From AGROVOC OWL Model towards AGROVOC SKOS Model

Ahsan MORSHED[1], Johannes KEIZER[1], Gudrun JOHANNSEN[1]
Armando STELLATO[2], Thomas Baker[3]

1. The Food and Agricultural Organization of UN (FAO), {name.surname@fao.org}
2. University of Rome, Tor Vergata {stellato@info.uniroma2.it}
3. Washington DC, USA, {tbaker@tbaker.de}

Background

AGROVOC is a multilingual structured thesaurus for the agricultural domain, which is owned and maintained by an international community of Agricultural Research Information Institutions. AGROVOC is used all over the world by researchers, librarians, information managers and others, for indexing, retrieving, and organizing data in Agricultural Information Systems. It currently includes 579523 terms in 19 different languages.

Since 2004 (1) the AGROVOC (within the Agricultural Ontology Service) Initiative underwent a process of semantic cleaning and enrichment to make it a base vocabulary for semantic web applications.

To maintain collectively such a semantically rich concept scheme, a tool called "AGROVOC concept server workbench" was developed. In early June 2010 the "AGROVOC concept server workbench" went into production (http://202.73.13.50:55234/agrovocdevv10/)

In 2004 there was clear intention to adhere to Semantic Web standards for Knowledge Representation and interchange, though there were no suitable model nor standard available for expressing a concept scheme without losing the richness of meaning contained in the original thesaurus. Namely, the first versions of SKOS (which, also, was still not recommended by W3C) were insufficient to express relations between terms.

For this reason the AGROVOC team at FAO developed a model that was based on 3 layers: concept to concept relations, term to term relations and string to string relations. This conceptualization was formalized in an OWL model.

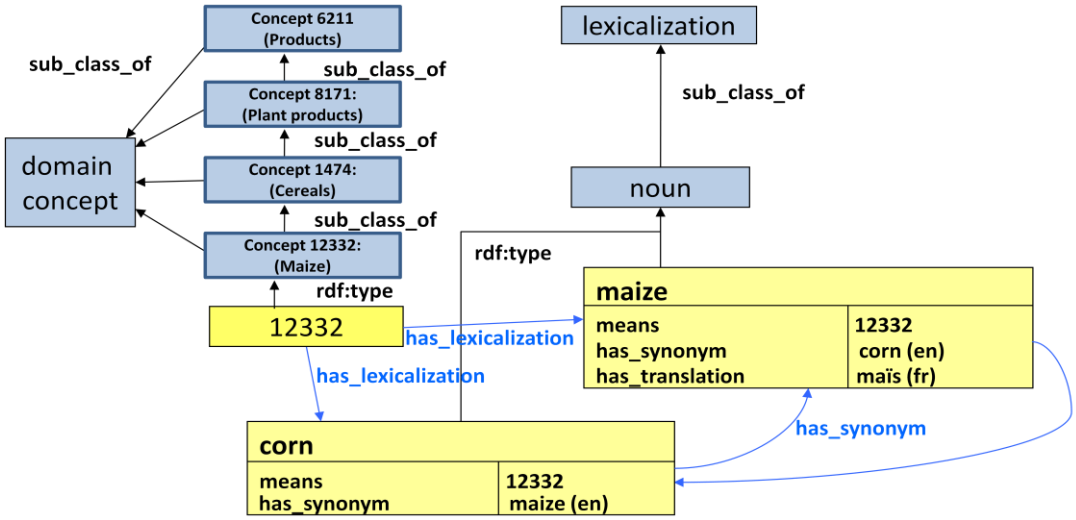

Figure 1 : AGROVOC RDF/OWL model

Figure 1 shows AGROVOC RDF/OWL model which is integrated in ACSW. The model can be described in the following way:

- The natural-language terms of the AGROVOC thesaurus were re-conceptualized as lexicalizations(Label) for underlying Concepts. Lexicalizations included preferred and alternative labels, synonyms, spelling variants, and translations. Descriptors were conceptualized as "preferred" Lexicalizations.
- Concepts were modelled as OWL Classes
- Each Concept-class was associated with one instance of that class (a sort of singleton) as a means of relating a concept to its lexicalizations. In fact this singleton instance was the ground anchor for establishing relationships with all the lexicalizations of the concept, while its associated class participated to the concepts taxonomy via rdfs:subClassOf relationships

  This modelling was chosen to meet a perceived need for description-logic-based computability, as declaring one Class to be an Instance of another Class sacrifices conformance with "OWL DL", a constrained, description-logic-conformant sub-set of the more expressive but computationally intractable variant "OWL Full".
- Relationships could also be specified between Concepts (such as "isUsedIn" or "causes") or between Lexicalizations (such as "hasAcronym") . In 2006 this was considered a significant and innovative feature of the meta model.

