantic Technologies

Raúl GARCÍA-CASTRO, Miguel ESTEBAN-GUTIÉRREZ, Asunción GÓMEZ-PÉREZ
Ontology Engineering Group, Universidad Politécnica de Madrid,
Campus de Montegancedo s/n, Boadilla del Monte, 28660, Spain
Tel: +34 913363670, Fax: + 34 913524819, Email: {rgarcia,mesteban,asun}@fi.upm.es

Abstract: This paper presents and discusses the current development status of the SEALS Platform, a lasting reference infrastructure for semantic technology evaluation. It describes the different entities managed by the platform and the ontology-based model that has been defined to represent them; it also provides an overview of the platform architecture. In addition, it presents the different challenges faced during the development of the SEALS Platform and a use scenario of the platform that supports the execution of evaluation campaigns over semantic technologies.

1. Introduction

One of the challenges of the future content and knowledge technologies is to manage and combine information about different digital and real-world entities and the characteristics of these entities. A way of having an effective representation and integration of this information is to use semantic technologies to correctly manage not just these heterogeneous content and data but also their associated metadata.

The SEALS European project (www.seals-project.eu) is developing an infrastructure for the evaluation of semantic technologies, named the SEALS Platform, that will offer independent computational and data resources for the evaluation of these technologies. This three-year project started in June 2009 and is funded by the Research Infrastructures area of the FP7 Capacities programme.

With the SEALS Platform users will define and execute evaluations on their own and will support the organization and execution of evaluation campaigns, i.e., worldwide activities in which a set of tools is evaluated according to a certain evaluation specification and using common test data.

This paper is structured as follows. Section 2 includes the objectives of the paper and the underlying methodology followed. Sections 3 and 4 present the different entities managed by the platform and the ontology-based model that has been defined to represent them, respectively. Section 5 provides an overview of the architecture of the SEALS Platform, whereas section 6 enumerates the challenges being faced during the platform development. Section 7 describes a use scenario of the SEALS Platform that supports the execution of evaluation campaigns over semantic technologies. Finally, section 8 draws the conclusions from this work.

2. Objectives and Methodology

The goal of the paper is to present and discuss the development of the SEALS Platform, a lasting reference infrastructure for semantic technology evaluation.

By developing an evaluation infrastructure for semantic technologies, we want to make sure that in the current explosion of semantic technologies we can support an

experimentation-driven research and development of these technologies and assess whether semantic technologies cover future requirements (in terms of their efficiency, scalability, interoperability, etc.).

In the past, from the first general framework for evaluating semantic technologies [1], researchers evaluated different types of semantic technologies, such as ontology engineering tools [2, 3], ontology repositories and reasoners [4, 5, 6], ontology matching tools [7, 8, 9], or semantic search tools [10, 11].

Besides, different community efforts have dealt with open evaluations of semantic technologies such as the RDF(S) and OWL Interoperability Benchmarking activities[i] [12], the Ontology Alignment Evaluation Initiative[ii], or the Semantic Web Service Challenge[iii] [13].

Now, by means of the SEALS Platform, we plan to connect all these evaluation initiatives by providing a common evaluation framework that supports the reuse and exploitation of the different available resources for semantic technology evaluation.

This paper presents the current development status of the platform, including the decisions made and the lessons learnt. This development has been performed regarding not only the technical requirements of the platform, but also the input from the top researchers in the semantic area, who will define their evaluations in the platform and organize the evaluation campaigns supported by it.

3. The SEALS Entities

One core result in the development of any research infrastructure is the definition of the data model to be used; in our case, this model is the one for representing software evaluations, evaluation campaigns and the rest of the entities managed by the SEALS Platform.

The SEALS Platform revolves around the notion of software evaluation, which is largely inspired by the ISO/IEC 14598 standard on software product evaluation [14]. Nevertheless, it is not our intention to fully cover this standard but to focus on the entities required to describe software evaluations and automatically execute them.

The entities included in our notion of evaluation are the following: in an *evaluation* a given set of *tools* is exercised, following the workflow defined by a given *evaluation description* and using determined *test data*. As a result of this process, a set of *evaluation results* is produced.

This high-level classification of entities can be further refined. Thus, semantic tools are classified in different types according to their functional scope, namely, ontology engineering tools, storage and reasoning systems, matching tools, semantic search tools, and semantic web service tools.

