Towards a novel content organisation in agriculture using semantic technologies: a study with Topic Maps as a tool

Shelly Patwar, Pritpal Kaur, Asil Gerard Sylvester and Venkataraman Balaji*

Knowledge Management and Sharing Group,
International Crops Research Institute for the Semi-Arid Tropics,
Patancheru – 502 324,
Andhra Pradesh, India
E-mail: shelly@cgiar.org
E-mail: pritpalkaur@cgiar.org
E-mail: a.sylvester@cgiar.org
E-mail: v.balaji@cgiar.org
Fax: +91 40 30713074
*Corresponding author

Abstract: Agricultural information management needs to be responsive to the needs of varied stakeholders, and content repurposing to suit multiple audiences is a key requirement. A framework that combines the semantic web techniques of Topic Maps (TM) and a global agricultural thesaurus, the FAO AGROVOC, has been developed on a pilot basis for five crops. The framework when overlaid on a web-based collection of information objects reveals interesting possibilities, especially in presenting the repository content in noticeably different ways to a subject-matter specialist or to an extension worker. Results are discussed and opportunities for further development are briefly described.

Keywords: semantics; ontology; knowledge model; TM; topic maps; content repurposing.

Biographical notes: Shelly Patwar works with the Knowledge Management and Sharing group of the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), which is an international agricultural research centre that works in sub-Saharan Africa, and in South and Southeast Asia. She conducts research in the area of content organisation using semantic technologies for the agricultural knowledge domain. She has degrees in agriculture and in IT in agriculture.

Pritpal Kaur is with the Knowledge Management and Sharing group of ICRISAT. Her interests include information and knowledge management for agricultural domain using semantic technologies. She has degrees in agriculture and in IT in agriculture.

Asil Gerard Sylvester is a researcher in applied IT working with the Knowledge Management and Sharing group at ICRISAT. He is involved in developing semantic technology-related applications in agricultural knowledge management. He has worked on projects in Afghanistan, India and the Philippines, and has presented papers in many international conferences. Also an expert on IT Security and Cyber Laws, he contributes regularly to online forums and magazines.

Venkataraman Balaji is the Global Leader for Knowledge Management and Sharing processes at ICRISAT. He has designed and implemented several IT for rural and agricultural development projects in countries in sub-Saharan Africa and South Asia (SA). His publications include a wide range of papers in many professional fora, book chapters and peer-reviewed journal articles. His work has been profiled in the *IEEE Spectrum* (February 2004 issue).

1   Introduction

The process of information generation in agricultural education and research, and its storage for online access has unique requirements because of the concurrent needs of domain matter specialists, educators and extension workers. Production or generation of content in different streams, and its storage for access in different containers are not feasible in an easy manner. An earlier study (Kaur et al., 2007) had brought out the relative paucity of agriculture-related content in the Wikipedia. It is our view that the relative inadequacy of agricultural content online is mainly due to the challenges associated with the management of content in multiple streams and containers to meet varied stakeholder needs and requirements. On the basis of this, and on the basis of the innovative manner in which semantic web techniques have been applied on a pilot basis in the agricultural domain (Rao et al., 2006), we have proceeded to perform a series of steps that give clear indications of how emerging semantic web techniques (Berners-Lee, 1999), especially that of Topic Maps (TM) (Biezunski and Newcomb, 2001), can be applied to overcome the content management challenges just outlined. In the study described here, which is part of a series of studies, we have made an effort to combine a global agricultural thesaurus, the FAO AGROVOC (http://www.fao.org/aims/ag_intro.htm), as the equivalent of ontology, with particular tools and techniques of TM.

The core approach of TM, namely, of decoupling the information layer from the knowledge layer makes possible to generate various meta-models or knowledge models that could be overlaid on a content repository (Ahmed, 2000). In this study, we start with an attempt to represent a part of the agricultural domain through a set of crop-based knowledge model that facilitate conceptual navigation, and presentation to suit different types of users. We have developed a process by which the AGROVOC terms and relationships are used in the design and construction of knowledge models.

