Knowledge Models in Agropedia Indica

Margherita Sini[1] Vimlesh Yadav[2] Jeetendra Singh[2]
Vikas Awasthi[2] Prabhakar TV[2]

[1] Food and Agriculture Organization
Viale terme di Caracalla
Rome, Italy
+39 0657056805
margherita.sini@fao.org

[2] Indian Institute of Technology Kanpur
Kanpur
India
+91 5122594005
{vimlesh, jsingh, vikasa, tvp}@iitk.ac.in

ABSTRACT
This paper refers to the creation of multilingual agricultural knowledge models (KMs) in the context of the Agropedia Indica project. We present the reasons and the requirements for the development of such KMs, the choices made in terms of modeling tools and modeling solutions, and we detail the content of some of the models.

Keywords
Knowledge modeling, Agropedia Indica, Agricultural Knowledge Models, Knowledge Maps, Ontology, CMap Tool, OWL, KMs.

1. INTRODUCTION
Several services are already available in India in order to help farmers and agricultural-related stakeholders find or disseminate information on their activities and products.

However, none of the existing services make extensive use of the new semantic techniques: no semantic searches are available allowing concept-based searches, language-independent platforms for knowledge navigation (most of the farmers do not speak English), no inferencing or reasoning capabilities are available.

The Agropedia Indica project was recently started in order to provide a national entry point for agricultural related information and be enriched with “smart” services with the use of semantic technologies.

Agropedia is a comprehensive, seamlessly integrated model of digital content organization in the agricultural domain. It aims to bring together a community of practice through an ICT mediated knowledge creating and organising platform with an effort to leverage the existing agricultural extension system.

Within Agropedia Indica there are several elements to consider: knowledge objects, knowledge models, interfaces, delivery mechanisms, etc. (see next section). This document explains in detail the issues related to the creation of the knowledge models.

2. AGROPEDIA INDICA
The Agropedia Indica project aims to develop a national reference portal for Agriculture in India, making use of modern technologies and providing domain-specific and user-specific services. The objective is to make available Agriculture Knowledge repositories of universal knowledge models and localized content (built in collaborative mode and in multiple languages such as English, Hindi and Telegu) for a variety of users, with appropriate interfaces.

Two different types of elements are key for the system:

- Knowledge Models: mainly used to navigate agricultural knowledge and to organize and search agricultural content; KMs have been designed with the intention of using them for indexing and browsing the content that we gather in the repository. These KMs are the structural representation of knowledge by using symbols to represent pieces of knowledge and relationships between them, which can be used to connect seamlessly to the knowledge base in Agropedia using semantic tools.
- Knowledge Objects: every type of resource related to agriculture, such as documents in various formats (pdf, word files, txt files, etc.), video files, audio files, pictures, etc.

Currently, a draft release of the first version of the system is available[1], but further developments are in progress. Services available in the first release of the portal include:

- Concept-base indexing and concept-base document retrieval: the KMs are used for this purpose.
- Navigation of knowledge maps: different formats have been used for processing (OWL, SVG) and visualization (jpeg, pdf).
- Extension material: this space of Agropedia portal deals with the certified contents added mainly by the agricultural scientists of the consortium partners of the project in the form of 'Library' content, 'Crop Calendar' or 'Dos and Don'ts' on the nine selected crops- Chickpea, Groundnut, Litchi, Pigeonpea, Rice, Sorghum, Sugarcane, Vegetablepea, and Wheat.
Library: The document like contents.
Crop Calendar: Month-wise package of practices of agricultural crops
"Dos and Don'ts: Important information about what should and should not be done during crop production & allied fields.

[1] http://www.agropedia.net

– Interactions: this is the social interactions space for the Agropedia users, which is based on web 2.0 technology. In contrast with the extension material it deals with contributed knowledge coming up through the active involvement of the users in agrowiki, agroblog, Forum and agrochat.

– Agrowiki: agrowiki is related to agricultural domain where Wikipedia is dealing with all types of information. agrowiki is easier to use than Wikipedia especially in case of create a wiki page. It provides users a word document with a toolbar having all possible functionalities which are commonly used by them in creating a document. The authenticity of content in case of wikipedia is doubtable whereas in Agropedia we have certified content under the category Extension Material.

– Personal space for agricultural experts: user can create their own profile and blog etc.

