Available online at www.sciencedirect.com
ScienceDirect
Journal of Integrative Agriculture
2012, 11(5): 694-699
May 2012

REVIEW

The AGROVOC Concept Scheme – A Walkthrough

Sachit Rajbhandari and Johannes Keizer

Food and Agriculture Organization of the United Nations, Rome 00153, Italy

Abstract

The Food and Agriculture Organization is developing a concept based multilingual vocabulary management tool to manage thesauri, authority lists and glossaries expressed as concept schemes ready to be used in a linked data environment. In this paper, we described the evolution of the AGROVOC thesaurus to AGROVOC Concept Scheme based on OWL (web ontology language) model and now shifting to SKOS (simple knowledge organization system) model. The paper explained why and how it evolved highlighting the key differences between different models. The system architecture and significant set of features available in the VocBench was discussed in the paper.

Key words: semantic web, ontologies, thesauri, OWL, SKOS, AGROVOC

INTRODUCTION – AGROVOC

AGROVOC is a multilingual structured thesaurus created by FAO and the Commission of the European Communities since 1980 covering the fields of food, agriculture, forestry, fisheries, and other related domains. In simple words, it consists of words or expressions (terms) in multiple languages and organized using equivalence (USE/UF), broader term (BT), narrower term (NT), and related term (RT) relationships. AGROVOC, available in 20 different languages is used worldwide to improve indexing or search information sources in their own language.

In thesaurus, one term could be associated with other terms using RT. However, it would be more meaningful if we could specify explicit relationships between the terms rather than simply portraying them as related terms. So there was a need of an extended set of relationships to perform a more granular and consistent indexing for effective searching and browsing. This inspired us for the transition of traditional thesaurus to the AGROVOC Concept Scheme (Dagobert *et al*. 2004).

THE AGROVOC CONCEPT SCHEME

The AGROVOC Concept Scheme was known as the Agricultural Ontology Service Concept Server (AOS/CS) in 2006 (Fisseha *et al*. 2001; Liang *et al*. 2006). The word AGROVOC was very common within the community and was like a brand name, so later it was named as the AGROVOC Concept Server. However, as the use of the word server created a confusion among users making them think it as a physical server machine, it has been renamed and also in the paper here after is addressed as the AGROVOC Concept Scheme.

The AGROVOC Concept Scheme is a semantically structured concept-based system consisting of concepts with their lexical representation, specific relationships between concepts, and also relationships between

Received 28 June, 2011 Accepted 30 September, 2011
Correspondence Sachit Rajbhandari, E-mail: sachit.rajbhandari@fao.org

their multilingual lexicalizations. The Concept Scheme is a one-stop shop for terminological knowledge facilitating its use for developing agricultural domain knowledge organization systems, including ontologies using semantic technologies.

The concept scheme has a knowledge base of around 30 000 concepts with 600 000 terms organized in ontological relationships (hierarchical, associative, equivalence). These concepts have been obtained by remodeling the traditional AGROVOC thesaurus. The AGROVOC Concept Scheme has been expressed in three different levels of representation (Liang *et al*. 2006): (1) Concept is the abstract meaning given to the group of the terms, e.g., rice in the sense of plant; (2) terms are the language specific lexical form of that concept, e.g., rice in English, arroz in Spanish, or riz in French; (3) term variants are the range of forms that can occur for each term, e.g., organization in American English or organisation in British English.

Alike in traditional thesaurus, now the concepts build the actual hierarchy and semantic structure of the ontology. Terms are no longer arranged in a hierarchy or related *via* semantic relationships, as is currently done in AGROVOC. Each term is a separate entity in every language that can be linked to concepts, to other terms and their variants. Additionally, there are lexical relationships like lexical equivalences (e.g., translations, synonyms) that should be captured in an accurate domain model.

THE AGROVOC OWL MODEL

The AGROVOC Meta-model (Liang *et al*. 2006) is based on OWL (Dean *et al*. 2004; McGuinness and Harmelen 2004) with three concepts at the top level: category, classification scheme, lexicalization. The domain concept is the sub class of category concept, which is the root of all domain concepts creating the hierarchical structure of the Concept Scheme. The concepts that are not domain concepts but belong to different classification schemes come directly under the category concept and are linked with the classification scheme class representing different classification schemes like AGRIS/CARIS classification.

