Lesson Learned for Ontology Construction with Thai Rice Case Study

Aree Thunkijjanukij¹, Asanee Kawtrakul², Supamard Panichsakpatana³, Uamporn Veesommai³

¹ Thai National AGRIS Centre, Kasetsart University, Bangkok, Thailand, libarn@ku.ac.th
² Department of Computer Engineering, Kasetsart University, Bangkok, Thailand
³ Faculty of Agriculture, Kasetsart University, Bangkok, Thailand

Abstract

Ontology plays a critical role in knowledge creation and knowledge management process. It helps make knowledge acquisition and retrieval process significantly more intelligent. The Thai AGRIS Centre, the agriculture knowledge service center of Thailand, has constructed ontology prototype for plant production using rice production issue as a test case. Our results showed that retrieval output from ontology query method delivered Recall value 3 folds as high as the Recall value delivered by the conventional query method. It is therefore advisable that experts or information specialist in each specific knowledge domain should collaboratively start developing their respective ontology in order to help increase efficiency of knowledge acquisition and retrieval process of their respective knowledge community. There are some lessons learned that we would like to share and we hope that they can be tips for those who want to speed up their ontology development process. The guidelines are as followed: Firstly, try to identify the purpose and scope of your ontology unambiguously; secondly, find out existing principles and knowledge base available in that knowledge domain for conceptualizing concept and relation; thirdly, try not to reinvent the wheel by studying and reuse existing ontology; fourthly, defines and categorizes concepts and relations in a clear hierarchically order and define relationship between concept as specific as require; fifthly, make consistency for term representing concept; sixthly; develop guideline and criteria for each ontology construction; seventhly; create multi-lingual ontology to enable knowledge sharing worldwide; eighthly; use expert approach to build a new ontology and use semi-automatic approach to update it; and finally evaluation should be done from the intended users point of view.

Keywords: Rice Production Ontology, Guide line, Lesson learned

Introduction

As rice is a major agriculture product of the country. The rice research database and related useful information are currently kept in each institutional knowledge repository and are mostly provided on world wild web. The retrieval process is directly carried through metadata. However, the existing methods and patterns for information retrieval do not allow us to accurately and comprehensively retrieve information. These constraints become critical if we are to perform agricultural knowledge management activity especially that process of knowledge retrieval. The conventional method of information retrieval process allows us to clumsily access each knowledge repository in a linear manner whereas, from the user’s point of view, an agricultural knowledge issue is always linked to network of related topics that
should be understood as interrelated subsystems. The efficient production of a plant, for example, is closely related to its cultivation method, varieties, pest and ecology as condition of soil, water, climate, and other factors in its environment. Therefore, naturally if users are to acquire knowledge on and learn about any plant production most efficiently, it means that they should be able to access to knowledge repository on that plant production topics in aforementioned integrative manner.

Ontology is an explicit specification of a conceptualization (Gruber, 1993). Ontology has hierarchically structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base. Ontology provided an organizational framework of the concepts in a system of hierarchical and associative relations that allows reasoning about knowledge. Basically, ontology is known as a medium of human expression and a man-made tool to communicate with machines. Presently ontology is one of the most interesting components for knowledge management, especially in terms of retrieval efficiency and knowledge sharing. Its ability to access knowledge through varieties of relationship between concepts serves well the need of users who want to access required information and knowledge, in a world so rich of information like now, more efficiently and comprehensively.

Thai National AGRIS Centre, as national agriculture knowledge service sector, have been embarking on the efforts to improve the efficiency of agricultural information and knowledge retrieval process, especially on those research repository, for the last decade. Our latest initiative was to develop ontology prototype for plant production using rice production issue as a test case. We anticipated that our lesson-learn on rice production ontology construction will be a useful starting point for other related issues in the broader agricultural domain. This paper aims to articulate those lessons learned and share them with people who are interested.