3. KNOWLEDGE MODELS FOR AGRICULTURE

3.1 Requirements and scope

Current, indexing systems frequently use keywords from traditional thesauri to tag resources and to allow easy retrieval. Agropedia Indica wishes to be a more modern system in which resources are tagged with the URI of concepts rather than terms (or keyword) in a particular language. This allows several benefits, between which, the use of any synonyms or any language in user searches.

Currently, with the aid of an agricultural expert we developed the following models:

1. A generic map, acting as a top level foundational crop ontology;
2. A specific map about the rice crop, including:
   a. pests;
   b. diseases;
3. A map on Pesticides;
4. A map on cropping systems and their agro-climatic zones;
5. A time-line based map for paddy crop activities.

Additional maps that are being planned are:

a. pests (generic map);
b. crop activities;
c. fertilizers;
d. Indian geographical areas;
e. agricultural experts;
f. others as needed.

3.1.1 Knowledge maps or Ontologies?

For the purpose of this paper, knowledge models and knowledge maps [3] are generally treated as synonyms. Currently the knowledge models at this stage are closer to topic-ontologies rather than schema-ontologies [5]. The modeling primitives include: classes, individuals, datatype properties, individual properties and literals [4]. We plan to increase the complexity of the model in our subsequent releases, to convert them into full fledged ontologies requiring the full modeling power of the OWL Web Ontology Language (OWL).

3.1.2 Representation languages

KMs can be represented in several ways: we can even just draw them on a piece of paper. However, when we move to the level of being able to programmatically access and exploit them, we need to represent them with a representation language.

Sometimes, a simple textual file may be enough: the requirements are the ones that drive us on the choice to make. In case of Agropedia Indica, we have decided to use OWL.

The main reason is that OWL is a W3C recommendation gaining increasing interest in the international community for semantic-based systems and it is machine-processable. Many ontologies are already available in this format, which makes an OWL-based system open to interoperability with those.

In addition to that, OWL allows straightforward data processing and visualization, with the many existing APIs able to process it (Jena[2], Protégé, etc.). Other existing tools and methods can be used to handle the KMs, allowing reuse for local applications. OWL is also web-enabled which permit easy data reuse and transfer, and therefore easy ontology linking and mapping.

Based on this choice, experiments were conducted for exporting from the suggested KMs editor to OWL [7].

3.2 Methodology

Building consistent and coherent knowledge models requires specific competences which cannot be found in many domain experts. We decided therefore to build a first set of models, and use them in workshops and training courses.

The choice of a standard language as the representation language for the system data (OWL), allows (in addition to other benefits) the team members that needs to contribute to the population of the maps, to be able to chose any editor able to save or export their data into OWL. Users are therefore not forced to use any specific system.

However, several tools have been investigated (such as Protégé or Swoop), but most of them were found to be too complex for non ontology-experts.

The suggested tool used during workshops and training to domain specialist was the Concept Ontology Editor (COE) tool version 4.11 available within the CmapTools software [6]. Cmap version 4.18 has also been tested, but due to its inability to export to OWL it has not been chosen.

The most important advantages of COE over other tools are:

- easy interface;
- easy of use;
- possibility to export data to TXT, XML, OWL, and SVG formats (see section 3.5).

In order to allow domain experts to use the COE tool properly (in the sense of allow correct representation of the information, and coherent feature use for a consistent export to OWL), specific guidelines have been prepared and distributed to project members.

[2] http://jena.sourceforge.net/

3.2.1 Guidelines

KMs or ontologies for the Agropedia Indica portal could be prepared using several tools. The only constraint, for easier integration with other models and incorporation in the Agropedia portal, is that those tools needs to able to save or export to OWL. However, due to lack of specific competences (on tools and semantic technologies) by domain specialists, the COE tool from the Institute for Human and Machine Cognition[3] (IHMC) has been chosen.

Domain experts include experts on pesticides, on specific crops, such as sorghum, or mango. Each of them has been trained on the creation of maps using the mentioned editor, and, in order to allow them to create consistent maps, they have been provided with specific COE-oriented guidelines.

These guidelines include:

- guidelines for maps in general: these explains what should be the objectives of the maps, who and why they should be developed. Indications on how to create links between maps is also included.
- guidelines for concepts and instances: includes indications on the types of concepts or instances should be created, how to distinguish them, what should be the mechanism for labeling, etc.
- guidelines for relationships: indicates how relationships should be created, named, and how to make the difference between concepts relationships and instances relationships.
- a section on "Good and Bad modeling": examples of real agricultural uses-cases are listed, with some examples on how experts have implemented them during workshops and how it is instead suggested to be.

