Margherita Sini (Food and Agriculture Organization of the United Nations, Rome, Italy)
Sachit Rajbhandari (Food and Agriculture Organization of the United Nations, Rome, Italy)
Jeetendra Singh (Indian Institute of Technology, Kanpur, India)
Johannes Keizer (Food and Agriculture Organization of the United Nations, Rome, Italy)
T.V. Prabhakar (Indian Institute of Technology, Kanpur, India)
Asanee Kawtrakul (Kasetsart University, Bangkok, Thailand)

Smart Organization of Agricultural Knowledge
The example of the AGROVOC Concept Server and Agropedia

Abstract: The tendency of representing information in a form that could be better elaborated by computers (the so called "machine readable format") (Berners-Lee 1998) initiated years ago, expanded to many domains, among which Agriculture. The Food and Agriculture Organization of the United Nations, The Kasetsart University and the Indian Institute of Technology Kanpur are pioneers in the representation of information and knowledge related to this domain using modern techniques such as ontology languages. This paper analyzes a couple of projects developed by these organizations, aiming to make use of a concept-oriented approach while describing agricultural topics. It is organized in two chapters each referred to each project, describing in particular the innovative aspects, the benefits, and the technology used.

1: The AGROVOC Concept Server

AGROVOC is a multilingual thesaurus for agriculture developed in the early '80s by the Food and Agriculture Organization and the European Commission. It was originally developed in English and translated in the other four official languages of the UN agency (i.e. Spanish, French, Arabic, Chinese). Due to its large use and success (in particular within the AGRIS network, www.fao.org/agris), it has currently reached 20 languages. As a thesaurus it has been used for indexing resources and for Information Retrieval (IR) on AGROVOC-indexed documents and on free text search engines. However, it is known how controlled vocabularies are of limited semantics and how their reengineering may help improving IR, information discovery and to contribute to the Semantic Web vision (Harper 2006).

In this line, the FAO started in 2001 the Agricultural Ontology Service Initiative (AOS) aiming to provide better services to users making use of semantic technologies. One of the services of the AOS is a concept-based system allowing users to organize and manage the agricultural keywords of the AGROVOC thesaurus into network of concepts. This idea led to the AGROVOC Concept Server, and therefore the following actions were initiated: identification of the new structure that should be given to the concept server (Soergel 2004), specification of the ontological model (Liang 2006), revision and refinement of the AGROVOC Thesaurus data to be compliant with the new structure (Kawtrakul 2005 and Sini 2009), and development of the collaborative platform for the management of the reengineered data (Sini 2008).

Innovative aspects

The reengineering of thesauri into ontologies is a complex matter discussed already by

several authors and similar approaches have been already implemented (van Assem 2006). The solution at FAO was to use the Ontology Web Language (OWL) to implement the new concept-based structure as indicated in Soergel (2004). The FAO recognized the importance of being able to represent specific types of relationships at concept level and at a term level, and this characterizes the AGROVOC Concept Server backbone structure: all terms have been represented as instances of the class "noun" which is a subclass of "lexicalization". The concepts, derived from the AGROVOC main descriptors, are then connected with the different instances with the hasLexicalization property. For a more detailed description of the full model see (Liang 2006). This fact led to the ability of representing all lexicalization and terminological information in addition to the primary conceptual structure of the concept hierarchy. This has been further elaborated in the Lexical Information Repository model of the NeOn EU project (Montiel-Ponsoda 2008).

Benefits

Once the AGROVOC Concept Server and its management tool, the Workbench, will be realized, many agricultural related concepts will be identified with a unique identifier and will be represented with multiple terms in many languages. The terms may also include spelling variants, acronyms, dialectal forms or local terms used in specific geographical area. This led to the ability of: a) realizing URI-based indexing systems, versus the traditional word-based indexing (generally using only English), with consequent ability of creating catalogues more machine-interpretable; b) allowing more interoperability with other systems using ontologies: the Concept Server may contain mapping and linking to other URI, if not reusing those; c) allowing to the users the freedom to use any language they wish or are used to, and any term they want to find agricultural information (no need to refer always to preferred terms as defined in controlled vocabularies).

Current efforts are undergoing in FAO in order to apply the Concept Server data to existing cataloguing systems (e.g. AGRIS, the FAO document repository), and Kasetsart University is also exploiting the same techniques and tools for improving access to information for example for farmers in Thailand.

