The AGROVOC Concept Server Workbench System: Empowering management of agricultural vocabularies with semantics

Margherita SINI[1], Sachit RAJBHANDARI[1], Maziar AMIRHOSSEINI[2], Gudrun JOHANNSEN[1], Ahsan MORSHED[1], Johannes KEIZER[1]

1. Food and Agricultural Organization of the UN (FAO), {name.surname@fao.org}
2. The Faculty of Information Science and Technology, Universiti Kebangsaan, Malaysia

Abstract

The Food and Agriculture Organization of the United Nations (FAO) is recognized as an information and knowledge-base organization. FAO's activities comprise four main areas which are closely related to various aspects of information and knowledge: capture and analyze, disseminate and share, localize and provide. The goal of developing and maintaining tools for information and knowledge management is attributed to the Office of Knowledge Exchange, Research and Extension (OEK) of FAO. One of the most important resources for covering the terminology of all subject fields in agriculture domain is the AGROVOC thesaurus, which evolved into a semantic system in order to provide ontology services. This newly reengineered system is called the “AGROVOC Concept Server Workbench (ACSW)”.

This article analyzes the different knowledge productions modes for the traditional AGROVOC and the new ACSW system: mode I assimilate to the traditional AGROVOC Thesaurus management and mode II to the ACSW system.

Keywords: ACSW, Epistemic approach, Knowledge production, Knowledge modes,

Introduction

Knowledge exchange and improving worldwide access to information in the agricultural domain by developing knowledge management resources, standards and tools is one of the main activities through which the Food and Agriculture Organization of the United Nations (FAO) aims to combat hunger and poverty in the world. One of the most important resources for covering the terminology of all subjects of interest to FAO (agriculture, forestry, fisheries, food and related domains, e.g. environment) is AGROVOC, the multilingual agricultural thesaurus, developed by FAO and the Commission of the European Communities in the early 80s. Since then it has continuously been updated by FAO in collaboration with partner organizations in different countries, and is now available online in 19 languages [1].

AGROVOC is currently being converted from a traditional term-based knowledge organization system (KOS) to a concept-based system [Soergel, 2004], the AGROVOC Concept Server (CS). The CS allows the representation of more semantics such as specific relationships between concepts as well as relationships between their multilingual lexicalizations. It will function as a resource to help structure and standardize agricultural terminology in multiple languages for use by any number of different users and systems around the world. A tool, the AGROVOC Concept Server Workbench (ACSW), has been developed by FAO in collaboration with Kasetsart University in Thailand and other partners, which supports the maintenance of the CS data in a distributed environment [Sini, 2008]. One of the goals is to set up a network of international experts who can share the collaborative maintenance and extension of the AGROVOC Concept Server, and thus enhance the creation of agricultural knowledge much more efficiently.

The ACSW is part of the larger Agricultural Ontology Service (AOS) initiative and the first major step towards an "Ontology Service" [Fisseha, 2001], which has the goal to provide semantic-based services to users in the agricultural domain. Once fully operational, the ACSW will offer a contextually rich and modern framework for modelling, serving, and managing agricultural terminology. When integrated with web-based search tools, it will greatly facilitate resource retrieval. It should provide access to document-like objects in a variety of languages and offer suggestions for other related resources that are potentially relevant to the topic of interest. The CS is foreseen to empower a variety of useful services such as automatic or semi-automatic translation services, information discovery and reasoning services, guided search services and concept disambiguation services. Such additional functionality will not only dramatically increase the scope of web-based search engines, but also revolutionize the way users interested in agricultural resources interact with the Web.

In this paper we describe the architecture of the new ACSW system (chapter 2), discussion the epistemic analysis (chapter 3), and provide conclusion with future studies (chapter 4).