The guidelines also guide users to obtain the maximum of expressivity (while exporting to OWL) keeping the complexity of the representation as minimal as possible.

This is why, despite the large types of relationships that are provided within the COE tool, we have decided to start with only four different types of relationships:

- "are", which correspond to the subclassOf;
- "is a", which represent an instance of a concept;
- DataTypeProperty, which link individuals to data values;
- ObjectProperty, which link individuals to individuals.

An object property is defined as an instance of the built-in OWL class owl:ObjectProperty. A datatype property is defined as an instance of the built-in OWL class owl:DatatypeProperty. Both owl:ObjectProperty and owl:DatatypeProperty are subclasses of the RDF class rdf:Property.

DataTypeProperties and ObjectProperties may be assigned with a meaningful name. A registry of already available defined agricultural related relationships such as "afflicts" "hasProductionPractice", "hasPostProductionPractice", "isDerivedFrom", "hasPest", has been provided to users.

The figure below gives an example of the guidelines for distinguishing relationships at concept level and at instance level within the CMap tool.

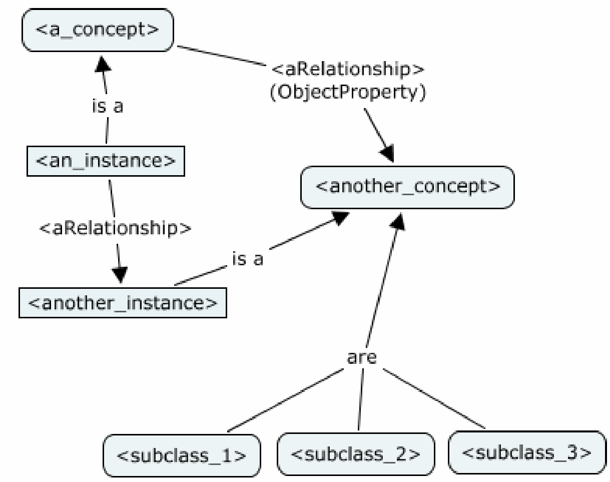

Figure 1: Example of different type of relationships in the COE tool: between two concepts, between two instances, subclassOf, and instanceOf.

A particular aspect of the provided guidelines covers the possibility of linking different maps: this is one of the important and crucial elements of ontologies or KMs, as it will allow the navigation of knowledge across maps developed for different purposes and by different domain experts.

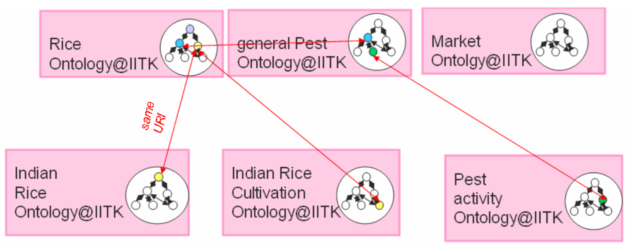

Figure 2: Reuse of URIs in different maps (in the COE tool realized by using the link functionality).

3.3 The Knowledge Models

Given the purposes of the KMs (see section 3.1), and for better organizations of the information, we have decided to organize the KMs in the following ways: a top level generic map for crops (which we call the "Foundational Agricultural Crop Ontology"), and many specialized maps based on specific topics, such as rice, diseases, pesticides, etc.

We could represent some maps in OWL-Lite, some in OWL-DL, but some of them have been exported to OWL-Full; all of them load correctly in Protégé. With some few refinements, OWL-Full KMs may be converted to OWL-Lite ones, in general preferable for the desirable level of representation and inference capabilities in Agropedia.

---

[3] http://www.ihmc.us/

3.3.1 The Foundational Agricultural Crop Ontology

The Foundational Agricultural Crop Ontology contains information common to every type of crops. This information include, among others:

- production practices (production technologies and protection technologies);
- post production practices (related to harvesting, threshing, post harvest technologies, marketing, etc.);
- environmental information (climate, soil);
- varieties and cropping systems;
- botanical description;
- origin (geographical areas).

Many of the Foundational Agricultural Crop Ontology's concepts are common to all crops. It contains also few instances common to many crops, such as plant nutrients (e.g. Nitrogen, Phosphorus, Potassium, etc.).

