Semantic Web Conference Ontology - A Refactoring Solution

Andrea Giovanni Nuzzolese[1], Anna Lisa Gentile[2(✉)], Valentina Presutti[1], and Aldo Gangemi[1,3]

[1] Semantic Technology Lab, ISTC-CNR, Roma, Italy
andrea.nuzzolese@istc.cnr.it, {valentina.presutti,aldo.gangemi}@cnr.it
[2] Data and Web Science Group, University of Mannheim, Mannheim, Germany
annalisa@informatik.uni-mannheim.de
[3] LIPN, Université Paris 13, Sorbone Cité, UMR CNRS, Paris, France

Abstract. The Semantic Web Dog Food (SWDF) is the reference linked dataset of Semantic Web community about papers, people, organisations, and events related to its academic conferences. In this paper we analyse the existing problems, of generating, representing and maintaining Linked Data for the SWDF. Accordingly, we discuss a refactoring of the Semantic Web Conference Ontology by adopting best ontology design practices (e.g., Ontology Design Patterns, ontology reuse and interlinking). We regenerate metadata for a set of conferences already existing in SWDF, using cLODg (conference Linked Open Data generator), an Open Source workflow which adopts the proposed refactoring.

1 Introduction

A good practise in the Semantic Web community is to encourage the publication of Linked Data about scientific conferences in the field, as a way of "eating our own dog food" [5]. The main example is the *Semantic Web Dog Food*[1] (SWDF), a corpus that collects Linked Data about papers, people, organisations, and events related to academic conferences. Currently, all main Semantic Web conferences and related events publish their data as Linked Data on SWDF, but for many other conferences, events and publication venues information is still not available in a structured and linked form. On the other hand the growth of available content with respect to the early times of SWDF poses data management issues and reveals design problems which where not foreseen when the dataset was at its initial stage.

There are several challenges to pursue the maintenance of a healthy and sustainable SWDF for the future: (i) the availability of appropriate vocabularies to express the current state of the data; (ii) the shared knowledge of such vocabularies; (iii) the availability of tools to ease the task of data acquisition, conversion, integration, augmentation, verification and finally publication; (iv) the ongoing maintenance of the dataset.

---

[1] SWDF: http://data.semanticweb.org.

In this work we focus on (i) and (ii) and use cLODg[2] (conference Linked Open Data generator) [4] to tackle (iii) (so far cLODg has been used to gather and publish metadata for ESWC2014 and ESWC2015 and it will be utilised for ESWC2016). As main contribution we identify issues of intentional nature on the SWDF dataset and we propose a refactoring solution for the Semantic Web Conference (SWC) ontology[3].

2   State of the Art

The first considerable effort to offer comprehensive semantic descriptions of conference events is represented by the metadata projects at ESWC 2006 and ISWC 2006 conferences [6], with the SWC Ontology being the vocabulary of choice to represent such data. Increasing number of initiatives are pursuing the publication about conferences data as Linked Data, mainly promoted by publishers such as Springer[4] or Elsevier[5] amongst many others. For example, the knowledge management of scholarly products is an emerging research area in the Semantic Web field known as Semantic Publishing [8]. Semantic Publishing aims at providing access to semantic enhanced scholarly products with the aim of enabling a variety of semantically oriented tasks, such as knowledge discovery, knowledge exploration and data integration. Despite these continuous efforts, it has been argued that lots of information about academic conferences is still missing or spread across several sources in a largely chaotic and non-structured way and a viable solution is a strong cooperation between researchers and publishers [1]. In this work we provide an analysis of existing modelling issues in the SWDF in order to provide a reference ontology and foster its adoption to close this gap.

3   Towards a Sustainable SWDF

The SWDF uses the SWC ontology as the reference ontology for modelling data about academic conferences. The SWC ontology combines existing widely accepted vocabularies (i.e., FOAF, SIOC and Dublin Core) and relies on the SWRC (Semantic Web for Research Communities) ontology for modelling entities of academic conferences, such as accepted papers, authors, their affiliations, talks and other events, the organizing committee and all other roles involved. Namely, the core types of SWC ontology are `foaf:Person` for describing people, `foaf:Organization` for describing the organisations (e.g., universities, research institutions, etc.) people are affiliated to, `foaf:Organization`, `swc:Artefact` for describing documents (e.g., papers, proceedings, etc.), `swc:OrganisedEvent` for describing any event related to an academic conference and `swc:Role` for describing the roles held by people at a conference. In the following we briefly list the main intentional issues affecting the data in the SWDF.

---

2 https://github.com/AnLiGentile/cLODg.
3 http://data.semanticweb.org/ns/swc/swc_2009-05-09.html.
4 http://lod.springer.com/wiki/bin/view/Linked+Open+Data/About.
5 http://data.elsevier.com/documentation/index.html.

Affiliations. Namely, the SWC ontology uses (i) the object property `swrc:affiliation` from the SWRC ontology in order to represent affiliations of people to organisations and (ii) the property `foaf:member` for expressing membership relations between organisations and people. Though this representation is very intuitive, it ignores the temporal dimension (i.e., the time when a given affiliation was held by an actor) that is relevant to interpret affiliations correctly. For example, it would be not possible to provide a correct answer to a simple competency question, such as "What was the affiliation of a person who authored a certain paper?".