2. Architecture of the System

The ACSW is a semantically structured system consisting of agricultural concepts with their lexical representation, and specific relationships. These concepts have been obtained remodeling the traditional AGROVOC thesaurus and using the web ontology language (OWL) [Lauser, 2006]. The Protégé OWL API has been used as the framework to manage the obtained ontology. Triples are stored in a MySQL database providing scalable persistence storage for large ontology like the CS.
The ACSW is a java based web application for collaborative building and structuring of multilingual ontologies and terminology systems. The graphical user interface of the system is developed using the Google Web Toolkit (GWT) providing AJAX environment decreasing the need to fetch new HTML page for each request. Client communicates with a server across a network using remote procedure call (RPC). Based on these server calls, the stored triples are accessed via the Protégé OWL API. Different widgets from GWT have been used to make the system more user-friendly and efficient. The Hibernate layer is used for the interaction with the database for easy retrieval/update of data, connection pooling and improved transaction management. Gilead (previously called as hibernate4gwt) has been used as a layer between GWT and Hibernate for the use of persistent entities outside the Java virtual machine (JVM).

The AGROVOC Concept Server model has three different level of representation: concept, term and term variants. Concept is only the abstract meaning given to the group of the terms. For e.g. ‘rice’ in the sense of plant. Terms are the language specific lexical form of that concept. For e.g. ‘rice’ in English, ‘arroz’ in Spanish, or ‘riz’ in French. Finally, term variants are the range of forms that can occur for each term. For e.g. ‘Organization’ or ‘Organization’.

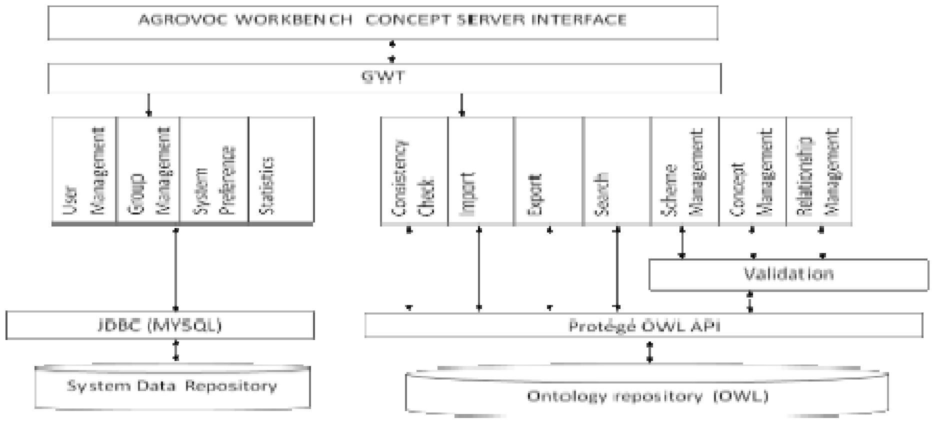

Figure 1: Architecture of AGROVOC Concept Server Workbench

The ACSW system provides the following modules for the management of ontologies :

Concept Management: This module allows concepts navigation, visualization, creation, deletion, and edits. After the creation of new concepts, user can add, edit or delete additional information associated to those concepts, as indicated below.
*Terms***:** List of words that represents the concepts in any language, e.g. "rice (en)" and "riz (fr)". Terms can be added, edited or deleted. Two terms can be related with each other using specific appropriated relationships.
*Definitions***:** It is the meaning of the concept which is valid across languages and lexical variances. For e.g. the definition of the concept 'Cycadaceae' (en) is "ancient palm like plants closely related to ferns in that fertilization is by means of spermatozoids (en)".
*Notes***:** May contain some important information about the concept for sharing with the other users in the community such as scope note or editorial note.
*Attributes***:** Domain specific information for the concept. For e.g. 'hasNumber' or 'hasCity'.
*Relationships***:** Relates concepts to each other. For e.g. 'has Pest'.
*Images***:** Images, pictures, or diagrams that describes the concept.
*History***:** Information such as creation date, latest modification date, status of the concept and the list of its changes for tracking concepts versions and updates.

Search: In the simple search, users can provide any keyword and the search result will return the concept that matches the keyword with the term's labels. In the advanced search feature more options are available, for providing a better result, e.g. using regular expressions, case sensitive function and search into descriptions. The result can be made more specific by filtering concepts using relationships, term codes, concept status, concept/term attributes, languages, or classification schemes.