Each term in a different language that describes the concept is modeled as an instance of the lexicalization concept. And the lexicalization instance is linked with concept instance using “has lexicalization” relationship. All the AGROVOC terms from the traditional thesaurus are modeled as lexicalization instance with all descriptors as preferred terms and non-descriptor as alternative terms.

Relationships that can be specified between concepts or terms (lexicalizations) are modeled as OWL object properties. And the relationships that relate concepts or terms to their variants are modeled as OWL data type properties (Liang *et al*. 2006) (Table and Fig. 1).

THE AGROVOC SKOS MODEL

FAO has to create its own unique AGROVOC OWL model, as at that time there was no standard model available capable of expressing the traditional thesaurus to concept scheme. SKOS (simple knowledge organization system) was in the early draft phase where concepts were described by literals without the ability to express relations between terms (Miles and Bechhofer 2009).

SKOS after its final release provided vocabulary for expressing the traditional thesaurus to AGROVOC Concept Scheme. Further SKOS-XL (SKOS eXtension for Labels) added the possibilities to express labels as first-class resources and interrelate with different relations (Antoine *et al*. 2009; Miles and Bechhofer 2009). SKOS has also become de-facto standard for sharing and linking knowledge organization systems as linked data. One more problem with OWL model is that every concept has to be conceptualized as instances, which was difficult to explain both in theory and practice. But in SKOS, every concept is by definition an instance of the class SKOS Concept. All the above facts encourage shifting AGROVOC from OWL to SKOS models.

In the new SKOS model, AGROVOC Concept Scheme is itself an instance of the class SKOS Concept Scheme, all terms are modeled as instances of the class SKOS label (classes lexicalization and noun replaced SKOS

Table Example on semantic relationships

Concept to concept	Superclass/subclass, causes, is caused by
Concept to term	hasLexicalization (links concepts to their lexical realizations)
Term to term	isSynonymOf, hasAcronym, hasAbbreviation
Concept/term to variants	hasStatus, hasSpellingVariant, hasSingular