2   Method

The underlying information resource layer for this study is the web-based learning resources collection of the Virtual Academy for the Semi-Arid Tropics (VASAT) (http://www.icrisat.org/vasat). The VASAT is a strategic coalition of field and research agencies for information, communication and capacity building, operating in South Asia (SA) and West and Central Africa (WCA) and focuses on drought matters. It is anchored by ICRISAT (http://www.icrisat.org/), an international agricultural research centre. VASAT generates learning-oriented extension material for the rural organisations of the semi-arid tropics. These resources are created in a granular fashion by experts and cover the production and protection aspects of ICRISAT's five mandate crops, namely, chickpea, pigeonpea, groundnut, sorghum and pearl millet.

As the information layer is meant for agricultural extension workers, the knowledge represented through the maps comprises various concepts that describe packages of production practices, diseases, weeds, pests etc. of a crop. The concepts were first manually collected from the VASAT learning resources collection online. To use a controlled vocabulary for these concepts in a knowledge model, we used an agricultural thesaurus. There are various thesauri available for the agricultural domain such as the CABI thesaurus (CABVOC), NALT, ASFA, GEMET, AGRIFOREST, USAID thesaurus, AGROVOC, etc. For this experiment, we decided to use the AGROVOC of the FAO, which is a multilingual structured thesaurus of all subject fields in agriculture covering forestry, fisheries, food security and related domains. AGROVOC is available in 20 languages and many more translations are being added. Ontology-explicit formal specifications of the terms in the domain, and relations among them used in this experiment were selected from the AGROVOC. Use of controlled vocabulary will help in translating the knowledge models in other languages in which AGROVOC is available, and we wanted to take advantage of this aspect. AGROVOC is thus our choice for this study.

To create knowledge models, most of the concepts, terms and relationships were defined using the ontology extracted from the AGROVOC. The ontology here has been used for 'categorisation', 'lexicalisation', and 'relationship legends' in the knowledge model. We have used the AGRIS/CARIS classification scheme (http://www.fao.org/aims/ag_classifschemes.jsp) of the AGROVOC as it covers general agriculture as well as related sectors such as fisheries, livestock, forestry, etc. The subjects covered in this study on the knowledge structures are related to crops. Thus, main categories populated are plant production, plant protection and general agriculture. The 17 categories thus identified are further divided into respective sub-categories as the second level hierarchy in the thesaurus. The concepts and terms used in the knowledge structure were then manually positioned under suitable subcategories. The terms populated cover the package of practices, diseases, and pests of the five mandate crops of ICRISAT (see above). The relationship legends of the AGROVOC are used as the associations in the knowledge map for each crop. These concepts and relationships are used to create the TM for the crops.

We experimented with the TM approach to create knowledge structures for the agricultural domain. According to Pepper (2000), TM are an ISO standard for describing knowledge structures and associating them with information resources. As such they constitute an enabling technology for knowledge management. Dubbed "the GPS (Global Positioning System) of the information universe", TM are meant to provide powerful new ways of navigating large and interconnected corpora (Rath, 2000). XML Topic Maps (XTM) (http://xml.coverpages.org/XTM-SC34-N0328.html), an XML-based interchange syntax is the standard for TM. Linear Topic Maps (LTM) provides a much simpler way of creating TM.

Figure 1 Modified from (Pepper, 2000) to describe a few sample concepts and relationships.
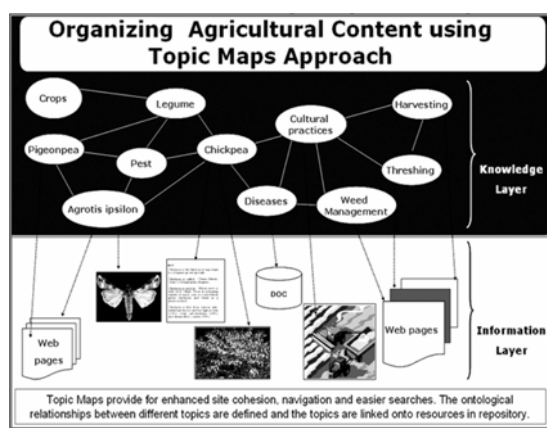
2.1 Software tool for creating the topic maps

The increasing interest on exploiting the potential of TM technology has resulted in several software applications to author, maintain and publish TM such as TMTab, TMNav, Topincs, Wandora, Mapalizer TM4L editor and viewer. We have conducted a brief review of a few of the available software tools for TM and a comparison is presented (Table 1).