Relationship Management: In this module, users can add, edit, or delete object properties or data type properties and their related information, such as labels in multiple languages, definitions, properties such as transitivity or symmetry, and domain and range.

Schemes: This module allow users to group concepts into user custom defined classification schemes. Each classification scheme has multiple level categories, created ex-novo or created reusing concepts from the CS data pool. For e.g. An example of pre-loaded classifications scheme is the AGRIS/CARIS Classification Scheme.

Validation: People can have their own way to construct ontology or maybe they have different background knowledge. If the user create a new concept, it has to be validated with domain/ontology expert before being published and used by others. Every action such as add/edit/delete of concept/term/relationships needs to be approved by two types of users, which are 'validators' and 'publishers'.

Consistency Check: Checks the validity and quality of the CS data. The system automatically investigate the data and return inconsistencies which would need to be manually fixed.

Export: Export ontology from OWL format to RDF, XML, TBX, SKOS, and SQL format. An import functionality is envisaged in subsequent releases.

Statistics: This module provides statistical information on frequency of system uses, e.g. count the total number of registered users, concepts, terms, relationships, and exports carried out.

User and Group management: These modules allows management of users and groups, (create, delete, etc.) and associate users to group(s) with different access permissions.

System Preferences: Sets preferences for future sessions, such as show/hide URIs, show/hide non-preferred terms, set languages or change password, set email for notifications, choose default ontology or the starting page after login.

The AGROVOC Concept Server Workbench also uses other web technologies like RSS and web services to expose its triple store data. This helps other third party applications to easily access the knowledge base. The system is available at http://naist.cpe.ku.ac.th/agrovoc.

3. Epistemic analysis of knowledge production modes within the AGROVOC systems

Epistemology can help us to identify things that are "epistemically valuable", in the sense that are relevant for taking particular information management decisions between available alternatives [Fallis, 2009]. The process of knowledge production is essential when considering the needs to take decisions, and epistemic values are one of the elements to consider. Knowledge production may occur under specific modes (identified as mode I and mode II), discussed through analyzing paradigms of the higher education functions. In this section we would like to demonstrate how Mode I is applicable to the traditional AGROVOC thesaurus, and Mode II is applicable to the ACSW system.

Mode I of knowledge production and AGROVOC

Mode I refers to a form of knowledge production – intended as a complex of ideas, methods, values, and norms – that is discipline-based [Gibbons, 1998]. In other words, Mode I of knowledge production is operationalized, likely to be conducted in the absence of a practical goal or application, driven by individuals [Moravec, 2007]. Mode I allows self-determination of research agendas and is limited to academic accountability and scientific disciplines rather than the public [Estabrooks, 2008].

Knowledge in AGROVOC has the same characteristics of others traditional thesauri, since traditional thesauri follow specific standards. Therefore, knowledge attributes of AGROVOC are mostly the same of others traditional thesauri and could be compared with knowledge production modes features. Below, we describe the features of knowledge production Mode I and we compare them with AGROVOC attributes:

1. Community interests: in knowledge production Mode I, problems are set and solved in a context – a particular discipline – governed by the interests of a specific community [Gibbons, 1998]. AGROVOC producers and users identify a specific community in the field of agricultural information. We can therefore assume that this attribute is matched in AGROVOC.

2. Disciplinary: Mode I of knowledge production is disciplinary; in brief, the disciplinary structure defines both what shall count as "good science" and prescribes as well, what users need to know [Gibbons, 1998]. In disciplinary structures, knowledge organizers make decisions on what the problems and solutions are. In constructing AGROVOC as a traditional thesaurus, producers had their own policies for developing controlled vocabulary. However, there is no capacity for moving toward problem solving in the context of applied sciences in the traditional AGROVOC.

3. Homogeneity: Mode I is characterized by relative homogeneity of skills [Gibbons, 1998]. Similarly, traditional thesauri have been developed by homogeneous skills as a team work, for instance, information scientists, librarians, documentalists, subject specialists and linguists.

4. Relation to public and social interests: This feature in knowledge production Mode I is not necessarily present [Gibbons, 1998]. Analogously, traditional thesauri usually follow their rules and policies in information storage and retrieval rather than public interests.