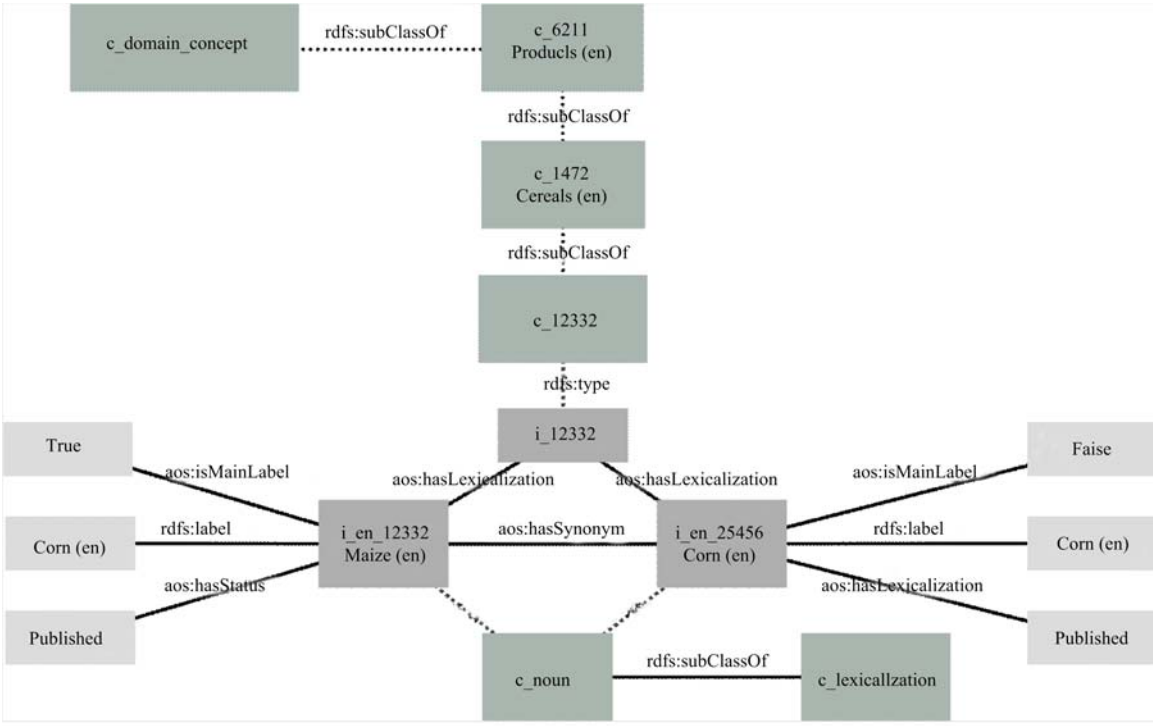

Fig. 1 Meta model for the AGROVOC ontology (2006).

labels), and concepts as instances of the class SKOS concept (class/instance pairs converted into SKOS concepts). An ontologically strong sub-class relationship between concepts is replaced by ontologically weaker broader and narrower relationships. Whenever needed, SKOS concepts may also be upgraded to OWL classes, with additional constraints, for use in local ontologies (Baker and Keizer 2010).

All the possible properties defined uniquely in OWL model are replaced by commonly used standard vocabulary (e.g., aos:c_image by foaf:Image, aos:hasDateCreated by dc:date, aos:hasDefinition by skos:definition) (Fig. 2).

THE VOCBENCH

The VocBench (also known as the Workbench, AOS/CS Workbench, AGROVOC Concept Server Workbench, Agricultural Ontology Server, or ACSW) is a web-based multilingual vocabulary management tool. It is relabeled with generic name VocBench, as this tool has been used to not only to transform AGROVOC but also other thesauri, authority lists and glossaries into SKOS/RDF concept schemes. It facilitates as a collaborative tool allowing experts to add or edit multilingual terminology and semantic concept information. Other outstanding feature of the VocBench is a built in workflow for the maintenance, validation, and consistency checks and further provides user and group administration (Suktarachan 2008; Yongyuth 2008; Sini 2010) (Fig. 3).

Concept management

Vocbench allows to create the concept with unique URI and can be positioned in the concept hierarchy. For each concept created, user can update more information within following components:

(1) Terms are the labels related to the concept in multiple languages. Vocbench allows adding relationships between the terms and specify different attributes for each term;

(2) definition is the meaning of the concept added in any language especially expressed in the technical terms;

(3) notes can be either scope or editorial notes to share more information related to the concept with other users;

