Sustainability of Evaluations Presented in Research Publications

Raúl García-Castro

Ontology Engineering Group,
Departamento de Lenguajes y Sistemas Informáticos e Ingeniería Software
Facultad de Informática, Universidad Politécnica de Madrid, Spain
rgarcia@fi.upm.es

Abstract. This position paper discusses how research publication would benefit of an infrastructure for evaluation entities that could be used to support documenting research efforts (e.g., in papers or blogs), analysing these efforts, and building upon them. As a concrete example in the domain of semantic technologies, the paper presents the SEALS Platform and discusses how such platform can promote research publication.

1 Introduction

The way of publishing evaluation-related information in research papers should be rethought to facilitate the use of such information.

The limited extension of research papers does not allow a full description of all the entities involved in evaluations (evaluation workflows, test data, tools and results) and of the concrete context in which evaluations were performed. Therefore, it is difficult if not impossible to reproduce evaluations described in research papers and to validate them; this forces researchers in most cases to blindly trust in the paper claims. Besides, both technologies and evaluation methods evolve over time and the evaluation data included in the paper becomes rapidly outdated.

Another perspective to take into account is that of advancing research by building upon existing one. Defining and performing evaluations is expensive and prone to errors. This is mainly because, besides the lack of full evaluation descriptions mentioned above, most lessons learnt during evaluation (both positive and negative) are not explicit in research papers. Furthermore, performing complex analyses across research papers (e.g., finding correlations between the results of different evaluations) is currently not possible.

The goal of this paper is to discuss how research would benefit of an infrastructure for evaluation entities that could be used to support documenting research efforts (e.g., in papers or blogs), analysing these efforts, and building upon them.

Such infrastructure would allow anyone reading or reviewing a paper to completely analyse the evaluations presented in the paper and to validate them, taking advantage of dynamic and enhanced result visualisations.

Besides, it would permit anyone to reproduce the evaluation presented in the paper under the same settings or using updated or alternative versions of the evaluation entities.

Furthermore, anyone interested in building upon existing evaluations could reuse the evaluation presented in the paper (fully or parts of it) and even combine the results from evaluations in different papers.

Clearly, someone could disagree with the above-mentioned claims; the main stands against them could be the following:

- *Refusal to unveil evaluation details.* In research environments, people are used to having other people review their work in detail, so this should be no problem. Besides, being against this would incline people to think that the researcher is hiding something.
- *Refusal to share work with others.* This opinion is also not expected since in research environments people are usually eager to be reused cited.
- *Refusal to devote effort to share evaluation details.* Even if researchers acknowledge the added value of sharing their evaluations, they will be reluctant to do so unless the benefits compensate their spent efforts.
- *Refusal to reuse work from others.* Even if the do-it-yourself attitude is characteristic of computer science researchers, they are also aware of the benefits of reuse; therefore, this is something that should not pose rejection.

The SEALS (Semantic Evaluation at Large Scale) European project[1] is developing an infrastructure (the SEALS Platform) that offers independent computational and data resources for the evaluation of semantic technologies [1].

Next, the paper presents an overview of the SEALS Platform and then discusses how such infrastructure could support the publishing and management of evaluation information in research papers, providing different benefits along the lines presented above.

2 An Infrastructure for Semantic Technology Evaluation

The idea of software evaluation followed in the SEALS Platform is largely inspired by the notion of evaluation as defined by the ISO/IEC 14598 standard on software product evaluation [2]. In any *evaluation* a given set of *tools* are executed, following a given *evaluation workflow* and using determined *test data*. As an outcome of this process, a set of *evaluation results* is produced.

This high-level classification of software evaluation entities can be further refined as needed; a detailed description of them and their life cycles can be found in [3]. For example, in accordance with the approach followed in the IEEE 1061 standard for a software quality metrics methodology [4], 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).

---

[1] http://www.seals-project.eu/

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 [5]. 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.

The SEALS Platform has been developed around these evaluation entities and following a service-oriented approach. The architecture of the platform comprises a number of components, shown in Figure 1, which are described below.
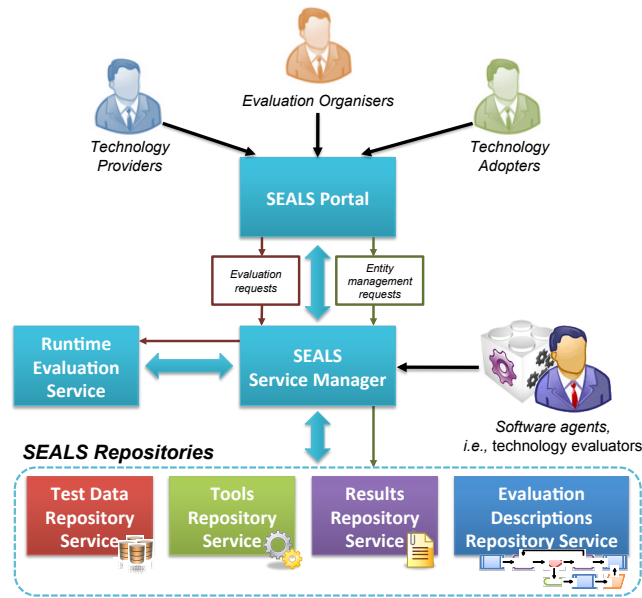

Fig. 1: Architecture of the SEALS Platform.