Table 1 Brief review of a few topic map software

| Tool	Language	License	Import/Export	Querying	Visualisation
TMTab	Java	Free	Available	Not available	Not available
Mapalizer	Perl	Free	Available	Not available	Not available
Wandora	Java	Open source	Available	Available	Available
Topincs	Java/Lisp	For non-commercial use	Not available	Not available	Not available
TMNav	Java	Open source	Not available	Not available	Not available
TM4L	Java	Open source	Available	Available	Available

To design knowledge models for crops we selected the Topic Map 4 Learning software (TM4L), which is an open source authoring environment that supports the creation, maintenance, and use of TM (Dicheva and Dichev, 2006). We decided to use this editor because of its easy-to-use interface for conceptual structure design and maintenance. Its functionalities include editing, browsing, and combining conceptual structures, coupled with support for relating concepts, linking concepts to resources, merging ontologies, external searching for resources, defining perspectives, etc. The TM4L functionality is enhanced by an interactive graphical user interface that combines a hierarchical layout with an animated graphical view, joined with context-sensitive features. It provides ontology and metadata engineering capabilities coupled with basic document management facilities. The TM4L Editor ingrains well the essential TM features to support easy and effective merger of existing information resources, even as it maintains the meaning structures. This allows for flexibility in the reuse and in extending existing repositories to additional purposes. The meta-structure created in the TM4L editor is fully compliant with the XTM standard and is thus interchangeable and interoperable with any standard XTM tool.

2.2 Topic map construction

The steps involved in creation of knowledge model using the topic map approach involves creation of topics, establishing associations, linking occurrences of topics and applying themes for different users.

Figure 2 Method used to generate crop knowledge maps
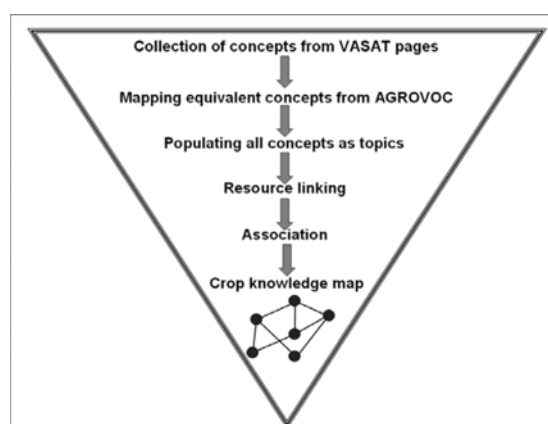
2.2.1 Topics

In our approach, the topic is the fundamental concept in all the knowledge structures. A topic, in its most generic sense, can be any "thing" whatsoever – a person, an entity, a concept, really anything – regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever (Pepper, 2000).

The VASAT learning resources served as the basis to create the crop-knowledge model. The content design in these resources was useful in capturing the concepts and topics covered under each crop. The concepts and terms for the knowledge structure were so extracted from the VASAT learning resources such that they would facilitate the users of the knowledge map in associative navigation and search in the content repository. The concepts then were manually mapped with AGROVOC and equivalent concepts, and terms were discovered from AGROVOC. All these concepts and terms were manually populated as topics covering different aspects of crops like diseases, pests, weeds and packages of practices and were added to the topics list in TM4L.

2.2.2 Occurrences

A topic may be linked to one or more information resources that are deemed to be relevant to the topic in some way.

Such resources are called occurrences of the topic. An occurrence could be a monograph devoted to a particular topic, for example, or an article about the topic in an encyclopedia; it could be a picture or video depicting the topic, a simple mention of the topic in the context of something else, a commentary on the topic (if the topic were a law, say), or any of a host of other forms in which an information resource might have some relevance to the subject in question (http://www.ontopia.net/topicmaps/materials/tao.html).