(4) attributes are used to better describe the charac-

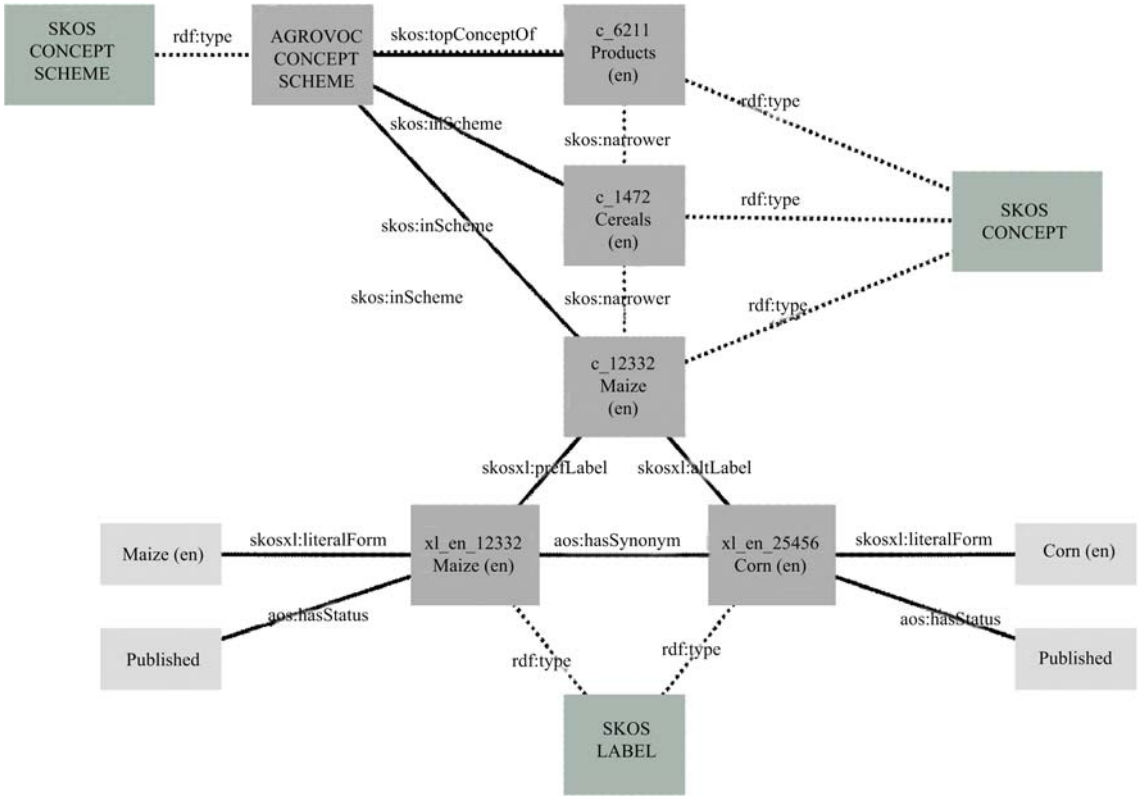

Fig. 2 Meta model for the SKOS ontology.

teristics of the concept;

(5) relationship shows the relation of the concept with other concepts in the hierarchy;

(6) history to show basic information like date created, date updated, status, and all the past actions of the selected concept in chronological order;

(7) image allows linking of the pictures to visualize the concept.

Relationship management

The VocBench data model is kept in OWL format, which is in triple pattern (subject-predicate-object). This module allows users to add, edit, or delete the predicate used in the model. The relationship can be either object or data type property and has following components to provide related information:

(1) Label in multiple languages can be assigned to identify the relationship;

(2) definition is the meaning given to the relationships in any language for describing its usage;

(3) property can be the characteristic of the relationship such as symmetric, transitive, or functional;

(4) inverse property shows the inverse relationship if exists for the selected relationship;

(5) domain & range specifies the boundary of the subject and object of the predicate.

Search

This function searches for the concept and is available as basic and advance search:

(1) Basic search is accessible from any module for quick search of the concept with key word filters (e.g., contains, exact match, or starts with);

(2) advance search allows making more accurate result by using the filters on relationships, languages, classification schemes, or concept status.

Classification management

Classification management helps to classify the concepts based on different schemes.

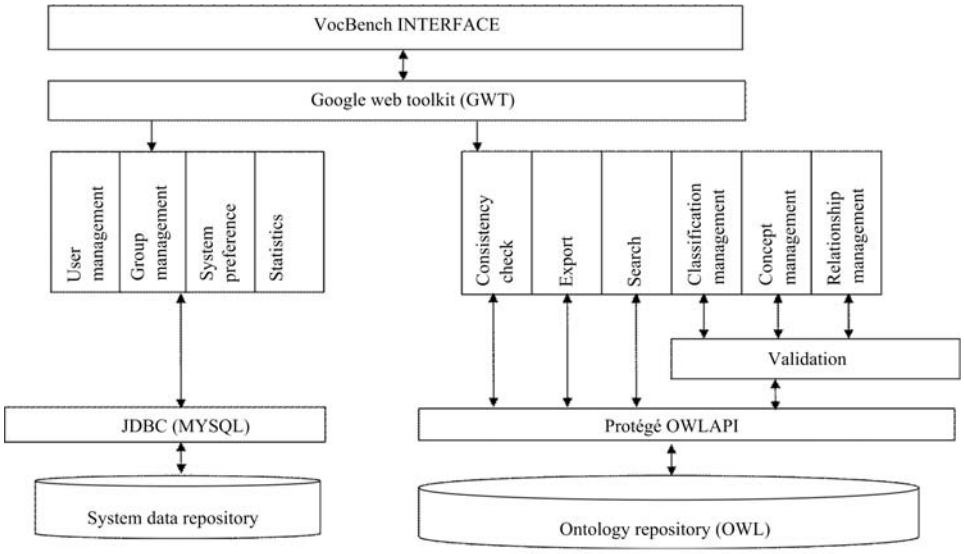

Fig. 3 System architecture of the VocBench.