Our result

We created Thai Rice Production Ontology (RPO) to help make process of knowledge acquisition and information retrieval more efficient and intelligent. The rice production ontology comprises more than 3,500 concepts, 5,500 terms and 45 relations. From the study, 9 criteria for constructing ontology and 5 rules for semi-automatic ontology maintenance were derived. We found that more than half of the concepts (68% of concept in RPO) were found in AGROVOC Thesaurus.

We evaluate RPO by using concept and relation for information retrieval. We collected search queries from research policy makers as test cases. The results showed that retrieval output from ontology query method delivered Recall value 3 folds as high as the Recall value delivered by the conventional query method. The Precision value from ontology query method is 1 and the Recall value is 0.9, whereby the Precision value and Recall value from the conventional method are 1 and 0.3 respectively. (Thunkijjanukij, 2008)

Our approach

Our experts have manually created Rice Production Ontology from scratch. We selected expert approach for ontology construction even though it was costly and time consuming because we would like our ontology prototype to be accurate, reliable and comprehensive.

The construction process was divided into 4 stages, as follow:
1. **Specification** - The scope of Rice Production Ontology (RPO) is defined by focusing solely on rice production. The objective is to increase the efficiency of information retrieval and the management of knowledge about rice production in Thailand.

2. **Conceptualization** – This is the stage at which the concepts have been defined. And needed conceptual frameworks have been applied to help define those concepts, terms, and their relations.

3. **Formalization** – This stage is to arrange concept and relation in the format that will clearly express its meaning.

4. **Evaluation** - The evaluation of rice production ontology has been carried out as followed:
   a. Evaluate ontology validity by experts’ consideration and correction
   b. Evaluate its utility for information retrieval by comparing its retrieval result between using ontology query and conventional query. The comparison is measured in terms of Precision and Recall.

**Lessons learned**

Having learned from our five-year experiences in developing Thai Agriculture Thesaurus by using English AGROVOC as prototype, we can short cut the process to one year when we assisted Laos to develop Lao AGROVOC. Some limitations of and problems with the AGROVOC we encountered during the last few years of development made us realized that it is vital to develop Thai Agriculture Ontology if we want to significantly improve the efficiency of agricultural knowledge acquisition and retrieval process (Thunkijjanukij, 2005), thus the initiative to pilot on rice production ontology construction. Having gone through the process, we are certain that our lessons learned from the construction will be an excellent guideline for developing other plant production ontology and, as a result, will significantly, help reduce time and cost for development. Our lesson-learned can be grouped into 5 topics namely: Specification, Conceptualization, Knowledge acquisition, Construction Process and Evaluation.

1. **Ontology Specification**
   - Domain and scope of ontology must be specific and clear cut.
   - At the start, usually, the most basic questions need be answer are: (Noy and McGuinness, 2001)
     - What is the domain that the ontology will cover?
     - For what purpose we are going to use the ontology?
     - For what typed of questions the information in the ontology should provide answers?
     - Who will use and maintain the ontology?

As for this study the scope is defined by focusing solely on rice production as a domain case study. The objective is to increase the efficiency of research information retrieval and the management of rice research knowledge in Thailand. Our targeted users are research administrators who always ask questions such as “Is there any research article concerned with the recent research problem?” , “Which issue is needed for further research?”, “Is the new research proposal a replicate or an overlap with any previous research?” and etc. In this case, therefore, the ontology should be included Scientific name and their Taxonomy of plant and
organism, physiology, anatomy, research methodology and all of plant cultivation and protection technical terms.

For the same scope of ontology but different targeted groups such as farmers who are looking for knowledge on how to deal with their critical problems about rice diseases and pest, additional development to the ontology would be focus on, for example, local names of diseases/pests and their synonym, techniques and method for protection and control, rice pest/diseases resistant variety name and picture represent concept of pest or diseases symptoms, etc.

2. Conceptualization
   • Conceptualizing concept and relation should begin with existing principles and knowledge base available in that knowledge domain. We found that to identify and define the concepts and relation pertinent to aspects of the knowledge domain specified, it is advisable to look for any form of knowledge repository as a good start for references.

