Collaborative Development of Multilingual Thesauri with VocBench (System Description and Demonstrator)

Armando Stellato[1], Sachit Rajbhandari[2], Andrea Turbati[1], Manuel Fiorelli[1],
Caterina Caracciolo[2], Tiziano Lorenzetti[1], Johannes Keizer[2], Maria Teresa Pazienza[1]

[1]ART Group, Dept. of Enterprise Engineering
University of Rome, Tor Vergata
Via del Politecnico 1, 00133 Rome, Italy
{stellato,turbati,fiorelli,pazienza,lorenzetti}@info.uniroma2.it
[2]Food and Agriculture Organization of the United Nations (FAO)
Viale delle Terme di Caracalla, 00153 Rome, Italy
{sachit.rajbhandari,caterina.caracciolo,johannes.keizer}@fao.org

Abstract. VocBench is an open source web application for editing of SKOS and SKOS-XL thesauri, with a strong focus on collaboration, supported by workflow management for content validation and publication. Dedicated user roles provide a clean separation of competences, addressing different specificities ranging from management aspects to vertical competences on content editing, such as conceptualization versus terminology editing. Extensive support for scheme management allows editors to fully exploit the possibilities of the SKOS model, as well as to fulfill its integrity constraints. We describe here the main features of VocBench, which will be shown along the demo held at the ESWC15 conference.

Keywords: Collaborative Thesaurus Management, SKOS, SKOS-XL

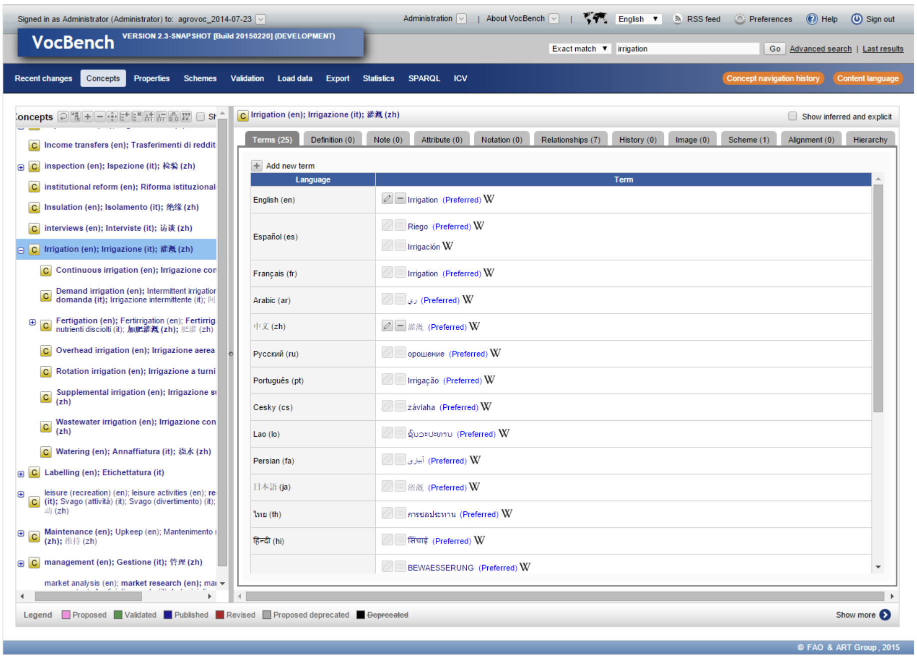
1 Introduction

In 2008, the AIMS group of the Food and Agriculture Organization of the United Nations (FAO, http://www.fao.org/) fostered the development of a collaborative platform for managing the Agrovoc thesaurus [1]: the "Agrovoc Workbench". Later on, in the context of a joint collaboration between FAO and the ART group of the University of Tor Vergata in Rome (http://art.uniroma2.it), the system has been completely rethought as a fully-fledged collaborative platform for thesaurus management, available free of charge and open source: VocBench. With respect to its predecessor, VocBench complies with standard Semantic Web technologies, by relying on Semantic Turkey [2], an RDF management platform already developed and currently maintained by the ART team. In particular, VocBench natively supports the SKOS[1] W3C vocabulary for representing thesauri and concept schemes, with its extension SKOS-XL[2] for extended labels (i.e. labels reified as RDF resources, which can be described in turn).