Validation

Validation allows experts to verify and validate newly updated data before publishing them.

Consistency checks

Consistency check allow checking the consistency of the data model.

Export

Export allows exporting of the ontology in different formats (currently supported ones are SKOS, SQL, Text).

Users/Group management

Users/Group management helps to assign users to different groups with different permissions.

Statistics

Statistics gives the statistical preview on count of concepts, terms, and relationships, frequency of system used, total number of users registered users, and exports carried out.

The system is available at http://aims.fao.org/website/VocBench/sub

FURTHER WORKS

Current released version of VocBench is based on OWL model and has above discussed features and functionalities. There is an on going development for the newer version, which will be based on native SKOS Meta model. The conversion from OWL to SKOS models for the AGROVOC dataset has been completed and has been exported as linked open data (LOD), which can be accessed through the SPARQL (Prud'hommeaux and Seaborne 2008) end point.

References

Antoine I, Phipps J, Rubin D. 2009. SKOS use cases and requirements. [2009-8-18]. http://www.w3.org/TR/skos-ucr/#UC-Aims

Baker T, Keizer J. 2010. Linked data for fighting global hunger: experiences in setting standards for agricultural information management. In: Wood D, ed., *Linking Enterprise Data*. Springer US, Washington, D.C. pp. 177-201.

Dagobert S, Lauser B, Liang A, Fisseha F, Keizer J, Katz S. 2004. Reengineering thesauri for new applications: the AGROVOC example. *Journal of Digital Information*, **4**, 1-23.

Dean M, Schreiber G, Bechhofer S, Harmelen F V, Hendler J, Horrocks I, McGuinness D L, Patel-Schneider P F, Stein L A. 2004. OWL web ontology language reference.

[2004-2-10]. http://www.w3.org/TR/owl-ref

Fisseha F, Hagedorn K, Keizer J, Katz S. 2001. Creating the semantic web: the role of an agricultural ontology server (AOS). In: *International Semantic Web Workshop (Infrastructure and Applications for the Semantic Web)*. Stanford, USA. pp. 1-2.

Liang A C, Lauser B, Sini M, Keizer J, Katz S. 2006. From AGROVOC to the agricultural ontology service/concept server an OWL model for managing ontologies in the agricultural domain. *Language*, **466**, 373-377.

McGuinness D L, Harmelen F V. 2004. OWL web ontology language overview. [2004-2-10]. http://www.w3.org/TR/owl-features

Miles A, Bechhofer S. 2009. SKOS simple knowledge organization system reference. [2009-8-18]. http://www.w3.org/TR/skos-reference

Miles A, Bechhofer S. 2009. SKOS simple knowledge organization system eXtension for labels (SKOS-XL). [2009-8-18]. http://www.w3.org/TR/skos-reference/skos-xl.html

Prud'hommeaux E, Seaborne A. 2008. SPARQL query language for RDF. [2008-1-15]. http://www.w3.org/TR/rdf-sparql-query

Sini M, Rajbhandari S, Amirhosseini M, Johannsen G, Morshed A, Keizer J. 2010. The AGROVOC concept server workbench system: empowering management of agricultural vocabularies with semantics. *IAALD 2010 World Congress*. Montpellier, France.

Suktarachan M, Thamvijit D, Rajbhandari S, Noikongka D, Mahasarakram P P N, Yongyuth P, Kawtrakul A, Sini M. 2008. Workbench with authoring tools for collaborative multi-lingual ontological knowledge construction and maintenance. [2010-03-25]. http://www.lrec-conf.org/proceedings/lrec2008/pdf/624_paper.pdf

Yongyuth P, Thamvijit D, Suksangsri T, Kawtrakul A, Rajbhandari S, Sini M, Keizer J. 2008. The AGROVOC concept server workbench: a collaborative tool for managing multilingual knowledge. In: *Proceedings of the 6th Conference of the Asian Federation of Information Technology in Agriculture*. Tokyo, Japan.

(Managing editor WANG Ning)