The occurrences for this experiment are the learning resources of VASAT. We have manually added the resources to the topics from the VASAT online collection as external occurrences. For a few topics in knowledge model, short descriptions, definitions, and names were provided as resources as well.

2.2.3 Associations

A topic association is (formally) a link element that asserts a relationship between two or more topics. The relationship legends provided by the AGROVOC are traditional thesaurus relationships, concept-to-concept relationships, term-to-term relationships, string-to-string relationships that are specially designed for agricultural domain and sub-domains like plant protection, plant production, etc. We have used a mix of these relationships to interlink topics in the knowledge structure.

2.2.4 Themes

The theme (known as scope in TM) enables us to define a context within which some topic characteristics are valid. Themes can help us to use one common knowledge structure, while themes further provide the option for customisation of a Topic Map to suit various audiences.

In this experiment, we have tried to separate content and knowledge structure for two different users, i.e., the subject-matter specialist and the extension worker. A theme named 'extension' has been created where the topic names and the information resources change when an extension worker is the user. The common names are displayed when the extension theme is activated. This way, a knowledge structure once created can be reused for subject-matter specialists or extension workers or other users by using the theme feature. The themed-topic map driven portal can be viewed in different ways by different categories of users (See below for an example).

Starting with a manual intervention and later upgrading to a semi-automatic process, the creation of this knowledge model makes it easier for content authors to be able to contribute to the repository, and a suitable interface such as that based on the Wikimedia could be an option.

3   Results

Key results of this experiment are in structural representation of agricultural knowledge as crop-knowledge models, and in the presentation of knowledge models to suit different types of users. Content aggregation from different online sources has also been tried.

The knowledge model created using the TM4L editor is presented on the browser interface using the Ontopia's Ontopia Knowledge Suite (OKS), which is a free downloadable topic map navigator, powered by Ontopia's engine. As the XTM file generated in the TM4L editor is interoperable, it was exported to the OKS to make a presentation in a web-interface. This interface displays all the topics, associations and occurrences for a crop on a single webpage. The OKS also has other facilities to visualise, author and edit TM.