Roles held by people at a conference. Roles such as program chair, track chair, etc. are model by using a commonly adopted ontology pattern based on the reification of a n-ary relation. Namely, the n-ary relation is identified by individuals of the class `swc:Role`. In fact, these individuals represents actual roles and relates people to events. The SWC ontology contains a very basic set of role classes (i.e., `swc:Chair`, `swc:Delegate`, `swc:Presenter` and `swc:ProgrammeCommitteeMember`) represented as sub-classes of `swc:Role`. This choice allows to instantiate the small set of different Role classes and cover the roles at specific events. For example, instead of sub-classing the `swc:Chair` class with MainChair, WorkshopChair, TutorialChair, etc., the different types of chairs should simply be instances of the generic Chair class and be labelled appropriately (e.g., `eswc2014:general-chair`[6]). The problem is that the individuals representing roles are defined locally to each conference. This means that for each conference, there is, for example, a different individual for representing the role "general chair". Hence, it is very complex to query the dataset in order to retrieve all the general chairs of the various editions of ESWC. More in detail, this comes from the erroneous reification of the n-ary relation on individuals of the class `swc:Role`, instead of using individuals of a different class representing the description of a role assignment situation.

We propose a refactoring of the SWC ontology, exploiting Ontology Design Patterns (ODP) [2]. We choose the Time indexed person role[7] ontology design pattern for modelling affiliations and assignment of roles to people. It is based on the reification of a n-ary relation, whose individuals are instances of the class `timeindexedpersonrole:TimeIndexedPersonRole`, representing the fact that a certain situation (e.g., a person affiliated to an organisation) occurs at a certain time interval. The classes `conf:Affiliation` and `conf:RoleAssignments` are defined as specializations of the class `timeindexedpersonrole:TimeIndexedSituation`[8]. According to this representation we are able to represent affiliations as n-ary relations having (i) a person (i.e., the agent holding the affiliation), (ii) an organisation and (ii) an associated time. The time can be an interval consisting of the conference dates or the instant when the paper was submitted. This allows us

---

[6] The prefix `eswc2014:` stands for the namespace http://data.semanticweb.org/conference/eswc/2014/.

[7] http://ontologydesignpatterns.org/wiki/Submissions:Time_indexed_person_role.

[8] http://ontologydesignpatterns.org/wiki/Submissions:TimeIndexedSituation.

to represent cases in which a person moves to another organisation in the time interval between the paper submission and the conference event.

Additionally, we aligned the ontology to (i) Dolce D0[9] classes that define the top level of the ontology and to (ii) FaBIO [7]. The defined ontology is available on-line for download[10] and the final data generation is performed using the cLODg Open Source workflow that uses our proposed refactored data model.

4   Conclusions and Future Work

This paper identifies current issues in the SWC ontology and proposes a refactoring solution. Given the new proposed data model we regenerate a subset of the SDWF dataset using the cLODg workflow [3] (the current cleaned dataset is provided on http://www.scholarlydata.org/). As future work we plan to collaborate for the cleansing of the official SWDF on http://data.semanticweb.org/ and to add more advanced alternative implementations for each step of cLODg and provide data maintenance services.

References

1. Bryl, V., Birukou, A., Eckert, K., Kessler, M.: What is in the proceedings? Combining publishers and researchers perspectives. In: Proceedings of SePublica 2014, Anissaras, 25 May 2014
2. Gangemi, A., Presutti, V.: Ontology design patterns. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies, 2nd edn, pp. 221–243. Springer, Heidelberg (2009). doi:10.1007/978-3-540-92673-3_10
3. Gentile, A.L., Acosta, M., Costabello, L., Nuzzolese, A.G., Presutti, V., Recupero, D.R., Live, C.: Accessible and sociable conference semantic data. In: Proceedings of WWW 2015 (Companion Volume), pp. 1007–1012. ACM (2015)
4. Gentile, A.L., Nuzzolese, A.G.: cLODg - conference linked open data generator. In: Proceedings of the ISWC 2015 Posters & Demonstrations Track co-located with the 14th International Semantic Web Conference (ISWC-2015), Bethlehem, 11 October 2015
5. Harrison, W.: Eating your own dog food. Industrial, Organizational Psychology, 5–7 June 2011
6. Möller, K., Heath, T., Handschuh, S., Domingue, J.: Recipes for semantic web dog food — the ESWC and ISWC metadata projects. In: Aberer, K., et al. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, pp. 802–815. Springer, Heidelberg (2007). doi:10.1007/978-3-540-76298-0_58
7. Peroni, S., Shotton, D.: FaBiO and CiTO: ontologies for describing bibliographic resources and citations. J. Web Semant. **17**, 33–43 (2012)
8. Shotton, D.: Semantic publishing: the coming revolution in scientific journal publishing. Learn. Publ. **22**(2), 85–94 (2009)

---

9 http://www.ontologydesignpatterns.org/ont/dul/d0.owl.
10 http://www.scholarlydata.org/ontology/conference-ontology.owl.