5. Quality control: This feature in knowledge production Mode I is determined essentially through the peer review judgments about the contributions made by individuals [Gibbons, 1998]. In the AGROVOC case also, peer review for controlling quality of terms and their relations have been done by prominent subject specialists.

Mode II of knowledge production and the ACSW system

In Mode II, knowledge production has become increasingly interrelated with technological, social and economical applications and take place outside the institutional regime by using Information and Communication Technologies (ICTs). Online data repositories, Internet archives, and the world wide web provide new ways of doing knowledge production. This mode focuses increasingly on monitoring, modeling, and mapping [Heimeriks, 2008], and can be described as transdisciplinary knowledge production [Estabrooks, 2008].

The ACSW system aims to be more than a tool for "collaborative building and structuring multilingual ontology" [Yongyuth, 2008], mostly because of its characteristics of supporting "knowledge of knowledge organization" [Gorman, 2005]. In fact, the ACSW is an intelligent behind-the-scenes support for powerful knowledge modeling [Lauser, 2001], with the aims of realize services for improving automatic indexing, free-text searching, artificial intelligence systems and Semantic Web applications [Soergel, 2004]. Here below, we intend to describe features of knowledge production mode II and compare them with the ACSW attributes.

1. Knowledge produced in the context of application: Problem-solving in knowledge production Mode II is organized around a particular application and knowledge results. Knowledge is always produced under an aspect of continuous negotiation, that is, it will not be produced unless and until the interests of the various actors are considered. Nonetheless, knowledge production in Mode II is the outcome of a process in which supply and demand factors can be said to operate [Gibbons, 1998]. Reengineering AGROVOC for fulfilling ACSW goals, for instance, is the best answer for social demands of more precise and unambiguous semantics, limited in a poorly defined KOS like a traditional thesaurus [Soergel, 2004]. In addition to this, the ACSW data model allows users to organize knowledge based on their own distinct application needs [Sini, 2008].

2. Transdisciplinarity: The research carried out in the context of application might be said to characterize a number of disciplines in the applied sciences and engineering (e.g., more recently, computer science). In Mode II, the shape of the final solution will normally be beyond that of any single contributing discipline. It will be transdisciplinary [Gibbons, 1998]. ACSW is closely related to context of application. This characteristic is operationalzed through ICTs capacities. ICTs influences can be seen in the ACSW data models, i.e. a concept based structure [Sini, 2008]. In fact, machine-processable formats and re-engineering of AGROVOC into an ontology and sharing ontologies across the web by using OWL [Lauser, 2006] to support artificial intelligence and semantic Web applications (text mining on the Web) [Soergel, 2004], demonstrate the transdisciplinarity roles of ACSW.

3. Heterogeneity and organizational diversity: Mode II is heterogeneous in terms of the skills and experienced that people bring to it [Gibbons, 1998]. Heterogeneous ontological domains which have been mapped with AGROVOC ontology, play roles of the skills and experienced people. Also, heterogeneous systems give heterogeneous interpretations [Gangemi, 2002] in ACSW knowledge production. We recognize the heterogeneity between ontologies while mapping them and when trying to find common points to navigate their various interpretations. The project of concept-based mapping of two agricultural thesauri, the Chinese Agricultural Thesaurus (CAT) and AGROVOC is the best example of the heterogeneity source in terms of language and cultural diversities and translation problems [Liang, Sini, 2006]. The mapping project between the ASFA thesaurus and the fishery part of the AGROVOC thesaurus [Gangemi, 2002], also demonstrate how heterogeneity in term of ontological structural diversities could be considered and resolved [Gliozzo, 1998].

4. Enhanced social accountability and reflexivity: Growing awareness about the variety of ways in which progress in science and technology can affect the public interest has increased the number of groups who wish to influence the outcome of the research process. Social accountability permeates the whole knowledge production process. In knowledge production Mode II sensitivity to the impact of the research is built in from the outset. It forms part of the context of application [Gibbons, 1998]. In the system we are presenting, social accountability and public interests are conducted by ACSW capacities.

