

A Collaborative Framework for Managing and Publishing KOS

Armando Stellato⁺, Ahsan Morshed^{*}, Gudrun Johannsen^{*}, Yves Jaques^{*},
Caterina Caracciolo^{*}, Sachit Rajbhandari^{*}, Imma Subirats^{*}, Johannes Keizer^{*}

^{*}The Food and Agricultural Organization of UN (FAO)
Viale delle Terme di Caracalla, 00153 Rome, Italy
{name.surname@fao.org}

⁺University of Rome, Tor Vergata
Via del Politecnico 1, 00133 Rome, Italy
stellato@info.uniroma2.it

Abstract. In the Food and Agriculture Organization of the United Nations (FAO), the need to revamp its popular agriculture vocabulary AGROVOC using Semantic Web knowledge representation standards combined with the need to provide a collaborative environment for development and maintenance purposes, pushed forward the realization of a dedicated AGROVOC thesaurus maintenance tool. With the progressive standardization of the AGROVOC knowledge model, following recent Simple Knowledge Organization System (SKOS) recommendations by the World Wide Web Consortium (W3C) and with the addition of more FAO-maintained vocabularies, the former “AGROVOC Concept Server Workbench” has become a general-purpose framework for thesauri and vocabulary development and is now reborn as “VocBench”. In this paper, we describe the path which led to its realization and its main features

1. Introduction

AGROVOC is a multilingual structured thesaurus for agriculture, owned and maintained by an international community of agricultural research information institutions and managed by the Food and Agriculture Organization of the United Nations (FAO). AGROVOC is used world-wide by researchers, librarians, information managers and others, for indexing, retrieving, and organizing data in agricultural information systems. Maintained since the 1970's, it currently includes 579523 terms in 19 different languages.

Due to the growth of the Semantic Web AGROVOC underwent a process of semantic refinement to shape it into a reusable Resource Description Framework (RDF) vocabulary that could fulfill its traditional functions of document indexing while also meeting the needs of a new generation of semantically-enabled applications. FAO made a first port of AGROVOC [1] using a customized model based on Ontology Web Language (OWL). The model was tailored to represent concept schemes and

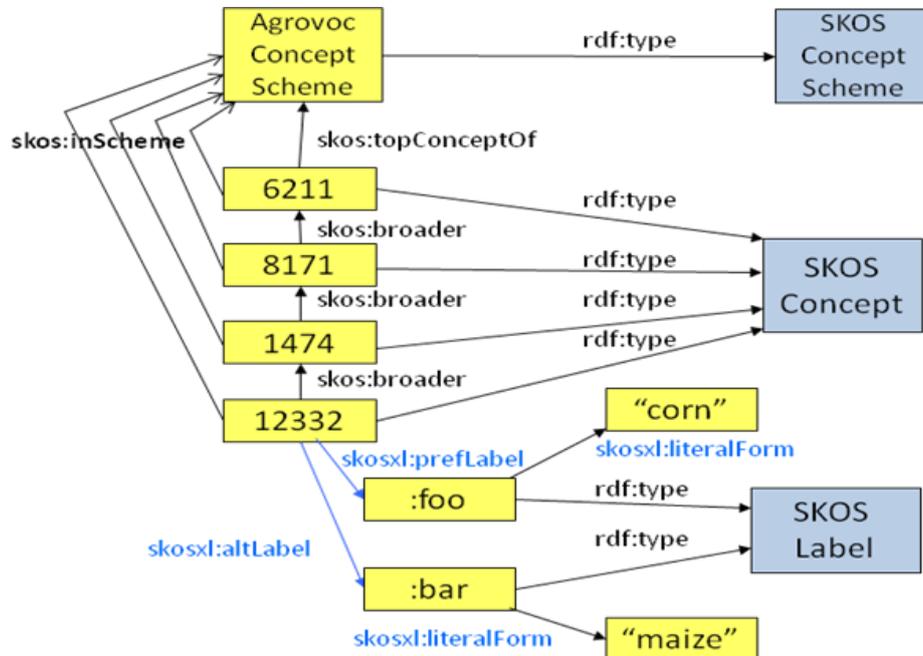


Figure 1 - AGROVOC SKOS Model

provide metadata descriptors about edited resources. It also put emphasis on multilinguality and the natural language description of resources. Such a customized model was difficult to manage using traditional ontology editing tools such as Protégé [2,3] (e.g. reified labels represent different resources to be edited, while users expect them to be editable inside a view of the concept to which they share attachment). The poor usability of traditional ontology editing tools combined with the need for a collaborative environment that supported roles-based authentication, editorial workflow, multilingual search and high installability (or, even better, no installation at all) led to the development of CS Workbench, a web application specifically tailored to the AGROVOC Vocabulary.