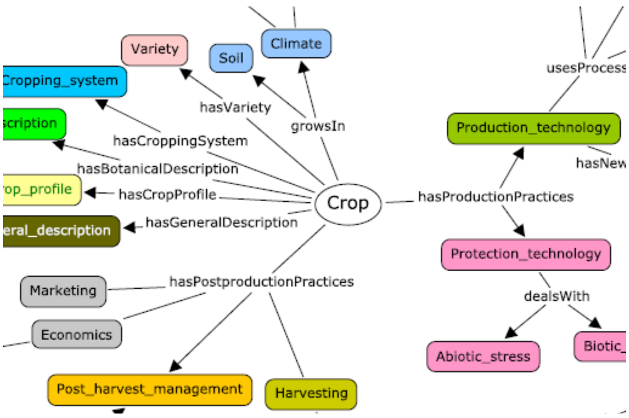

Figure 3: Snapshot of a part of the Foundational Agricultural Crop ontology in Agropedia Indica.

The map is OWL full, contains 76 classes, 16 individuals, 19 object properties, and 5 datatype properties.

Concepts and instances from this map are then re-used in specialized maps, such as the Rice map.

3.3.2 The Rice Knowledge Model

The KM about Rice [1 and 2], is currently structured in three different maps, in order to better organize the different areas related to rice: a generic map on rice crop (including information such as production and post production practices, including rice field preparation and seed and sowing of rice, harvesting and threshing, rice post harvest management, and rice cropping systems); rice pests; and rice diseases. With the help of other scientists we can enrich those maps or prepare new ones with other rice related information such as rice products, rice varieties, rice nutrient management, etc.

All maps related to different aspects of rice (see pictures below) reuse many concepts from the Foundational Agricultural Crop Ontology and extend them with specific information related to this crop.

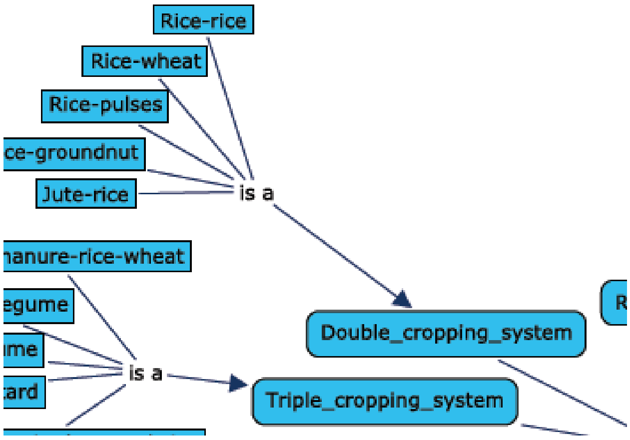

Figure 4: Rice cropping systems in Agropedia Indica.

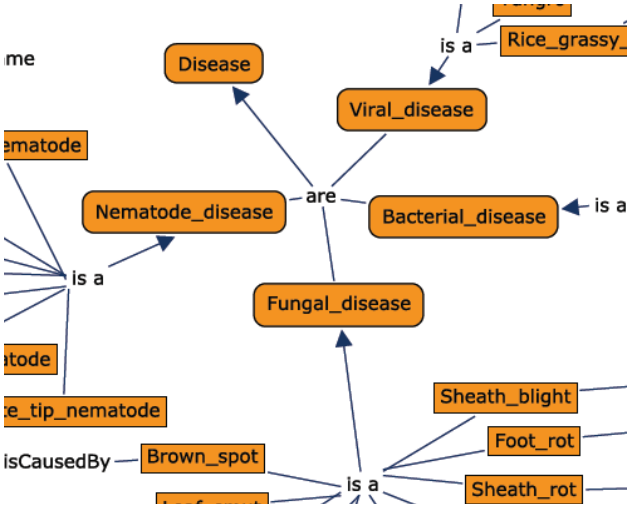

Figure 5: Rice disease management in Agropedia Indica.