The Technology

The AGROVOC Concept Server model had been designed using OWL DL. A basic model (www.fao.org/aims/aos) acts as a Foundational Agricultural Ontology identifying how the agricultural concepts and their lexical representation should be represented.

The original AGROVOC Thesaurus (www.fao.org/agrovoc) is stored in MySQL database. It has been converted to OWL using a java-based routine and imported in a triple store using the Protégé API[1].

The Workbench system is a Java-based web application. This application accesses the triple store via the Protégé OWL API and outputs the results to the users in an AJAX style using the Google Web Toolkit (GWT). Additional widgets from GWT incubator have been used to design a more user friendly system.

Hibernate has been used to interact with the database providing high performance object/relational persistence, easy retrieval/update of data, transaction management and database connection pooling. Gilead (previously known as hibernate4gwt) has been used

---

[1] http://protege.stanford.edu/plugins/owl/api/

as a layer between GWT and Hibernate, allowing the use of persistent entities outside the JVM.

AGROVOC Workbench also exposes its triple store data via web services and RSS feeds. With these technologies, other applications can easily access data and news about changes, thus reducing time and resources to download whole database and update the latest version of the AGROVOC Concept Server data into their applications.

The system is available at naist.cpe.ku.ac.th/agrovoc.

2: The Agropedia project

Agropedia can be defined as an agriculture knowledge repository of universal meta models and localized content for a variety of users with appropriate interfaces built in collaborative mode in multiple languages.

The primary objective of the Agropedia project (agropedia.iitk.ac.in) is to build an infrastructure of agricultural knowledge presented in various customized ways to different stakeholders, being those scientists, students, extension workers, farmers, or policy makers. Multilingual and localized information is very important in India.

In order to have a backbone system "smart" enough to allow easy interaction with other data, e.g. with the agricultural data already available in several Indian institutes such as the G. B. Pant University Of Agriculture & Technology, Patnagar, or the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), partners of the project, the Indian Institute of Technology Kanpur (IITK) decided to develop Knowledge Models (KMs), which should act as conceptual reference for many domains starting from different crops such as Chickpea, Groundnut, Litchi, Pigeon pea, Rice, Sorghum, Sugarcane, Vegetable pea, and Wheat.

While working with domain experts, an easy tool for knowledge representation to work with is essential. Elaborated ontology editors such as Protégé, the NeOn Toolkit, Swoop or others are too complex for non Information Management expert. Consequently, IITK initiated a series of workshops in order to explain how to represent knowledge models using the CMap tool (cmap.ihmc.us/coe) to agronomists, soil scientists, plant breeders or geneticists, farm managers, and other experts. A series of maps were prepared where agricultural concepts were represented in a consistent way across multiple maps. Specific guidelines were elaborated so that "Seed_treatment" in the chickpea KM results the same concept of the one used in the Sorghum KM.

Although the KMs were prepared with the CMap tool, known more as a visual tool for human consumption, more than machine consumption, the maps were prepared in an intelligent way: the guidelines homogenized the representation of the URI, allowing easy linkage or mapping between maps. Instances and classes were differentiated thanks to the functionalities of the tool, and object properties were taken from a registry for easy reuse.

The registry of relationships (containing object properties and data type properties) has been build reusing some relationships from the AOS ontological relationships used to refine the AGROVOC Thesaurus in the AGROVOC Concept Server (aims.fao.org/en/website/Ontology-relationships) and has extended them, with consequent mutual enrichment between the two registries.

However, not a unique registry could have been built because of the need of keeping modularized agricultural information: the AGROVOC Concept Server aims to represent general agricultural related information and be the base for the development of more domain specific ontologies, the Agropedia KMs aim to represent more in detail

information related to local fertilizers,   soil, cropping techniques and methods, etc.

Innovative aspects

Agropedia presents to users different semantically oriented tools: textual and audio blogs, wikis, forums, and the KMs presented in different formats (pdf, static or context-sensitive images). Users have the possibility to choose a preferred way of navigating the KMs: as a simplified hierarchical list or as a network of concepts and instances. In both cases, when the user clicks an element, he/she get as results all resources in the knowledge base (catalogue) related to the selected item. In fact, resources from the library catalogue are already tagged with the concepts from the KM, making the system highly innovative compared to the traditional string search. No matter what languages the maps are displayed, the results will be always the same (currently, KMs exists in English and Hindi).