Similarly, it is also possible to distinguish different types of test data: persistent test data (data whose contents are stored in and physically managed by the SEALS Platform), external test data (data whose contents have an existence outside the platform and whose lifecycle is not controlled by it), and synthetic test data generators (pieces of software that can generate synthetic test data on-the-fly according to some determined parameters).

In accordance with the approach followed in the IEEE 1061 standard for a software quality metrics methodology [15], evaluation results are classified according to their provenance, differentiating raw results (those evaluation results directly generated by tools) from interpreted results (those generated from other evaluation results).

Moreover, our entities include not only the results obtained in the evaluation but also any contextual information related to such evaluation, a need also acknowledged by other authors [16]. To this end, we also represent the information required for automating the execution of an evaluation description in the platform, which, with the rest of the entities presented, yields traceable and reproducible evaluation results.

4. The SEALS Metadata

Our design principles when defining the metadata model for representing these entities and the relationships between them were that the model should be

- *machine-processable* so that it supports the automation of the evaluation process,
- *exhaustive* so that evaluations can be reproducible,
- *interoperable* so that it can interchange evaluation-related information between different systems, and
- *extensible* so that it will have to be expanded and then used in concrete evaluations and evaluation campaigns.

To cover these requirements we decided to use ontologies for representing such model. Ontologies are formal and explicit specifications of a conceptualization [17] for representing consensual knowledge; they are easily extensible, and support interoperability at the knowledge level. The OWL ontology language [18] has been used for implementing these ontologies.

Since the entities presented above share a number of common properties, we developed an upper ontology (shown in Figure 1) to represent them, as well as different ontologies covering each entity domain.

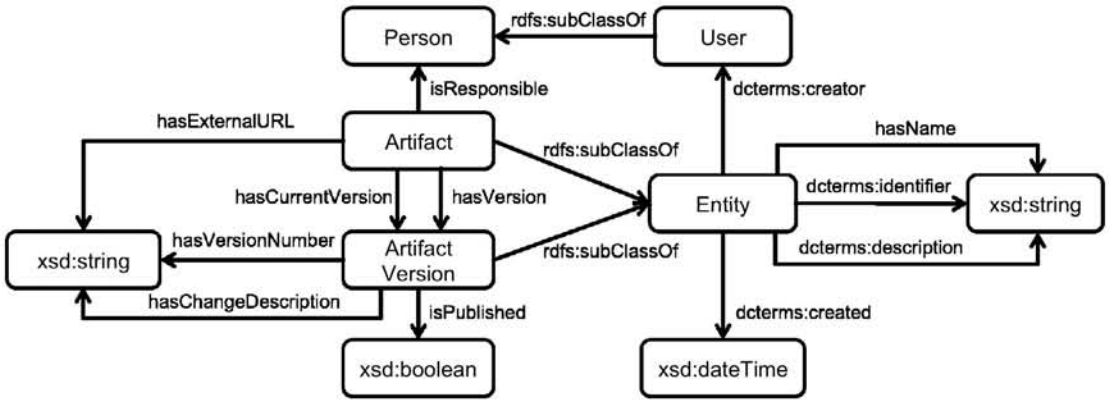

Figure 1: Overview of the SEALS Upper Ontology

During the definition of the ontologies we tried, when possible, to reuse current standards and models. Thus now we reuse existing ontologies such as the Dublin Core [19], FOAF [20] or VCard [21] ones.

5. Architecture of the SEALS Platform

The SEALS Platform will be an independent, open, scalable, extensible, and sustainable infrastructure with which semantic technologies can be evaluated remotely by providing an integrated set of evaluation services and test data. The SEALS Platform will be used in public world-wide evaluation campaigns, and the results of these evaluation campaigns will be employed in creating semantic technology roadmaps, identifying sets of efficient and compatible tools for developing large-scale semantic applications.

The architecture of the SEALS Platform comprises a number of components in charge of providing different functionalities and managing the platform entities (both their data and metadata). These components, shown in Figure 2, are described next.

