Using an AGROVOC-based ontology for the description of learning resources on organic agriculture

Salvador Sánchez-Alonso and Miguel-Angel Sicilia

Abstract. Education is a critical requirement for the development of sustainable agriculture. Learning resources available through the Web can be described with metadata for enhanced availability. The provision of semantic metadata describing these resources further facilitates retrieval and selection of resources based on richer annotations that use ontologies. This paper sketches the potential use of ontologies related to organic agriculture for the description of learning resources.

1 Introduction

Organic agriculture is a form of agriculture whose main objective is obtaining food efficiently while respecting the environment and preserving Earth's natural fertility, which is attained through the optimization of the resources available as well as avoiding synthetic pesticides and fertilizers. Although organic farming is nowadays widespread in most developed countries, the promotion of the ecological practices in agriculture requires much effort in terms of education. Different institutions and organizations provide educational resources on the topic, some of them openly available through the Web. However, locating those resources with conventional search engines is complicated, mainly due to noise in the results of common input terms. Learning object repositories provide an alternative –which can be seen as an extension–, which enables more relevant results at the cost of developing (and packing together with the learning objects) a few metadata records. The IEEE LOM standard[1] can be used to provide metadata to learning resources thus facilitating their retrieval.

There exists, however, an additional extension that would provide richer means of browsing, navigating and searching for educational resources: the use of formal annotation based on (formal) ontologies. This approach is specially suited to the field of agriculture, since the large and mature thesaurus AGROVOC[2], a vocabulary covering the terminology of subject fields in agriculture, forestry, fisheries, food and related domains, is widely used in practice, which represents a degree of consensus regarding terminology.

Moving to a Semantic Web practice requires essentially three elements: (i) the elaboration of a formal ontology from the thesaurus, (ii) the provision of specialized semantic search software and (iii) the change in current indexing practices, from the traditional use of thesauri to the new semantic annotation. This paper focuses on element (i), providing a tentative mapping technique for the specifics of AGROVOC. The technique is targeted to the annotation of learning objects, reusing the effort carried out in ontologies for learning resources (Sicilia et al., 2004).

AGROVOC is a structured, controlled vocabulary used for indexing and retrieving data in agricultural information systems. It consists of organized terms (in different languages) covering not only the terminology of agriculture, but also terms in forestry, fisheries, food and other related domains. These terms are used to unambiguously identify resources. Indeed, the knowledge contained in the vocabulary allows standardizing indexing processes, making searching simpler and more efficient.

As in other thesauri, terms are related in AGROVOC, but the even though the kind of relationships supported in thesauri is generally very limited, AGROVOC includes a richer set of relationships classified in "traditional thesaurus relationahisps", "concept-to-concept relationships", "term-to-term relationships" and "String-to-String relationships". The most important among traditional relationships in AGROVOC are:

- *Broader term* (BT) relationships link a general term to other(s) more specific. Thus, the concept `Soil` is BT related to more general terms such as `Land cover`.
- *Narrower term* (NT) relationships represent the opposite of BT. The concept `Soil` is NT related to the three more specific concepts `Top soil`, `Rhizosphere` and `Subsoil`.
- *Related term* (RT) relationships link any two concepts holding a non hierarchical relationship. The term `Fish`, for instance, is RT related

[1] http://ltsc.ieee.org/wg12/

[2] http://www.fao.org/agrovoc/

to terms as varied as `Foods`, `Perishable products`, `Seafoods`, `Fresh products` or `Postmortem changes`.

Other less relevant relationships in AGROVOC are: *Is Referenced in Scope Note* (SNX), *Scope Note Reference* (SNR), *See* (SEE), *Seen for* (SF), *Use* (USE) and *Used for* (UF). But despite the interesting set of relationships included in the thesaurus (those classified as "concept-to-concept relationships", "term-to-term relationships") if resources are to be meta-tagged using the terms and relationships in AGROVOC, these relationships should be unambiguously defined for data to be the basis of advanced management processes based on computational semantics. This unambiguous definition is possible only through a formal representation in an ontology language.

AGROVOC was developed by FAO (*Food and Agriculture Organization of the United Nations*) and the *Commission of the European Communities* in the early 1980s, being updated since then on a regular basis and extensively used for indexing and retrieving data in agricultural information systems. Understanding that similar efforts such as the European GEMET exist (de Lavieter, 1995), the number of terms included (up to 40.000 terms per language), the continuous support and funding from FAO, as well as its generally acceptance, number of active partners and widespread use, make of AGROVOC an outstanding resource in the field of vocabularies and certainly a point of reference for the subject fields covered.