5. Quality control: Although, peer review still exists, in knowledge production Mode II additional criteria are added through the context of application which now incorporates a diverse range of intellectual interests as well as other social, economic or political ones [Gibbons,1998].

The study of ACSW's previous attributes in comparison to the features of knowledge production Mode II shows that the ACSW has appropriate answers to participants, end users in the term of social accountability and reflexivity. This means that the ACSW has been successfully in problem solving in the context of application and collaborative functions. On the other hand, it still follows peer review methods (through ontology editors, validators, and publishers) to ensure quality control [Sini, 2008].

Towards Mode III of knowledge production: the future of the ACSW

Mode III of knowledge production focuses on enhance of personal and social capitals which are the sources of knowledge creation. Personal capital, which allows to create a vision of what one wants to be, is a holistic approach to describe individual potential. In addition, personal capital is about innovating the self in continuous growth and improved capacity to catalyze other forms of capital, such as social capital in organizations, institutions, and corporations. Mode III of knowledge production is generated within a knowledge creation based on intellectually proactive individuals which permits the identification, measurement, and utilization of tacit knowledge [Harkins, 2005].

The capacities of the ACSW system, in terms of ontology management, prepare an appropriate base for users collaborations in ontology development. User's participation in ontology development can occur in various fields: expansion of the agricultural knowledge, heterogeneous interpretations in the agricultural domain, extension of agricultural tacit knowledge use in conceptual and semantic relations, and development of semantic web applications. Therefore, the ACSW has a strong potential for moving toward Mode III of knowledge creation by enhancing personal capital and social capital. The Ontology Game, for instance, as a future plan, which will be a source for terms acquisition and word relations [Yongyuth, 2008], will be the basis for using tacit knowledge in knowledge production Mode III.

4. Conclusion and future studies

The AGROVOC Concept Server Workbench system appears to be an innovative and powerful tool for organizing knowledge in the agricultural domain empowering the use of agricultural knowledge in the community. The legacy AGROVOC traditional thesaurus is under restructuring and will be available through this new collaborative tool so that AGROVOC experts can directly modify information with no further delay, provide suggestions on agricultural-related concepts from anywhere in the world, and requests for improvements on content and services.

Future works include a revision of the backend model to allow use of inferencing and reasoning mechanisms, system enrichment with automatic knowledge extraction services, further improvements in functionalities, and incorporation of KOS mappings. Subject indexing or information retrieval will be therefore further enhanced.

The ACSW systems, while ICTs evolves, aims to be an efficient and effective tool for knowledge management and knowledge production Mode III to satisfy multiple "ways of knowing" and serve computer consumption for better user services.

References

- [Soergel, 2004] Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., Katz, S. (2004), "Reengineering thesauri for new applications: the AGROVOC example", *Journal of Digital Information*, Vol. 4. <ftp://ftp.fao.org/docrep/fao/008/af234e/af234e00.pdf>