The distinction between lexicalization, and string had already been dropped , thus reducing complexity of the model.

The model described above, which is currently reflected in the RDF/OWL representation of AGROVOC has several problems:

- Lexicalizations are related to Concepts only by means of a parallel and artificially redundant set of Instances, which is both conceptually problematic and poses practical difficulties for software developers designing queries, mappings and display interfaces.

- A concept such as Maize, modelled as a Class, is declared to be a sub-Class of the Class Cereals as well as a sub-Class of the Class Domain Concepts, whereas conceptually, Cereals may more properly be seen as an Instance of Domain Concept and Mais as an instance of Cereals.

- Interpreting the Broader Term and Narrower Term relationships used between Concepts in the original AGROVOC Thesaurus as Sub-Class relationships between OWL classes arguably constitutes "ontological overcommitment" and is also prone to force undesirable interpretations (in AGROVOC broader/narrower relationship, type and IS-A relationship are often mixed up, which is not allowed in a relationship, such as rdfs:subClassOf, which has clear and formally defined semantics.

- Furthermore the use of such a "proprietary" model would make it much more difficult in future to use standard tools to manage AGROVOC

AGROVOC WB as SKOS

In the meantime SKOS had evolved to more complexity and completeness and the SKOS version with its extensions, published as W3C standard in 2009 completely catered for all the needs of AGROVOC. In order to overcome the above problems, AGROVOC can be presented in SKOS format (see in Figure 2). The objective of the project is to migrate OWL format to SKOS format for broader usages of information technologies and provide the Linkdata services to Agricultural community all over the world.

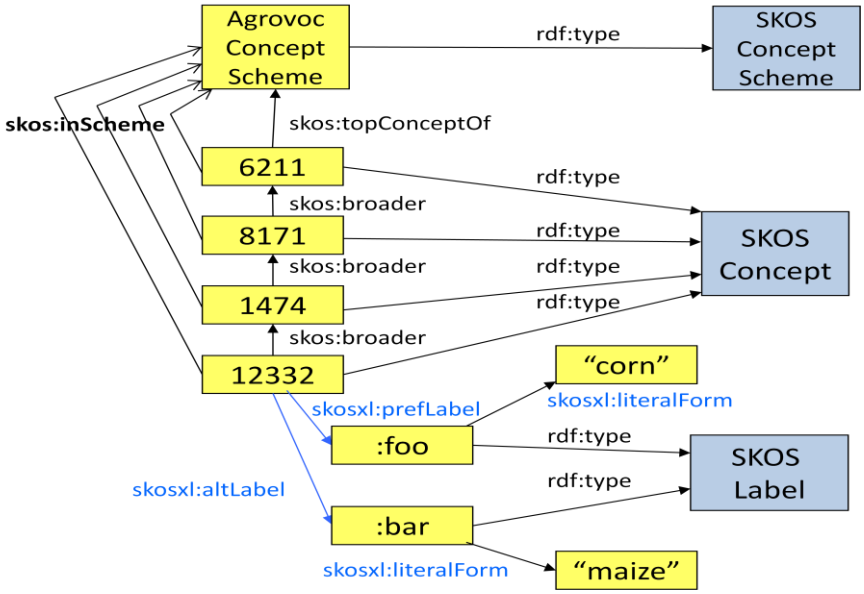

Figure 2 : AGROVOC SKOS model (proposed)

The usages of SKOS model give more flexibility of semantics than RDF/OWL model. Furthermore, SKOS will be easier platform to teach the people and will be used for AGROVOC Linkeddata.

References

[1]. Dagobert Soergel, Boris Lauser, Anita Liang, Frehiwot Fisseha, Johannes Keizer, and Stephen Katz. 2004. Reengineering thesauri for new applications: the AGROVOC example. Journal of Digital Information 4(4).http://journals.tdl.org/jodi/article/view/112/111.

[2]. Alistair Miles, Sean Bechhofer, eds. 2009. SKOS Simple Knowledge Organization System Reference. [W3C Recommendation, 18 August 2009]. http://www.w3.org/TR/skos-reference/.

[3]. AGROVOC Concept Server at http://aims.fao.org/community/group/agrovoc-concept-server-group

[4]. Thomas Baker and Johannes Keizer. Linked data for Fighting Global Hunger: Experiences in setting standards for Agricultural Information Management. (to be published in early 2011)