For example, to develop RPO, we created the rice production conceptual framework based upon framework presented in Apply theories of plant production and rice ecology (De Datta, 1961 and Rice Department, 2007), then, applied agriculture information classification system of AGRIS (AGRIS/CARIS Subject Categorization Schemes, FAO) as a guideline for defining concept categories.

3. Knowledge acquisition
   • Try not to reinvent the wheel

At present, there is a number of ontology built for different purposes. Studying their patterns and concepts will facilitate our conceptualization process for there will be samples of guidelines and criteria ready for use.

The construction of RPO has gone through that process. We studied from existing ontology, related literature and accessed to rice knowledge resources; such as

- Related text books, web sites, interview and capture knowledge from rice experts.
- Agriculture Thesaurus such as: AGROVOC Thesaurus (http://www.fao.org/agrovoc) and Thai Agriculture Thesaurus (http://pikul.lib.ku.ac.th)
- Related ontology
  o Rice ontology : ontology specialized for genome informatics of rice (http://www.ro.dna.affrc.go.jp/)
  o Plant ontology : ontology of plant-based knowledge domains, such as anatomy, morphology, development, traits and phenotypes (http://www.plantontology.org)
  o CWR ontology: Ontology for crop wild relatives (http://www.cropwildrelatives.org/)
Reuse existing ontology and thesaurus
Although we started building Rice Production Ontology from scratch, we nevertheless concluded that reusing the existing ontology and thesaurus can be productive, and a lot more cost and time saving. Especially, as we found that more than half of concepts (68% of concept in RPO) were found in AGROVOC Thesaurus. Therefore, it should be more efficient if we try to borrow and apply concept and term from the AGROVOC Thesaurus in the first place. By the same analogy, a really rich of combined vocabularies from the AGROVOC and the existing plant, animal, fungi, bacteria, virus and organism thesaurus will be the best resourceful place where we can start building an agriculture domain specific ontology.

4. Ontology construction

- Avoid confusion by starting defines concepts and relations in a clear hierarchically order.

For example, the categorization of concept and relation of RPO was performed as below;

  a. Concept categorization.
     - Divide concept in to 2 groups as concepts of organisms and non organisms as name entity and concept that represents things in classes or type as common word.
     - Concept of organism, such as plant, animal, bacteria, fungi, and virus should be defined according to their ordering taxonomy.
     - Concept of thing or non organism which is not classified in taxonomy will be made into concept according to its class ordering that it belongs

  b. Relationship categorization.
     Relation in RPO were defined in to 3 group, as following
     - Relationship between concept and concept as Vertical relation and Horizontal relation
     - Relationship between concept and term.
     - Relationship between term and term

- Define relationship between concept and concept as specific as need.
  Too broad relationship between concept, such as isSubClassOf or isPartOf, is sometimes not specific enough for utilize ontology to answer question like “Are there any pest insects has the same Order of rice green leafhopper (has scientific name Nephotettix virescens), and what are they?” (Nephotettix virescens is an important rice pest insect, its Order is Homoptera. There are many kind of pest insect in this Order). If we use the general relation such as “isSubClassOf” to define the relationship between concept of Order, Family and Genus, we will unable to answer this question easily.

- Make consistency for term representing concept
As it is necessary to use word or term to represent concept and we must avoid confusion and miscommunication, we therefore must have a consistent criterion in selecting a proper term. We should not use both a singular and plural as main term for representing concept. And we suggest using a singular noun as a representation.

- Develop guideline and criteria for each ontology construction
  Having a set of guideline and criteria for ontology construction can be a big advantage. It can accelerate the construction process significantly. Besides, it will help set up standard for merging similar ontology in the future. Therefore, it is reasonable then to expect that a guideline and criteria for create a specific ontology will considerably help facilitate the ontology construction for a related issue in the broader domain. For instance, ontology for rice production of Thailand will be a very useful prototype for the construction of ontology for the whole Thailand’s plant production knowledge base etc.