The maturity of AGROVOC makes it a good candidate to become a point of departure for an effort of formalization. In fact, this has already been approached; see for example (Soergel et al., 2004). However, this paper focuses on semantic annotation for learning resources, which would be a particular application of the development of an AGROVOC-based ontology.

The rest of this paper is structured as follows. Section 2 reviews previous work in the field, before section 3 exemplifies the extension of AGROVOC to formal ontology focusing on a particular aspect of organic agriculture, fertilization, which is covered by the thesaurus. Then, section 4 sketches some guidelines on how annotation of learning resources on organic farming can be annotating by making use of the ontology. Finally, conclusions and outlook are provided in section 5.

2 Timeline for the creation of an AGROVOC ontology

From 2003, several efforts have attempted to convert the AGROVOC thesaurus into an ontology. And even though the full conversion of AGROVOC to a formal ontology is currently ongoing work, probably due to the massive effort necessary to properly address this issue, it is fair to recognize that some interesting work has been done during the years.

In 2003, the thesaurus was converted into an RDFS file. Later on, several efforts coordinated to redesign the traditional thesaurus into an ontology-based relational database available to download. Subsequent efforts used AGROVOC in the design of ontologies specific to given domains, such as food safety (Lauser, 2001), or bibliographic information on agriculture (Wildemann, Salokhe and Keizer, 2004), among others. In a more general domain, Kashyap (1999) proposed an approach for designing an ontology for information retrieval based on databases' schemas and a collection of queries that are of interest to the users. However, advanced semantic inference is not considered an issue and thus not dealt with in depth, as the main goal is to reduce the involvement of domain experts. Closer to AGROVOC, an interesting effort by Gangemi et al. (2002) aims at building a fishery ontology by reengineering and integrating the fisheries terminologies from the several systems, one of which is AGROVOC. On the other hand, Fisseha, Liang, and Keizer (2003) sketched some ideas to transform AGROVOC to an ontology, while created a root application ontology based on an application profile addressing the problem of heterogeneity due to differences in terminologies (Liang et al., 2006). These were important moves towards a proper ontology of AGROVOC, even though important issues such as providing "intelligent behind-the-scenes support for query expansion" (in their own words), or creating the complete inventory of domain-relevant entity types and relationship types, to name a few, remain unattained.

In spite of the several ongoing works on developing agriculture ontologies, and the existence of several application profiles and other IEEE LOM-based schemas specific to agriculture such as CGIAR's LOM Core (Beniest & Zschocke, 2005), FAO's learning resources application profile (FAO 2006), Rural-eGov's LOM application profile (Tzikopoulos et al., 2007) and others, there are no reports on the development of specific schemas for the semantic annotation of learning resources for agriculture education.

3 Converting AGROVOC to ontological languages: the case of fertilization

At present, AGROVOC contains close to 30,000 descriptors and more than 10,000 non-descriptors called synonyms (terms which help the user to find a specific descriptor). Given the size and scope of AGROVOC, any ontology created from it will likely include thousands of concepts and relationships, and consequently will not be feasible to manage. Here we will focus on the concrete case of fertilization in organic agriculture as a simple application of a part of the thesaurus.

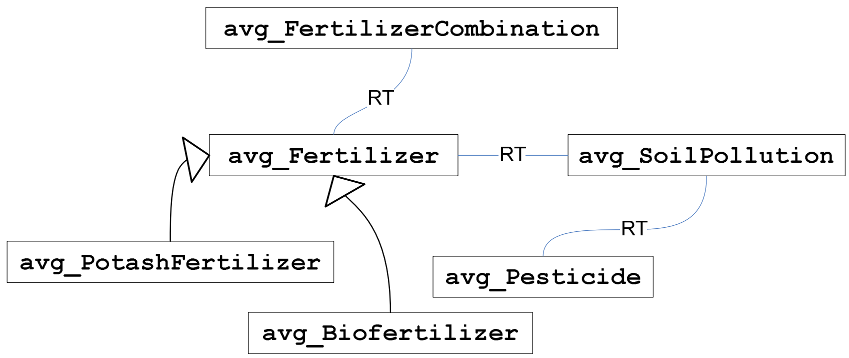

Figure 1. A fragment of the terms in AGROVOC related to fertilization.