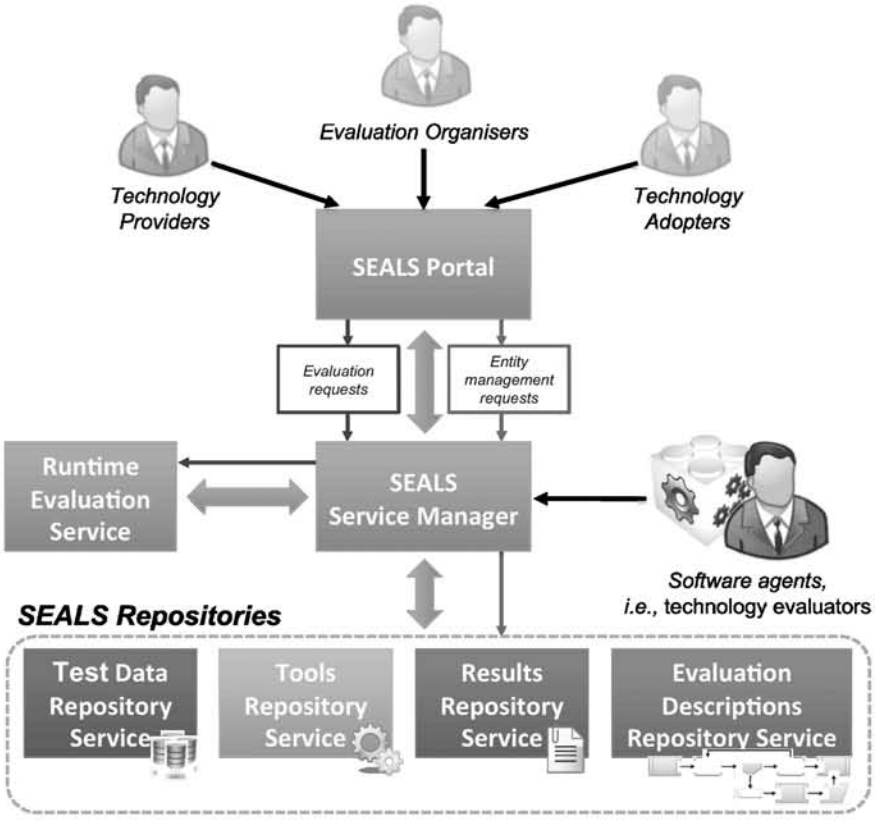

Figure 2: Architecture of the SEALS Platform

- **SEALS Portal**. The SEALS Portal provides a user interface for interacting with the SEALS Platform. Portal users can use the portal for managing contents in the SEALS Platform and requesting the execution of evaluations. The portal relies on the SEALS Service Manager for carrying out the users' requests.
- **SEALS Service Manager**. The SEALS Service Manager is the core module of the platform and is responsible for coordinating the other components within the platform and for maintaining consistency within the platform. This component exposes a series of services that provide programmatic interfaces to the SEALS Platform. Thus, apart from the SEALS Portal, the services offered can also be used by third party software agents.
- **Evaluation Descriptions Repository**. This repository manages the descriptions of the different evaluations that can be conducted.
- **Test Data Repository**. This repository manages the test data that can be used with the different evaluation descriptions.
- **Tools Repository**. This repository manages the tools that can be subject to an evaluation.
- **Runtime Evaluation Service**. The Runtime Evaluation Service is used to automatically evaluate a certain tool according to a particular evaluation description and using some specific test data.
- **Result Repository**. This repository is used to store the evaluation results produced by the Runtime Evaluation Service. These results are raw result data and any interpretation produced over these data.

6. Development Challenges

Developing the SEALS Platform requires facing numerous challenges since the platform is expected to be

- **Generic**. The platform needs to cope with the different heterogeneous semantic technologies and with the different evaluations that could be performed over them. On the one hand, this requires reconciling heterogeneity in the technical level, where we need to execute evaluations by uniformly accessing semantic technologies with different hardware and software requirements. On the other hand, in the information level we need to achieve a common understanding of all the platform entities.
- **Extensible and interoperable**. These characteristics are also required at all levels. The platform follows a service-oriented architecture and exposes its public services through standard formats (e.g., WSDL, XML Schema, WADL) and protocols (e.g., HTTP, SOAP over HTTP and JMS). Besides, the SEALS metadata schemas have been defined as ontologies and the metadata of the different platform entities will be publicly available.
- **Dynamic**. The platform development process is based on evolving prototypes since we need to cope with evolving requirements; and this is so because, on the one hand, not every requirement was known at design time and, on the other hand, we expected changes in the requirements as a result of the lessons learnt during development.
- **Robust**. The quality assurance process involves the participation of all the relevant stakeholders in testing, including not only developers but also end users; this process covers testing at different levels (i.e., unit, integration and acceptance). It must be added that the Continuous Integration practice is followed to improve quality and reduce risks during development.
- **Efficient**. The platform needs to run in an efficient hardware infrastructure. Besides, it must scale in terms of users and of the size of the data managed. To this end, the platform needs to ensure high-availability and to have load-balancing capabilities.
- **Open**. The platform will be openly available to anyone interested in evaluating semantic technologies or in the evaluation results of such technologies. However, special attention should be paid to data licensing issues, not only regarding test data but also regarding the evaluation results obtained by the tools.
- **Independent**. Users must have trust in the platform and in the evaluation results obtained with it. To this end, users will be able to analyse all the entities used during evaluations as well as to reproduce evaluation results.
- **Sustainable**. The platform is being developed under an open source approach because the platform and its development should remain beyond the end of the SEALS project. Furthermore, special attention has to be drawn to topics such as facilitating the deployment and the administration of the infrastructure.