[1] http://www.w3.org/TR/skos-reference/

[2] http://www.w3.org/TR/skos-reference/skos-xl.html

adfa, p. 1, 2011.



- Income transfers (en); Trasferimenti di reddit
- inspection (en); Ispezione (it); 检验 (zh)
- institutional reform (en); Riforma istituzional
- Insulation (en); Isolamento (it); 绝缘 (zh)
- interviews (en); Interviste (it); 访谈 (zh)
- Irrigation (en); Irrigazione (it); 灌溉 (zh)
  - Continuous irrigation (en); Irrigazione con
  - Demand irrigation (en); Intermittent irrigation domanda (it); Irrigazione intermittente (it); 间
  - Fertigation (en); Fertirrigation (en); Fertirrig nutrienti disciolti (it); 加肥灌溉 (zh); 肥灌 (zh)
  - Overhead irrigation (en); Irrigazione aerea
  - Rotation irrigation (en); Irrigazione a turni
  - Supplemental irrigation (en); Irrigazione su (zh)
  - Wastewater irrigation (en); Irrigazione con (zh)
  - Watering (en); Annaffiatura (it); 浇水 (zh)
- Labelling (en); Etichettatura (it)
- leisure (recreation) (en); leisure activities (en); re (it); Svago (attività) (it); Svago (divertimento) (it); 动 (zh)
- Maintenance (en); Upkeep (en); Mantenimento (zh); 维持 (zh)
- management (en); Gestione (it); 管理 (zh)
- market analysis (en); market research (en); mar

Irrigation (en); Irrigazione (it); 灌溉 (zh)

Language	Term
English (en)	Irrigation (Preferred) W
Español (es)	Riego (Preferred) W Irrigación W
Français (fr)	Irrigation (Preferred) W
Arabic (ar)	ري (Preferred) W
中文 (zh)	灌溉 (Preferred) W
Русский (ru)	орошение (Preferred) W
Português (pt)	Irrigação (Preferred) W
Cesky (cs)	závlaha (Preferred) W
Lao (lo)	ຊົນລະປະທານ (Preferred) W
Persian (fa)	آبیاری (Preferred) W
日本語 (ja)	灌漑 (Preferred) W
ไทย (th)	การชลประทาน (Preferred) W
हिन्दी (hi)	सिंचाई (Preferred) W
	BEWAESSERUNG (Preferred) W

Figure 1. VocBench User interface showing a fragment of the AGROVOC thesaurus

While providing a more thorough support for RDF, VocBench retains the focus on multilingualism, collaboration and on a structured content validation & publication workflow that characterized it yet from its infancy. The demo will provide a guided tour through all of VocBench features and will let the user experience the editorial process that accompanies the development of an authoritative resource.

2 A Quick Glance at VocBench Features

The feedback gathered from real thesaurus publishers guided the development of VocBench: FAO and its partners provided great support for shaping interaction and collaboration capabilities. Here follow the features that mostly characterize the system.

User Interface. VocBench has been conceived as a web application accessible through any modern browser, therefore disburdening end users from software installation and configuration. The user interface consists of multiple tabs, each one associated with specific information and functionalities. A quick exploration of the available tabs is sufficient to discover most of the VocBench functionalities. Figure 1 offers a typical view of VocBench, with the concept tree on the left, and the description of the selected concept on the right, centered on the term tab, listing all terms in the different languages available for the resource. Concepts in the tree may be shown through their labels in all of the selected languages for visualization. An option toggles between a view of preferred labels only, and all labels. The multilingual characteristics of VocBench are not limited to content management, as its interface is also localized in different languages, currently: English, Spanish, Dutch and Thai.