Benefits

Agropedia is an attempt to inject social networking and semantic technologies into agriculture in general and Indian agriculture in particular.

Agriculture knowledge (for that matter any knowledge) can be said to be of two kinds – the well articulated, expert certified knowledge what we call *Gyandhara* and the tacit, folk knowledge, called *Jandhara*. Agropedia attempts to provide a platform for hosting both kinds of knowledge and also linking amalgamating them in an interesting way.

The Library section of the Agropedia is the expert certified knowledge. Wiki, blogs, Forum provide the platform for un-regulated people-created content/knowledge. In addition, Agropedia permits users to comment upon certified knowledge in the library. These comments, usually in the nature of how they benefited from that piece, ratings, variations in the advice that have been innovated by the user, etc., are un-moderated and create a situation where the certified knowledge and experience by the end-user sits together greatly increasing its utility.

The Technology

The first release of Agropedia was implemented using the Alfresco suite. Subsequently, because of the need of incorporating other functionalities, the Drupal Content Management System (CMS) has been used as the main technological backbone.

The KMs, as mentioned, are created with CMap, and exported in SVG format to make them available as context-sensitive images. Other formats (pdf, jpg) are used for visualization only. A Java routine customize the OWL version of the KMs (obtained with a simple export from the CMap tool) as a hierarchical tree as a content index of the catalogue.

Drupal is used for implementing the blogs, the chats, the forums, the Q/A area, the user management area, etc. The taxonomy module is used for tagging and searching the content.

A Java module for automatic tagging using an the KMs is in process of implementation.

3: Conclusions and Future Works

The work undergoing by FAO and other AOS partners for making better use of traditional thesauri is inline with the current strategies of making data more

"processable". Similarly, the Agropedia project opens the road to the representation of agricultural knowledge in the form of concept based maps.

A lot still remains to do: for the AGROVOC Concept Server investigations on the role of OWL2 and OWL rules should be carried out, as well as the completion of the collaborative tool to maintain the data pool. Further extensions of this tool consider the exploitation of knowledge extraction from corpus text in different languages to enrich the concepts with more terms or synonyms.

In Agropedia further work is planned to develop more KM, and export the CMap formatted data to OWL for further incorporation and better services through the Agropedia portal, in particular making use of reasoning tools between the different maps.

For both systems, mutual integration can be further investigated, being them both addressed to the same area of users.

Finally, Linked Data (linkeddata.org) exposure for both systems is a must.

References

Berners-Lee T., September 1998, *Semantic Web Road map,*
    <www.w3.org/DesignIssues/Semantic.html>

Harper C.A., Tillett B., 2006, *Library of Congress controlled vocabularies and their application to the Semantic Web*

Soergel D., Lauser B., Liang A., Keizer J., Katz S., 2004, *Reengineering thesauri for new applications. The AGROVOC example*, in Journal of Digital Information, Volume 4 Issue 4, Article No. 257

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*, OWL workshop 2006, Athens, Georgia, U. S. A., and International Conference on Dublin Core and Metadata Applications, DC-2006--Colima, Mexico Proceedings

Kawtrakul A., Imsombut A., Thunyakijjanukit A., Soergel D., Liang A.C., Sini M., Johannsen G., Keizer J., 2005, *Automatic Term Relationship Cleaning and Refinement for AGROVOC*, < www.fao.org/docrep/008/af240e/af240e00.htm>

Sini M., Soergel D., Johannsen G., 2009, *Proposal for restructuring Scientific Names and Common Names of Organisms in AGROVOC*, <ftp.fao.org/docrep/fao/011/ak283e/ak283e00.pdf>, <agrovoc-revision-refinement.blogspot.com>, <www.icrisat.org/vasat/agrovoc.htm>

Sini M., Lauser B., Salokhe G., Keizer J., Katz S., 2008, *The AGROVOC Concept Server: rationale, goals and usage*, eds. Emerald Group Publishing Limited, Journal: Library Review, Year: 2008, Volume: 57, Issue: 3, Page: 200 – 212

Montiel-Ponsoda E., Aguado de Cea G., Gómez-Pérez A., Peters W., Coling 2008, *Modelling Multilinguality in Ontologies*, Companion volume, Posters and Demonstrations, pages 67-70, Manchester, August 2008

van Assem M., Malaise V., Miles A., Schreiber G., 2006, *A Method to Convert Thesauri to SKOS*, in European Semantic Web Conference, pp. 95-109, ESWC 2006