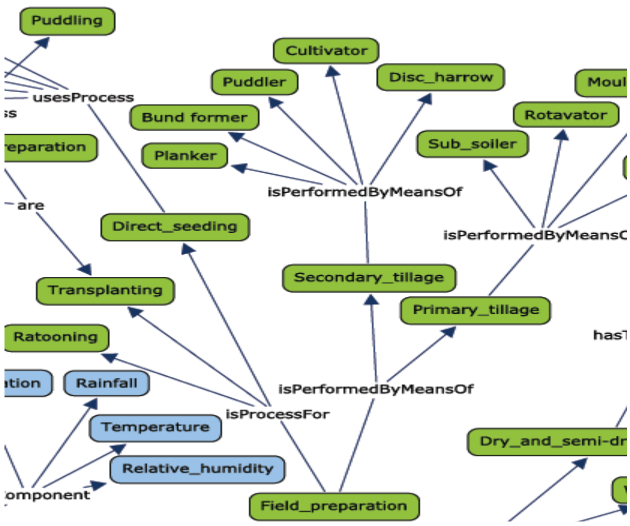

Figure 6: Rice field preparation in Agropedia Indica.

The combined rice map is OWL full, contains 179 classes, 116 individuals, 34 object properties, and 3 datatype properties.

3.3.3 The Pesticide Knowledge Model

Pesticides can be used in many different crops. For this reason they have been organized in a domain-specific map (the pesticide map). Here, the main three subclasses are the three different types of Pesticides: Herbicide, Fungicide, and Insecticide. Each of them has been further divided in many subclasses (e.g. Thiocarbamate herbicide, Phenoxy carboxylic, and Dinitroaniline as subclasses of Herbicide; Fumigant, Chlorinated hydrocarbon, and Pyrethroid as subclasses of Insecticide). Each subclass is also expanded with the list of individuals (e.g. Chloropicrin, Methyl bromide, and Phosphine as instances of Fumigant).

Note that for the provided guidelines all concepts and instances names have been given in singular and capitalized.

An advantage of this map, is not only to list and categorize hierarchically all types of pesticides used in agriculture, but also to act as a reference when preparing specific crops' maps. For example, we can link rice diseases (from the Rice disease map) to specific pesticides from this map, by indicating which disease can be controlled by which fungicide.

The pesticide map has also two other important features: it contains the indication of the name of the pesticide used in the market (e.g. Bromacil hasTradeName Hyvar-X). This information can be used by farmers when they need to buy a specific product against a specific pest. Also, it contains information about which bio-pesticide derived from specific plants.

For the purpose of this map, the trade name has been created as literal. Note that we may have multiple trade names for a specific element (e.g. DDT hasTradeName Tafarol, which is the main DDT produced and available in India, but it has other names given by other companies that produce it).

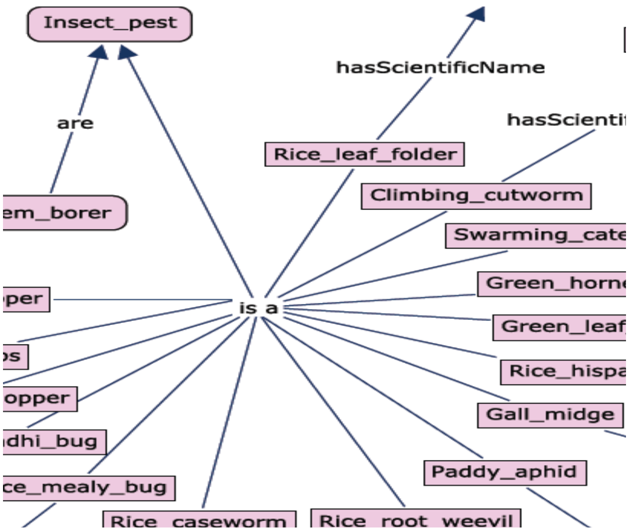

Figure 7: Snapshot of a part of the Pesticide KM in Agropedia Indica.

The map, OWL Full, contains 37 classes, 109 individuals, 4 object properties, and 1 datatype property.

An extension of this map would be to convert the trade names of the products into individuals and connect them to the name of the company which produces them.

3.3.4 A temporal-based Knowledge Model: the paddy crop cycle

As mentioned in section two, one of the objectives of Agropedia Indica is the provision of (language-independent) advanced services to different type of users.

It is known that farmers are interested in the activities associated with crops and especially to the temporal aspect for those activities; competency questions [8] for those farmers may be: when do I have to first irrigate the crop after seed sowing?, when will the seedling of my rice crop be ready for planting in the main field if I use the dry bed method for nursery raising?, my crop is 30 days old - what should I do now?