Role-based Access Control. VocBench promotes the separation of responsibilities through a role-based access control mechanism, checking user privileges for requested functionalities through *roles* that users assume. A completely customizable access policy specifies roles and their assigned privileges. New roles can be created and existing ones can be modified. The default policy recognizes typical roles and their acknowledged responsibilities: *Administrators, Ontology editors, Term editors* (Terminologists), *Validators* and *Publishers*.

Formal Workflow and Recent Changes. Collaboration is essential for distributing effort and reaching consensus on the thesaurus being developed. To facilitate collaboration, VocBench provides an editorial workflow in which editors' changes are tracked and stored for approval by content validators. This workflow management is supported by role-based access control, by providing users with different roles so to enforce the separation between their responsibilities. In a collaborative environment, where users may proactively edit a shared resource, it is important to have means for monitoring the situation. Regarding this aspect, the ability to control recent changes to the thesaurus is useful for detecting hot sections and coordinating with other editors. In VocBench, users can see recent changes both in the Web user interface and as an RSS feed.

Advanced Scheme Management. VocBench allows to manage thesauri organized around multiple concept schemes. Users can switch across schemes by selecting them through the relevant *Schemes* tab in the user interface. VocBench functionalities are well-behaved with respect to schemes, as actions that would generate *dangling* concepts (concepts not reachable through any tree-view) are forbidden, detailing the cause of the impediment to the users. In any case, since data can be loaded from pre-existing sources developed outside of VocBench, a fixing utility for dangling concept is available through the UI, and will be part of a larger section dedicated to Integrity Constraints Validation, especially thought for fixing violated SKOS constraints.

Metrics & SPARQL Querying. VocBench reports several metrics concerning the thesaurus itself and the collaborative workflow. In addition to statistics and visualizations provided by VocBench, users may formulate SPARQL 1.1 queries/updates to select precise information, perform custom analytical tasks or modify the thesaurus bypassing the standard editing functionalities. The SPARQL editor is based on the open source Flint SPARQL Editor (`https://github.com/TSO-Openup/FlintSparqlEditor`), which provides syntax highlighting and completion, and has been customized to be fed with information from the edited thesaurus.

Alignment. From version 2.3, VocBench supports alignments to other thesauri. Currently, the creation of alignments can either be performed manually, by inserting URIs as values of the various SKOS mapping properties, or be assisted in case of mappings to other thesauri managed by the same instance of VocBench. In the latter case, a concept-tree browser with advanced search interfaces facilitates the identification of the best matching concepts from the targeted datasets.

3 Some Notes on Architecture and Technologies

Semantic Turkey, the RDF backbone of VocBench, offers an OSGi service-based layer for designing and developing OWL ontologies and SKOS(XL) thesauri. A lightweight Firefox interface is available for use as a desktop tool, which is now complemented by VocBench, mainly differentiating for its collaborative nature and its focus on thesauri.

VocBench has a layered architecture consisting of a presentation and multi-user management layer, a service layer and a data management layer. The first layer is implemented as a Web application, powered by GWT (Google Web Toolkit, `http://www.gwtproject.org/`). The other layers coincide with the Semantic Turkey RDF management platform, equipped with an extension providing additional services expressly developed for VocBench. VocBench is also in charge of user and workflow management, since these aspects are not covered by Semantic Turkey. User accounts and tracked changes are stored in a relational database accessed through a JDBC connector. The adoption of OSGi allows for plugging of extensions: in particular, other than realizing additional services, different connectors for specific RDF middleware and triple storage technologies can be provided. VocBench is currently shipped with a connector for Sesame2 [3], supporting all of its storage/connection possibilities: in memory, native, remote connection and their respective configurations. The remote connection is particularly useful, as it allows VocBench to connect to Sesame2 compliant triple stores (e.g. GraphDB [4]) without need for a dedicated connector. VocBench RDF API are based on OWL ART (`http://art.uniroma2.it/owlart/`), an abstraction layer supporting access to different triple stores. Different connectors can be implemented from scratch in terms of those API, or by reusing middleware already bridged through other existing connectors. For instance, the Virtuoso triplestore [5] is compatible with the Sesame API, but requires a dedicated client library: it thus needs to be introduced by a specific connector, though its implementation may be largely realized as an extension of the already existing Sesame connector. Finally, particular attention has been paid to system scalability, both on performance and maintenance aspects. To this end, information is provided to the frontend as much as possible in an incremental fashion (e.g. each level of the concept hierarchy, as nodes are expanded).