Fertilization, understood as "*any aspect of the use of fertilizers to improve crop growth and soil fertility*", is a central concept in organic agriculture. It is a fact that no farming exploitation can be considered –and consequently certified as– *organic*, unless a strict use of fertilizers and fertilizing techniques, specific to organic agriculture, is followed. In Europe, for instance, regulation on organic certification enforces fertilizers and soil conditioners to be composed only of substances listed, which particularly involves varied cultivation practices, and the rigorous limitation on the use of non-synthetic fertilizers. In AGROVOC, the term `Fertilization`[3] is considered a physiological concept, so the AGROVOC term defining the application of fertilizers to soil or plants that should be used is, instead, `Fertilizer application`. This latter term is related to (RT) `Fertilizers`, a concept which includes any natural or synthetic compound spread on the plant or worked into soil to increase the plant's natural capacity to grow.

A particular kind of fertilizers is `Biofertilizers`, a NT related concept describing "any naturally occurring organic substances applied to soil for the purpose of maintaining or improving fertility". As `Biofertilizers` are exclusively composed of natural substances such as animal manures, composts, nitrogen fixing bacteria and mycorrhizae, it is suitable (from the point of view of current regulations) for organic farming

[3] To enhance readability, AGROVOC terms are in `courier` font. In figures, agrovoc terms are prefixed by `agv_` as a form of classifying terms from different sources.

practices. On the contrary, other types of NT related `Fertilizers` such as `Nitrogen fertilizers` and `Phosphate fertilizers` —and their derivatives `Superphosphate` and `Rock phosphate`— are generally prohibited for organic farming. This is because they have been named as factors in the over enrichment of surface waters (excessive nutrients in ponds and lakes cause over growth of algae) and phosphate accumulation in subterranean waters (extremely soluble nitrogen in its nitrate form not sufficiently absorbed by plants can leach into groundwater). Others types of Fertilizers (i.e. NT related to this term) are `Biofertilizers`, `Calcium fertilizers`, `Inorganic fertilizers`, `Liquid gas fertilizers` or `Organic fertilizers`, among others (see Figure 1). Figure 1 shows a fragment of AGROVOC with NT/BT relations interpreted as class/subclass.

An interesting term in AGROVOC is `Fertilizer combinations`. It represents any fertilizer mixture with agents such as herbicides, plant growth substances or pesticides. This is obviously not suitable for organic agriculture according to the definition provided at the beginning of this article. At the moment, this is one of the many terms related to `Fertilizers` (i.e. `Fertilizer combinations` is NT related to `Fertilizers`). Another interesting concept to note is that of `Soil pollution`, a general concept RT related to `Soil degradation` which is not certainly applicable to given types of `Fertilizers` such as `Biological fertilizers`, a powerful reason to better profile the relationship holding between `Fertilizers` and `Soil Pollution`.

Current relationships in AGROVOC suggest the following:

- All kinds of `Fertilizers` can be involved in `Soil degradation`, a concept RT related to `Soil Pollution` which is in turn related to `Fertilizers`.
- Any instance of `Fertilizers` can be part of `Fertilizers combinations`, which is probably false for some kinds of `Biological fertilizers`.

Figure 1 shows the abovementioned concepts and the relations holding between them as a previous step for its formulation and clarification in ontological terms.

In AGROVOC, four relationships showing the hierarchical links between terms exist: BT (broader term), NT (narrower term), RT (related term) and UF (non-descriptor). These relationships are specific to thesauri and, as such, can not be considered a limitation or disadvantage. However, an ontology takes this conceptual framework one step further by structuring the terms more formally, and by providing richer relationships between concepts than what is currently provided in thesauri. An AGROVOC ontology should ideally clarify the hierarchies defined to better prepare the knowledge base for inferences. Thus, a number of basic actions should be taken such as:

- Creating two new sub-concepts of `Fertilizer applications` namely `General fertilizer applications` and `Specific fertilizer applications`. `General fertilizer applications` would give support to fertilizer techniques suitable for all kinds of farming practices (organic and non organic). These techniques will likely include the utilization of permitted Fertilizers and will explicitly avoid non permitted compounds. On the other hand, `Specific fertilizer applications`, would provide support to techniques suitable for specific forms of agriculture. This can be subsequently divided into `Organic agriculture fertilizers applications` (for organic agriculture techniques), extensive agriculture techniques, etc.