2. From AGROVOC Workbench to FAO VocBench

Following this first re-modeling of AGROVOC, new knowledge representation languages emerged for the web that were expressly intended to represent thesauri and concept schemes. In 2009, the W3C approved Simple Knowledge Organization System SKOS [4] as a recommendation. SKOS, together with the SKOS-XL [5] model (SKOS Extension for Labels), were a good fit to the modeling requirements of

AGROVOC (see Figure 1). Moreover, the inclusion of additional vocabularies¹ in CS Workbench highlighted the need to rally around a standard model such as SKOS. In 2010, the decision was made to transform CS Workbench into a general-purpose, SKOS-compliant platform for collaborative knowledge management. CS Workbench thus gave way to VocBench, a collaborative editing framework for thesauri and concept schemes. The current version, VocBench 1.1, still relies internally to some extent on the customized OWL knowledge model though it offers facilities for a clean SKOS/SKOSXL export of its content. The next major release (2.0) will feature a native interface for SKOS and SKOSXL, based on the OWL ART API² abstraction layer and SKOSXL interfaces, and configurable support for a host of different triple store technologies. Support for generic OWL ontologies is also on the roadmap, considering the necessity to model and export as linked open data many FAO repositories such as the FAO Journal Data records (JAD), best modeled through commonly adopted ontology vocabularies for bibliographic recording (such as the Bibliographic Ontology [6]).

2.1. AGROVOC Linked Open Data (AGROVOC LOD)

Every AGROVOC concept is identified by a unique, web-accessible Uniform Resource Identifier (URI). AGROVOC is published as linked open data (LOD) using human-readable web pages and also using RDF representations in XML, and Turtle. To fulfill the LOD vision of interconnected, standardized, machine-readable linked resources, AGROVOC is converted from its native OWL ontology to SKOS-XL. We have also aligned seven thesauri (i.e. EUROVOC, GEMET, LCSH, NAL, STW, RAMAEU, and TheSoz) with AGROVOC, generating 18,800 terminologist-verified out-links and 2000 in-links (GEMET and RAMAEU). The syntactic matching approach was used and all exact matches were evaluated by a domain expert.

3. Conclusion

We have presented FAO VocBench, an editing and management framework for thesauri and concept schemes, which thanks to its emphasis on collaborative workflow and its adherence to W3C standards is unique among open-source tools for thesauri development.

Other than its native compliance to SKOS (and SKOSXL), the next major version of VocBench will provide support for linking external resources or aligning their concepts, possibly by dynamically accessing them through their LOD endpoints and presenting their content in a uniform and easy way for the user. Currently the

¹ *Biotech Glossary*: it is a flat (not hierarchically organized) vocabulary with concepts and relation between them. It currently contains descriptors for a total of 3000 concepts about biotechnologies

² <http://art.uniroma2.it/owlart/>

linking/mapping process is performed manually or with support from automatic matching tools embedded in the Eclipse integrated development environment.

FAO VocBench is available at: <http://aims.fao.org/website/The-VocBench/>
while the LOD dataset is available at: <http://aims.fao.org/website/Linked-Open-Data/>

References

1. Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., Katz, S.: Reengineering Thesauri for New Applications: The AGROVOC Example. *Journal of Digital Information - JODI* 4 (2004)
2. Gennari, J., Musen, M., Fergerson, R., Grosso, W., Crubézy, M., Eriksson, H., Noy, N., Tu, S.: The evolution of Protégé-2000: An environment for knowledge-based systems development. *International Journal of Human-Computer Studies* 58(1), 89–123 (2003) Protege.
3. Knublauch, H., Fergerson, R., Friedman Noy, N., Musen, M.: The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. In : *Third International Semantic Web Conference - ISWC 2004, Hiroshima, Japan (2004)*
4. W3C: SKOS Simple Knowledge Organization System Reference. In: *World Wide Web Consortium (W3C)*. (Accessed August 18, 2009) Available at: <http://www.w3.org/TR/skos-reference/>
5. W3C: SKOS Simple Knowledge Organization System eXtension for Labels (SKOS-XL). In: *World Wide Web Consortium (W3C)*. (Accessed August 18, 2009) Available at: <http://www.w3.org/TR/skos-reference/skos-xl.html>
6. D'Arcus, B., Giasson, F.: *Bibliographic Ontology Specification*. (Accessed 2008) Available at: <http://bibliontology.com/>
7. Baker, T., Keizer, J.: *Linked Data for Fighting Global Hunger: Experiences in setting standards for Agricultural Information Management*. In Wood, D., ed. : *Linking Enterprise Data*. Springer US, Washington 177-201 10.1007/978-1-4419-7665-9_9.