4 System Demo

In the demonstration, visitors will be guided through all of VocBench features, experiencing the editorial process that accompanies the development of an authoritative resource. The audience will initially be acquainted with the UI of the environment and learn how to browse the loaded dataset in order to explore its content. Later on, they will try the more common editing operations for creating, modifying and relating concepts and (SKOS-XL reified) labels. Interested people will go through the full editorial workflow, seeing how different roles will contribute to the evolution of the thesaurus.

The demo will be carried on real thesauri from a few of the large organizations that are already using VocBench for maintaining their resources. These thesauri include:

- Agrovoc (Food and Agriculture Organization)

- Eurovoc (EU Documentation Office)
- Unified Astronomy Thesaurus (Harvard-Smithsonian Center for Astrophysics)
- Teseo (Italian Senate)

5 More about VocBench

This paper accompanies the demo of VocBench being held at the 12th Extended Semantic Web Conference. More information about VocBench, an in-depth comparison with other systems, user evaluation, lessons learned and insights on the future of the system, can be found in [6], an article presented at the Research Track of this same conference.

5.1 Availability

VocBench is distributed as open-source under the Mozilla Public License (`https://www.mozilla.org/MPL/2.0/`).

VocBench home page: `http://vocbench.uniroma2.it/`

Source code on Bitbucket: `https://bitbucket.org/art-uniroma2/vocbench`

A sandbox server for testing VocBench capabilities is hosted by courtesy of the Malaysian research center MIMOS Berhad at: `http://202.73.13.50:55481/vocbench/`

References

1. Caracciolo, C., Stellato, A., Morshed, A., Johannsen, G., Rajbhandari, S., Jaques, Y., Keizer, J.: The AGROVOC Linked Dataset. Semantic Web Journal 4(3), 341–348 (2013)
2. Pazienza, M.T., Scarpato, N., Stellato, A., Turbati, A.: Semantic Turkey: A Browser-Integrated Environment for Knowledge Acquisition and Management. Semantic Web Journal 3(3), 279-292 (2012)
3. Broekstra, J., Kampman, A., van Harmelen, F.: Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema. In : The Semantic Web - ISWC 2002: First International Semantic Web Conference, Sardinia, Italy, pp.54-68 (2002) June 9-12.
4. Kiryakov, A., Ognyanov, D., Manov, D.: OWLIM – a Pragmatic Semantic Repository for OWL. In : Int. Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2005), WISE 2005, New York City, USA (2005) 20 November.
5. Erling, O., Mikhailov, I.: RDF Support in the Virtuoso DBMS. In Pellegrini, T., Auer, S., Tochterman, K., Schaffert, S., eds. : Networked Knowledge - Networked Media, in Studies in Computational Intelligence 221. Springer Berlin Heidelberg (2009) 7-24
6. Stellato, A., Rajbhandari, S., Turbati, A., Fiorelli, M., Caracciolo, C., Lorenzetti, T., Keizer, J., Pazienza, M.T.: VocBench: a Web Application for Collaborative Development of Multilingual Thesauri. In : The Semantic Web. Latest Advances and New Domains:12th Extended Semantic Web Conference, ESWC 2015, Portoroz, Slovenia, 31 May - 4 June 2015. Springer International Publishing (2015) (accepted for publication).