7. Using the SEALS Platform in Evaluation Campaigns

The semantic technology evaluation services will initially be available for five different types of technologies (ontology engineering tools, ontology storage and reasoning systems, ontology matching tools, semantic search tools, and semantic web service tools) and for different evaluation criteria (interoperability, scalability, etc.).

During the SEALS project, we have planned to organize public world-wide evaluation campaigns that cover the five different types of technologies mentioned above, and these will follow other well-known evaluation campaigns in the Semantic Web area such as the Ontology Alignment Evaluation Initiative, the RDF(S) and OWL Interoperability Benchmarking activities[iv], the Semantic Web Service Challenge[v], and the PASCAL Challenges[vi].

All the evaluation campaigns will contain a set of evaluation scenarios that will, in turn, use different test data. The use of the SEALS Platform in these evaluation scenarios will be the following:

1. The evaluation campaign organizers will populate the platform with the different evaluation descriptions and test data to be used in the evaluation scenarios.
2. Technology providers will be able to participate in the evaluation campaigns by uploading their tools into the platform and registering them to one or more evaluation scenarios.
3. All the evaluation scenarios will be executed using the participating tools with the predetermined evaluation descriptions and test data. This will produce different evaluation results (both raw results and interpretations of them).
4. The results of the evaluation scenarios will be available through the SEALS Portal. This way, the evaluation campaign organizers will obtain feedback about the evaluation campaigns, evaluation descriptions and test data; technology providers will be able to know the capabilities of their tools and to compare them with others; and technology adopters will be able to see and compare the results of the different existing technologies.

8. Conclusions

The development of the SEALS Platform and the definition of the different evaluation descriptions and test data to be used in the platform are currently taking place in parallel with the organization of the different evaluation campaigns, which will take place during the summer of 2010.

At the end of the evaluation campaigns, all the resources used (from test data to results) will be publicly available, so people will be able to perform their own evaluations through the SEALS Platform or to browse and compare the different results. Additionally, we plan to publish all the data on evaluations and evaluation campaigns stored the SEALS Platform as RDF data.

Furthermore, we expect that the evaluations and test data in the platform will be extended by the community; users will be able to define and execute their own evaluations, either with the test data provided or with their own, and to reproduce the results of any existing evaluation.

SEALS will work towards the creation of an open and sustainable worldwide community focused on the evaluation and progressive development of semantic technology, which will survive the SEALS project. Our long-term goal is that the SEALS Platform be actively used and managed by the semantic community.

SEALS will innovate the way in which semantic technology is evaluated. The infrastructure developed within SEALS is, therefore, expected to provide future benchmarks for both industry and academia to evaluate their applications/innovations.

Previously we mentioned some of the challenges being faced during the development of the SEALS Platform, which any other similar initiative has met or will probably meet. It should be added that the long-term success of a research infrastructure like the SEALS Platform depends on facing also other types of challenges, for example, social (e.g., involving the community in the evaluations and evaluation campaigns in order to obtain relevant and agreed upon outcomes) or financial (e.g., getting funding for the platform after the end of the project) ones.

Acknowledgements

This work is supported by the SEALS European project (FP7-238975). Thanks to Rosario Plaza for reviewing the grammar of this paper.

References

[1] OntoWeb Deliverable 1.3. A survey on ontology tools. Technical report, OntoWeb Thematic Network, May 2002.