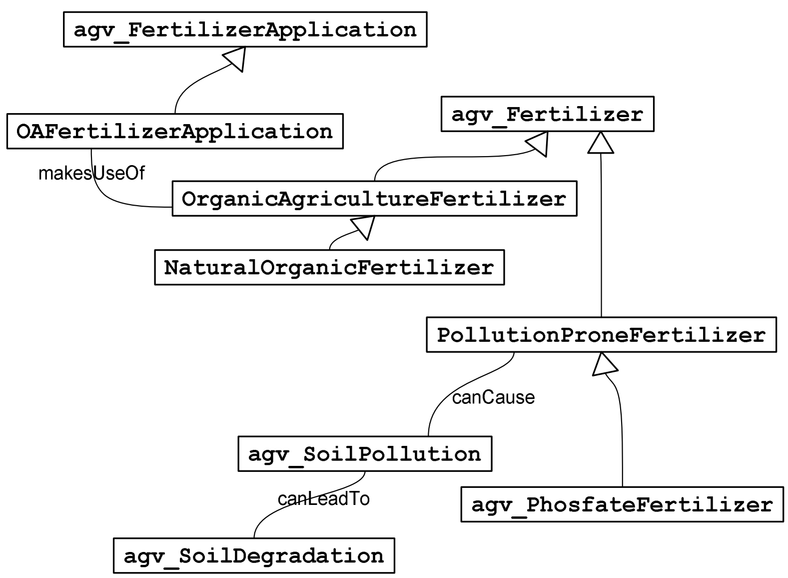

Figure 2. Fragment of an AGROVOC-based ontology for fertilization.

- A subclass of `Fertilizers`, should be created to give support to the concept of Organic agriculture-suitable fertilizer. Let us name this term `Organic agriculture-suitable fertilizers`.
- Remove the relationship between `Fertilizer application` and `Fertilizers`, as in its current for it provides room for errors in selecting the appropriate resource if the selection is based on the

ontology knowledge. A good solution (but more complex than current situation) is to link `Organic agriculture fertilizer applications` to `Organic agriculture-suitable fertilizers` through an ontology property named `makesUseOf`.

- Link `Soil pollution` only to those fertilizers which might be cause of pollution; let us name them `Pollutant-prone fertilizers`. In this manner, the rest of the different types of `Fertilizers` will no longer be able to be associated to `Soil degradation` through `Soil pollution` as this relationship will not hold to all the `Fertilizers` but only to those *pollution-prone*.

Figure 2 summarizes all the above discussion.

4 Case study: annotating resources on fertilization

Learning resources on organic agriculture may be targeted to different kinds of learners. Here we will provide two cases that illustrate such diversity of needs. The first will be a scientific paper while the second is a resource oriented to dissemination of technical practices.

Annotating technical resources

In general, any digital resource can be explicitly declared as an instance of a `LearningObject` class –or any of its subclasses. Then, there is a family of properties (and sub-properties) starting from a generic `about` that connects them to any other concept. The `about` predicates play the same role of a "keywords" marker, but it can be specialized. One important educational issue is the production of organic agriculture fertilizers. However, the notion of what an OA fertilizer is can be defined in several ways. A possible definition can be provided by the next two SWRL rules:

```
NaturalOrganicFertilizer(?x) → OrganicAgricultureFertilizer(?x)

agv_Fertilizer(?x)
∧ defines(?y, ?x)
∧ OrganicAgricultureStandard(?y)
       → OrganicAgricultureFertilizer(?x)
```

Rules in SWRL (Horrocks et al., 2004) are of the form of an implication in the form antecedent → consequent, so that each rule can be read as:

"whenever the conditions specified in the antecedent hold, then the conditions specified in the consequent must also hold". In the examples above, natural organic fertilizers are suitable to organic agriculture but also any fertilizer defined as such by standards or recommendations is.

Composting is one of the most widespread of these production techniques. The *Compost happens!* tutorial (CHT[4]) is an online resource with a common sequential structure. This resource can be declared as an instance `TutorialLearningObject(cht)`. The tutorial can be annotated as `describes(cht, composting)` where `OAFertilizerProductionTechnique(composting)`.

A user searching for techniques for producing organic agriculture fertilizers (anything `about OAFertilizerProductionTechnique`) will match the resource since `produces(composting, compost)` and `compost` will be classified as OA, for example, because it is defined by the BSI's[5] `BSI_PAS_100` standard.