Based on these and similar competency questions the knowledge map that we prepared details:

- the phases or activities from nursery bed preparation to storage of the final product;
- the days in which these activities should take place;
- the possible methods or instruments or system that can be used for a specific activity;
- quantities (of seed per method of sowing, of chemicals for disease control, of moisture contained in grains at the time of storage, etc.);

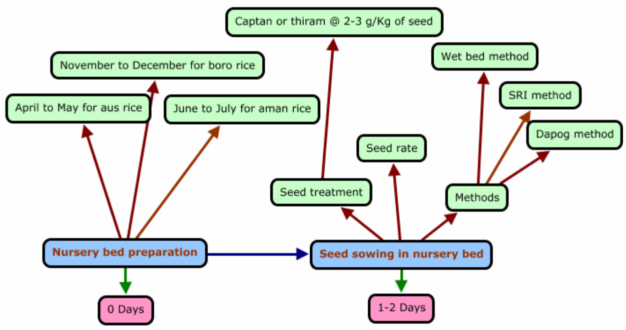

Figure 8: Snapshot of a part of the Paddy crop cycle timeline KM.

Based on this map, and by specifying temporal relationships between the activities such as "precedes" and "follows", it would be easy to infer how many days an activity has to be carried out compared to previous activities (e.g. First weeding occurs 25-30 days after transplanting, First irrigation occurs 35-45 days after transplanting and 10-20 days before First weeding).

3.4 Modeling solutions

During the preparation of the KMs, we had to take some modeling decisions.

Generic maps will be mainly composed of those concepts which are common to many specific domains; e.g. crops contains concepts such as "Seed", "Irrigation", "Drainage", as they are common to many crops such as Rice, Sorghum, or Mango.

Specific maps such as Rice will then contains specific elements associated to this type of crop, but which can be related to a low-level concept in the Foundational Agricultural Crop Ontology (see figure below).

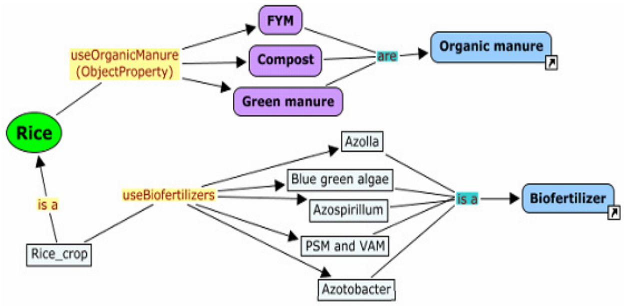

Figure 9: Connecting Rice map to generic concepts from the Foundational Agricultural Crop Ontology map (e.g. Organic manure and Biofertilizer).

Other modeling decisions were taken for more complex modeling issues, such as mathematical expressions. In the map related to Marked (just started and not completed) we had to identify these situations:

- "gross income" calculatedBy ("yield" multipliedBy "price")
- "net income" calculatedBy ("gross income" minus "cost of production")

In this case every element of the expression has been identified as an operand. Being concepts, it will always be possible to reuse them in other maps.

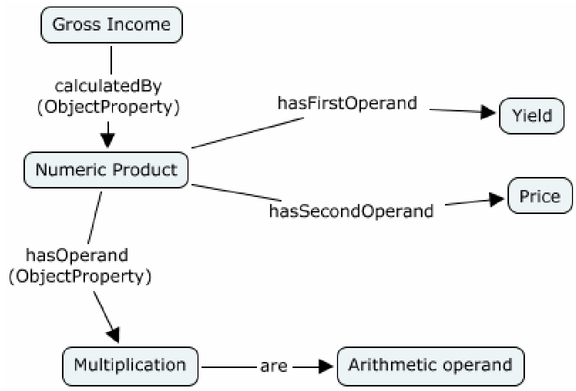

Figure 10: Representing mathematical expressions in Agropedia Indica.

3.5 Visualizing Knowledge Models

We are currently investigating several ways for processing the KMs, for the purpose of giving easy graphical navigation of those to Agropedia users.

The current technological solutions approached are:

1. display the KMs as a tree: this solution has the disadvantage that we are constrained to display only hierarchical relationships;
2. display the KMs as a picture: export all maps to the Scalable Vector Graphics format (.svg) and allow users to graphically navigate them;
3. display the KMs using "ontology-oriented" graphical solutions: objects of the KMs (concepts and instances) will be dynamic and would allow further visualization of related objects (this solution will allow not only graphical navigation but also the visualization of additional information associated to concepts or instances); this solution can technically be implemented through:
   a. use Jena to process the exported OWL files and visualize them in Adobe Flex for graphical navigation;
   b. use Jena to extract KMs information and use the open source Prefuse[4] visualization toolkit for image-based graphical navigation.