[2] Y. Sure and O. Corcho (eds.). Proceedings of the 2nd International Workshop on Evaluation of Ontology-based Tools (EON 2003). Volume 87 of CEUR-WS, Florida, USA, October 2003.

[3] R. García-Castro and A. Gómez-Pérez. Guidelines for benchmarking the performance of ontology management APIs. Proceedings of the 4th International Semantic Web Conference (ISWC 2005). Galway, Ireland. LNCS 3729, pp 277-292. 6th-10th November, 2005.

[4] Y. Guo, Z. Pan, and J. Heflin. LUBM: A Benchmark for OWL Knowledge Base Systems. Journal of Web Semantics 3(2), pp 158-182. 2005.

[5] L. Ma, Y. Yang, Z. Qiu, G. Xie, Y. Pan, and S. Liu. Towards a Complete OWL Ontology Benchmark. Proceedings of the 3rd European Semantic Web Conference (ESWC 2006), Budva, Montenegro, 2006.

[6] T. Gardiner, D. Tsarkov, and I. Horrocks. Framework For an Automated Comparison of Description Logic Reasoners. Proceedings of the 5th International Semantic Web Conference (ISWC 2006), LNCS 4273, pp 654-667. Springer, 2006.

[7] H. Do, S. Melnik, and E. Rahm. Comparison of schema matching evaluations. Proceedings of the Workshop on Web, Web-Services, and Database Systems, pp 221– 237, 2002.

[8] P. Avesani, F. Giunchiglia, and M. Yatskevich. A large scale taxonomy mapping evaluation. Proceedings of the 4th International Semantic Web Conference (ISWC 2005), pp 67– 81, 2005.

[9] J. Euzenat and P. Shvaiko. Ontology matching. Springer, 2007.

[10] E. Kaufmann, A. Bernstein, L. Fischer. Nlp-reduce: A "naïve" but domain independent natural language interface for querying ontologies. Proceedings of the 4th European Semantic Web Conference (ESWC2007), Innsbruck, Austria, 2007.

[11] R. Bhagdev, S. Chapman, F. Ciravegna, V. Lanfranchi, and D. Petrelli. Hybrid Search: Effectively Combining Keywords and Semantic Searches. Proceedings of the 5th European Semantic Web Conference (ESWC 2008), Tenerife, June 2008.

[12] R. García-Castro. Benchmarking Semantic Web technology. AKA - IOS Press. 2010.

[13] C. Petrie, T. Margaria, H. Lausen, and M. Zaremba (eds.). Semantic Web Services Challenge. Springer, 2008.

[14] ISO/IEC 14598-6: Software product evaluation - Part 6: Documentation of evaluation modules. ISO/IEC, 2001.

[15] IEEE 1061-1998. IEEE Standard for a Software Quality Metrics Methodology. IEEE, December 1998.

[16] B. A. Kitchenham, R. T. Hughes, and S. G. Linkman. Modeling software measurement data. IEEE Transactions on Software Engineering, vol. 27, no. 9, pp 788–804, 2001.

[17] R. Studer, V. R. Benjamins, and D. Fensel. Knowledge engineering: Principles and methods. IEEE Transactions on Data and Knowledge Engineering, vol. 25, no. 1-2, pp. 161–197, 1998.

[18] D. McGuiness and F. van Harmelen. OWL Web Ontology Language Overview. W3C, Recommendation, 10 February 2004. Available: http://www.w3.org/TR/owl-features/

[19] DCMI Usage Board. DCMI Metadata Terms. Dublin Core Metadata Initiative, Recommendation, 14 January 2008. Available: http://dublincore.org/documents/dcmi-terms/

[20] D. Brickley and L. Miller. FOAF vocabulary specification 0.97. FOAF Project, Namespace Document - 3D Edition, 1 January 2010. Available: http://xmlns.com/foaf/spec/

[21] H. Halpin, R. Ianella, B. Suda, and N. Walsh. Representing vCard objects in RDF. W3C, Member Submission, 20 January 2010. Available: http://www.w3.org/TR/vcard-rdf/

---

[i] http://knowledgeweb.semanticweb.org/benchmarking_interoperability/

[ii] http://oaei.ontologymatching.org/

[iii] http://sws-challenge.org/

[iv] http://knowledgeweb.semanticweb.org/benchmarking_interoperability/

[v] http://sws-challenge.org/

[vi] http://pascallin.ecs.soton.ac.uk/Challenges/