Annotating scientific literature

Palomäki et al (2002) reported a typical comparative study of organic versus traditional techniques, in this case for the particular kind of *Elsanta* strawberries, a variety of strawberry with a particular flavour. This kind of resources considered as LO are regarded as `ScientificReportLO`, and they can be annotated with respect to the concrete research methods employed.

A user might first search for reports regarding `StrawberryPlantType` that report on `Experiments` (specified in the `researchMethodUsed` property). The report mentioned will be of the kind `ComparativeExperiment`. A more detailed specification may specify as `measuredVariable fruitProductivity`. This kind of measures can be used for a systematization of research evidence, but it also can be retrieved as supplementary material for expositive or tutorial learning objects, which are related to the learning topics underway.

5 Conclusions and outlook

An example on how an ontology fragment derived from AGROVOC can be used to annotate resources has been provided. Further work will deal with

[4] http://www.compostinfo.com/main/intro.htm
[5] http://www.bsi-global.com/

advancing the conversion of AGROVOC to formal ontology, and with the provision of tools and improved techniques for semantic annotation specific to learning resources.

Acknowledgements

This work has been supported by EU funded projects LUISA (Learning content management system Using Innovative Semantic web services Architecture, FP6−2004−IST−4 027149), and Organic.Edunet (ECP-2006-EDU-410012).

References

Beniest, J. & Zschocke, T. (2005). Developing a learning object repository for international agricultural research. In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2005 (pp. 4553-4555).

FAO (2006). Metadata Application Profile for FAO's Learning Resources (draft). Online at: ftp://ftp.fao.org/gi/gil/gilws/aims/metadata/docs/learnap.doc

Fisseha, F., Liang, A. and Keizer, J. (2003). Reengineering AGROVOC to Ontologies Step towards better semantic structure. In *Proceedings of the NKOS Workshop, Rice University, Houston, Texas, USA*.

Gangemi, A., Fisseha, F., Pettman, I. Jand Keizer, J. (2002). Building an Integrated Formal Ontology for Semantic Interoperability in the Fishery Domain. In *Proceedings of the First International Semantic Web Conference, ISWC 2002*.

Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B. and Dean, M. (2004) SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C Member Submission, http://www.w3.org/Submission/SWRL/

Kashyap V. (1999) Design and creation of ontologies for environmental information retrieval. In *Proceedings of the 12th Workshop on Knowledge Acquisition, Modeling and Management (KAW'99)*, Banff, Canada.

Lauser, B. (2001). From thesauri to Ontologies: A short case study in the food safety area in how ontologies are more powerful than thesauri - From thesauri to RDFS to OWL. In proceedings of the ECDL 2001 OAI Workshop.

de Lavieter, L. (Ed.) (1995). Multilingual Environmental Thesaurus. Part 1, English; Part 2, Français; Part 3, Deutsch; Part 4, Nederlands; Part 5, Italiano; Part 6, Norsk; Part 7, Dansk; Part 8, Español. NBOI, Nederlandse Bureau voor Onderzoek Informatie / EEA-TF - European Environment Agency - Task Force, Amsterdam, November 1995, pp. (English) vi + A-78; B-112; C-56; D-199, total 445.

Liang, A., Salokhe, G, Sini, M. and Keizer, J. (2007). Towards an infrastructure for semantic applications: Methodologies for semantic integration of heterogeneous resources. In *Cataloging & Classification Quarterly*, 43 (3/4).

Palomäki, V., Mansikka-aho, A.-.M. and Etelämäki, M. 2002. Organic fertilization and cultivation technique of strawberry grown in greenhouse. Acta Hort. (ISHS) 567:597-599 http://www.actahort.org/books/567/567_129.htm

Sicilia, M.A., García, E., Sánchez-Alonso, S. and Rodríguez, E. 2004. Describing learning object types in ontological structures: towards specialized pedagogical selection. In *Proceedings of ED-MEDIA 2004 - World conference on educational multimedia, hypermedia and telecommunications*, pp. 2093-2097.

Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J. and Katz, S. (2004). Reengineering thesauri for new applications: the Agrovoc example. *Journal of Digital Information*, 4(4).

Tzikopoulos, A., Manouselis, N., Yialouris, C.P., Sideridis, A.B. (2007). Using educational metadata in a learning repository that supports lifelong learning needs of rural SMEs. In *Proceedings of the 2007 EFITA Conference*, Glasgow, UK.

Wildemann, T., Salokhe, G. and Keizer, J. (2004). Applying new trends to the management of bibliographic information on agriculture. *Zeitschrift für Agrarinformatik*, 12(1), 9-13.