Solution (1) can be easily rendered in technologies like JavaScript and is the most efficient. Solutions (3) seem to require large downloads and higher frontend clients. Performances and usability of these solutions are still under investigation.

3.6 What's next: synonyms and multilingualism

As mentioned, multilingualism and richness of lexicon are keys issues within the Agropedia Indica project. The current KMs are being developed in English. Synonyms may be represented as literals associated to existing concepts or instances. However, translations of maps will be done via the use of the AGROVOC thesaurus and AGROVOC Concept Server [9 and 10] which contains all agricultural terminology in Hindi. A Telegu release will be also ready soon.

AGROVOC already provides web services which can be used for our purposes.

4. A concrete application of the KM

In this project, KMs are created by domain expert using concept ontology editor tools. These models and then used as the essential element to organize the data pool: indexing is performed by the user after uploading a document. The user is presented with a list of concepts from the Knowledge Models from where he can select one or more concepts to associate to the uploaded resource. The system will relate the URI of the selected concept(s) to the document.

Similarly when the user search for information, can select concepts from the model, and the system will retrieve all resources which have been previously associated to these concepts.

Note that the user can navigate the KMs in his preferred language: the result of the indexi8ng or searching will remain the same, because concept-indexing is language-independent.

5. Conclusions

Agropedia Indica is an attempt to build a national repository for agricultural information based on semantic technologies. In this paper we described the knowledge models that have been developed in Agropedia Indica: a foundational crop generic ontology, and a crop specific ontology (rice); generic models for pesticides and diseases and their crop specific counterparts. Some of the guidelines for developing these models have been described.

[4] http://prefuse.org/

Several possible visualizations of the KMs in the Agropedia portal are under investigation.

The representation language used as "backbone" of the KMs and the tool used to produce them, turned out to be good enough for the purpose of the project.

6. ACKNOWLEDGMENTS

The authors would like to thank the National Agricultural Innovation Project (NAIP), Indian Council of Agricultural Research (ICAR) for funding the project.

7. REFERENCES

[1] *Rice research in India*, Indian Council of Agricultural Research, 1985

[2] *Rice in Indian perspective*, Sharma, S. D., Nayak, B. C. (Eds.), Vol. 1 and 2, 2005, ISBN 81-7019-493-3

[3] Novak, J. D., Cañas, A. J., *The Theory Underlying Concept Maps and How to Construct and Use Them*, Florida Institute for Human and Machine Cognition, Technical Report IHMC CmapTools, Rev 01-2008

[4] OWL Web Ontology Language Reference, W3C Recommendation, http://www.w3.org/TR/owl-ref/

[5] Kiryakov, A., *Ontologies for Knowledge Management*, Semantic Web Technologies: Trends and Research in Ontology-based Systems, John Davies, Rudi Studer, Paul Warren, 2006, John Wiley & Sons, Ltd

[6] Institute for Human and Machine Cognition (IHMC), CMap (Concept Ontology Editor, COE), http://cmap.ihmc.us/

[7] Hayes, P., Eskridge, T. C., Saavedra, R., Reichherzer, T., Mehrotra, M., Bobrovnikoff, D. *Collaborative Knowledge Capture In Ontologies*, Proceedings of the Third International Conference on Knowledge Capture (K-Cap'05), Banff, Canada, pp 99-106.

[8] Allemang, D., *Ontologies, Reuse and Domain Analysis*, TopQuadrant, Inc.

[9] Sini, M., Lauser, B., Salokhe, G., Keizer, J., Katz, S., The *AGROVOC Concept Server: rationale, goals and usage*, Library Review, 2008, Vol. 57, Issue 3, pp.200 – 212, ISSN 0024-2535, Emerald Group Publishing Limited

[10] Yongyuth, P., Thamvijit, D., Suksangsri, T., Kawtrakul, A., Rajbhandari, S., Sini, M., Keizer, J., The AGROVOC Concept Server Workbench: a *Collaborative Tool for Managing Multilingual Knowledge*, Proceedings of IAALD AFITA WCCA, 2008, Tokio, Japan