- **SEALS Portal.** The SEALS Portal provides a web user interface for interacting with the SEALS Platform. Thus, the portal will be used by the users for the management of the entities in the SEALS Platform, as well as for requesting the execution of evaluations.
- **SEALS Service Manager.** The SEALS Service Manager is the core module of the platform and is responsible for coordinating the other platform components and for maintaining consistency within the platform. This component exposes a series of services that provide programmatic interfaces for the SEALS Platform. Thus, apart from the SEALS Portal, the services offered may be also used by third party software agents.
- **SEALS Repositories.** These repositories manage the entities used in the platform: test data, tools, results, and evaluation workflows.

– **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.

All the evaluation entities stored in the platform are described according to
a set of OWL ontologies[2] [3]. Since the entities presented above share a number
of common properties, we developed an upper ontology to represent them, as
well as different ontologies covering each entity domain. During the definition of
the ontologies we tried, when possible, to reuse current standards and models
(i.e., Dublin Core, FOAF, VCard).

3 Publication and Management of Evaluation Information

This section discusses how the SEALS Platform could support the publication
and management of evaluation information.

The SEALS Platform offers manual and programmatic access to the evalua-
tion entities stored in its repositories. This allows linking the evaluation resources
mentioned in research papers to the actual resources stored in the platform. Be-
sides, if reverse links were created from the evaluation entities to research papers,
networks of papers around concrete evaluations could be built.

The SEALS Platform also allows storing different versions of tools, evaluation
workflows and test data. This way, it maintains the traceability from the concrete
evaluation used in one paper to those evaluations that include updated versions
of tools, evaluation workflows or test data.

All the evaluation entities stored in the SEALS Platform are described us-
ing ontologies with the aim of having consensual and interoperable descriptions.
These machine-processable descriptions can be published in the Web or be em-
bedded in research papers. Furthermore, it provides dynamic and interactive
visualisations of evaluation results that could be used in non-standard research
papers (e.g., multimedia or interactive documents).

In the SEALS Platform, evaluation reproducibility is a main requirement. To
this end, evaluations are only executed over persistent (i.e., unmodifiable) entities
and the whole evaluation execution context is stored. This allows replicating the
concrete evaluation presented in a research paper at any moment and by anyone.

All the evaluation entities can not only be accessed but also be reused both
inside and outside the SEALS Platform. This reuse can be performed as a whole
(e.g., reusing some test data in another evaluation infrastructure) or partially
(e.g., evaluation workflows are defined with the BPEL language and new work-
flows can be defined from existing workflows and services).

Finally, since evaluation results are represented following common schemas
(i.e., ontologies), researchers could exploit these results in unexpected ways. To
allow this, we have defined a quality model for semantic technologies that defines
the main quality characteristics of such technologies and allows the combination
and comparison of results from different evaluations [6].

---

[2] http://www.seals-project.eu/ontologies/

4 Conclusions

This paper proposes to support research publications (or any other type of research documentation) through an infrastructure for evaluation entities. Having such infrastructure would allow, on the one hand, connecting research publications with the actual evaluations used in them and, on the other hand, interconnecting different research efforts.

The SEALS Platform aims to support these ideas in the domain of semantic technologies with the ultimate goal of increasing the maturity of the semantic research community by enriching the body of knowledge on semantic technology evaluation and by encouraging an experimentation-based research.

However, the project is still in its way to achieve the approach presented in this paper since functionalities for linking evaluations with publications are not planned yet. To this end, future challenges to be faced are not only technological but also social (e.g., it requires greater commitment since researchers have to invest more effort than they are now) or legal (e.g., important issues are the access and use policies for evaluation data).

Furthermore, the success of such approach will depend on the existence of software technologies that are coupled to researchers' working environments and that leverage the effort of using an infrastructure such as the SEALS Platform in day-to-day research.

References

1. García-Castro, R., Esteban-Gutiérrez, M., Gómez-Pérez, A.: Towards an infrastructure for the evaluation of semantic technologies. In: Proceedings of the eChallenges 2010 Conference, Warsaw, Poland (2010)
2. ISO/IEC: ISO/IEC 14598-6: Software product evaluation - Part 6: Documentation of evaluation modules (2001)
3. García-Castro, R., Esteban-Gutiérrez, M., Kerrigan, M., Grimm, S.: An ontology model to support the automatic evaluation of software. In: Proceedings of the 22nd International Conference on Software Engineering and Knowledge Engineering (SEKE 2010), Redwood City, CA, USA (2010) 129–134
4. IEEE: IEEE 1061-1998. IEEE Standard for a Software Quality Metrics Methodology (1998)
5. Kitchenham, B.A., Hughes, R.T., Linkman, S.G.: Modeling software measurement data. IEEE Trans. Softw. Eng. **27** (2001) 788–804
6. Radulovic, F., García-Castro, R.: Towards a Quality Model for Semantic Technologies. In: Proceedings of the 11th International Conference on Computational Science and Its Applications (ICCSA 2011), Santander, Spain, (To be published) (2011)