- Create bilingual ontology to enable knowledge sharing worldwide
  It is advisable that we should develop bilingual ontology based upon terms in local and international languages, so that it can be communicated, shared and applied extensively. As RPO we created a bilingual ontology in Thai and English.

- Create the first ontology by expert and update ontology by semi-automatic approach
  The high quality ontology should first be built by experts and then updated by semi-automatic approach. This is because ontology is a representative for specific knowledge domain; hence the necessity to bring in experts in that respective domain to build quality ontology that can accurately represent knowledge. However, as the cost of constructing and maintenance ontology via expert is especially high in terms of both time and expenses, the semi-automatic approaches will be attractive for updating and maintaining ontology. By semi-automatic approach also uses computer program together with man who is domain specific expert or information specialist to accurate rules and verify or confirm in process of ontology maintenance.

- Formalize ontology requires efficient tools
  In the stage of arranging concept and relation in the format that will clearly express its meaning, we need efficiency and user friendly tools. In our case study, to formalize concepts and relation for RPO, we used AGROVOC CS Workbench, FAO. The system provides a web-based integrated environment for the development and management of agricultural concepts. The workbench allows multiple distributed users to accomplish a range of tasks covering every stage of terminology development. The web-based workbench is intended for distributed information management specialists like terminology experts, editors and ontology experts to provide the means to maintain the CS. It shall facilitate the editing of concepts, terms and relationships, terminology and translation work as well as the organization of the workflow and communication amongst the editors. (http://naist.cpe.ku.ac.th/agrovocworkbench/)
The AGROVOC CS Workbench is very user friendly. However, since the tool is still in development process it lacked of some necessary functions such as "drag and drop" functionality for changing or moving position of the whole hierarchical nodes within the same ontology, graphical visualization which able to view all necessary ontology nodes in the same window.

5. Evaluation

The evaluation should be done from the user point of view. Evaluation is important to guarantee the quality of the resulting to the ontology’s users. There are two kinds of evaluation, Technical evaluation and User assessment (Pinto and Martins, 2004). Ontology which is created by expert should make a focal point for evaluation to the user and be judged from the user point of view. Since evaluation process for verification to guarantee correctness according to the accepted understanding about the domain of specialized knowledge sources, has already proved by the experts during their construction process. For RPO, the usability and usefulness is to increase retrieval efficiency for rice research knowledge management. Results of the study showed that the RPO can increase efficiency as the completeness of information retrieval up to 3 folds from the conventional method and enable to answer the questions of research administrator while the conventional search query can not.

Discussion

Finally, we would like to propose a few ideas for those who are embarking on building ontology. Our experience demonstrated that Ontology plays a critical role in knowledge acquisition and knowledge management process. It helps make knowledge storage and retrieval process significantly more intelligent. So it is very reasonable to encourage the construction of many more ontology. Experts and information specialist in each specific knowledge domain should collaboratively start developing their respective ontology and also relevant criteria and rules for construction and maintenance. What then follows is the need for efficiency tools for construction a new ontology, by doing the transfer and merging the existing ontology. Those are the fields of research we should further focus.

Acknowledgement

The content of this paper was actually derived from our experience in rice production ontology development. We don’t think it is possible to finish this paper without valuable cooperative efforts from some of our advisors and colleagues. We thank Johannes Keizer for his invaluable advices. We do very much appreciate technological supports from Margherita Sini, Thiranan Damrongson, Dussadee Thamvijit and Aurawan Imsombat. We thank Director of Plant Varieties Protection Division, Surakrai Sungkasubuan and experts from Rice Department for their information support and also Sudjai Lochaloen, Ratchanee Pattaravayo and Chanattika Kamdee for assisting us in data collection process. Finally, we are grateful to Suriyon Thunkijjanukij for his help in editing. Without his assistance, this paper would be unreadable by targeted readers. Be that as it may, it was our team’s grave responsibility to articulate and present the lessons learned in this paper, therefore if there is any error or suggestions please feel free to it with us.
References