- [Sini 2008] Sini, Margherita, Boris Lauser, Gauri Salokhe, Johannes Keizer, Stephen Katz (2008) "The AGROVOC Concept Server: rationale, goals and usage" *Library Review*, vol. 57, Issue 3, p. 200 – 212. <http://agrovoc-cs-workbench.googlecode.com/files/Final-322881-1.doc>
- [Fisseha, 2001] Fisseha, Frehiwot; Hagedorn, Kat; Keizer, Johannes and Katz, Stephen (2001) "Creating the Semantic Web: the Role of an Agricultural Ontology Server (AOS)" *International Semantic Web Workshop (Infrastructure and Applications for the Semantic Web*. <ftp://ftp.fao.org/docrep/fao/008/af228e/af228e00.pdf>
- [Lauser, 2006] Lauser, Boris, Margherita Sini, Anita Liang, Johannes Keizer, Stephen Katz (2006) "From AGROVOC to the Agricultural Ontology Service / Concept Server An OWL model for creating ontologies in the agricultural domain" Proceedings of the OWLED*06 Workshop on OWL: Experiences and Directions, Athens, Georgia, USA, November 10-11, 2006. Also available <ftp://ftp.fao.org/docrep/fao/009/ah801e/ah801e00.pdf>
- [Fallis, 2009] Fallis, Don, Dennis Whitcomb (2009) "Epistemic Values and Information Management". *The Information Society*, Volume 25, Issue 3 (May 2009), p. 175-189. <http://sirls.arizona.edu/files/14/TIS_fallis_whitcomb.pdf>
- [Gibbons, 1998] Gibbons, Michael (1998) Higher Education Relevance in the 21st Century Gibbons. Paris: World Bank, Human Development Network/Education, 1998. 62p. [Online] <http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079956815/Higher_ed_relevance_in_21st_century_En98.pdf>
- [Moravec, 2007] Moravec, John W. (2007) A New Paradigm of Knowledge Production in Minnesota Higher Education: A Delphi Study. Advisor Arthur M. Harkins [Dissertation, Doctor of philosophy] Faculty of the graduate school, University of Minnesota. 300p
- [Estabrooks, 2008] Estabrooks, Carole A., Peter Norton, Judy M. Birdsell, Mandi S. Newtond, Adeniyi J. Adewale, Richard Thornley (2008) "Knowledge translation and research careers: Mode I and Mode II activity among health researchers". *Research Policy* 37 (2008) 1066–1078.
- [Heimeriks, 2008] Heimeriks, Gaston, Peter van den Besselaar, Koen Frenken (2008) "Digital disciplinary differences: An analysis of computer-mediated science and 'Mode 2' knowledge production". Research Policy, 37 (2008) 1602–1615
- [Yongyuth, 2008] Yongyuth, Panita, Dussadee Thamvijit, Thanapat Suksangsri, Asanee Kawtrakul, Sachit Rajbhandari, Margherita Sini, Johannes Keizer (2008) The AGROVOC Concept Server Workbench: A collaborative tool for managing multilingual knowledge. [Online] <ftp://ftp.fao.org/docrep/fao/010/ai168e/ai168e00.pdf>
- [Gorman, 2005] Gorman, Michael (2005) A paper on education for librarianship: ALA'S Standards for Accreditation of Master's Programs in Library and Information Studies, 1992. [Online] <http://mg.csufresno.edu/PAPERS/Accreditation_standards.pdf>
- [Lauser, 2001] Lauser, Boris, Margherita Sini, Anita Liang, Johannes Keizer, Stephen Katz (2006) "From AGROVOC to the Agricultural Ontology Service / Concept Server An OWL model for creating ontologies in the agricultural domain" Proceedings of the OWLED*06 Workshop on OWL: Experiences and Directions, Athens, Georgia, USA, November 10-11, 2006. <ftp://ftp.fao.org/docrep/fao/009/ah801e/ah801e00.pdf>
- [Gangemi, 2002] Gangemi, Aldo, Frehiwot Fisseha, Ian Pettman, Johannes Keizer "Building (2002) "An Integrated Formal Ontology for Semantic: Interoperability in the Fishery Domain". Agricultural Information and Knowledge Management Papers 13p. [Online] <ftp://ftp.fao.org/docrep/fao/008/af242e/af242e00.pdf>
- [Liang, Sini, 2006] Liang, A. C., M. Sini (2006) "Mapping AGROVOC and the Chinese Agricultural Thesaurus: Definitions, tools, procedures" New Review of Hypermedia and Multimedia, Volume 12, Issue 1, June 2006 , pages 51 – 62. Also available <ftp://ftp.fao.org/docrep/fao/009/ag862e/ag862e00.pdf>
- [Gliozzo, 1998] Gliozzo Alfio Massimiliano, Caterina Caracciolo, Mathieu d'Aquin, Marta Sabou, Wim Peters, Johanna Voelker (2007) 7.3.1: Results from Experiments in Ontology Learning including Evaluation and Recommendation. [Online] <ftp://ftp.fao.org/docrep/fao/field/009/ai163e/ai163e00.pdf>
- [Harkins, 2005] Harkins, A, Vitaskova, A., and Mollberg, L.B. (2005). Strategies for innovation in tertiary education: Producing mode III knowledge & personal capital. Theory of Science, XIV/XXVII, 1/2005, 69-88

---

[1] AGROVOC: http://www.fao.org/agrovoc/