Figure 3   Generalised topic map structure for the agriculture domain (based on AGRIS-CARIS classification)
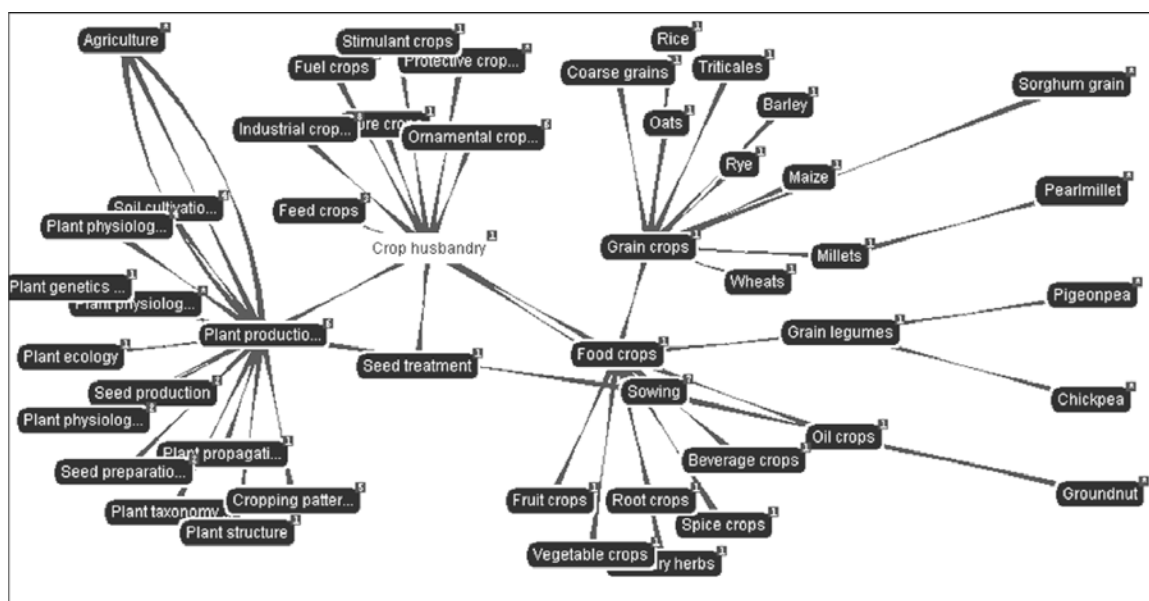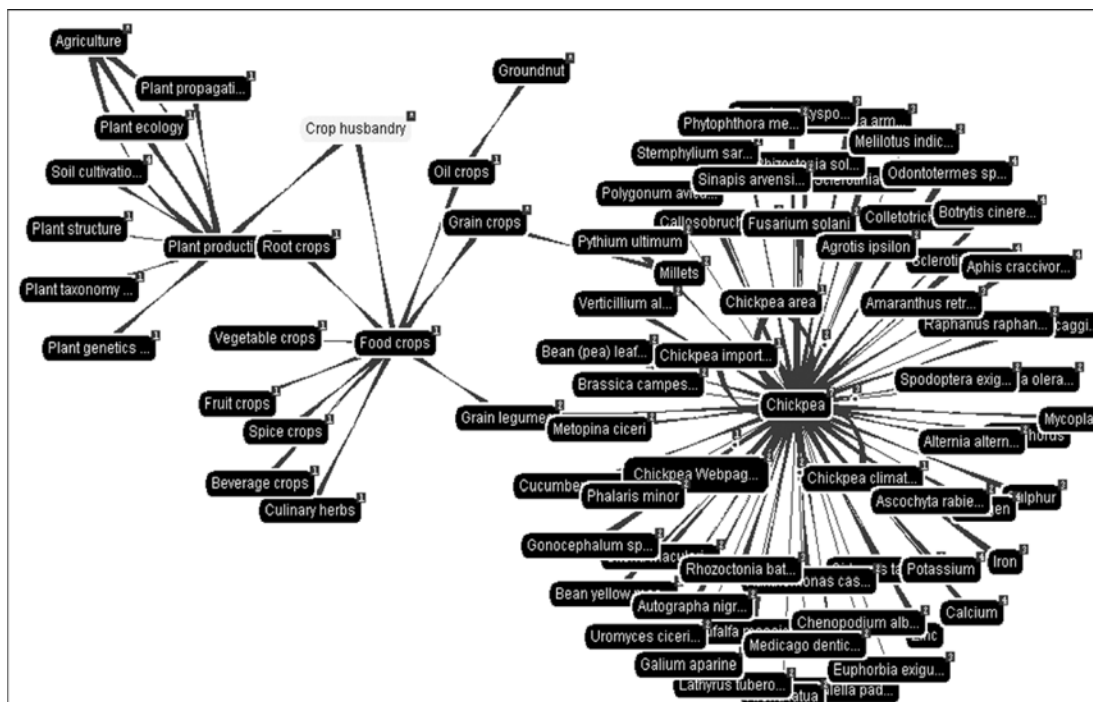
3.1 Knowledge models for five crops

In the initial steps, we created a generic topic map for crops.

Inheriting from the basic knowledge model, more specific crop-knowledge maps for chickpea, pigeonpea, sorghum, groundnut and pearlmillet were developed. These five knowledge maps formed the meta-structure for VASAT learning resources. Individual crop TM can be browsed on the web at http://test2.icrisat.org/.

Figure 4 Chickpea knowledge map generated from generalised topic map

3.2 Customisation according to users (Scope)

With the help of themes and scope, it was possible to customise the content according to the users. The scope feature incorporated into the knowledge map enables dynamic changes in the navigational structure for various types of users thus providing different occurrences or resources for different users. This feature can be said to offer 'personalised' version of the knowledge repository. In this experiment, we have used this feature for two types of users – the agricultural extension worker and the subject-matter expert.

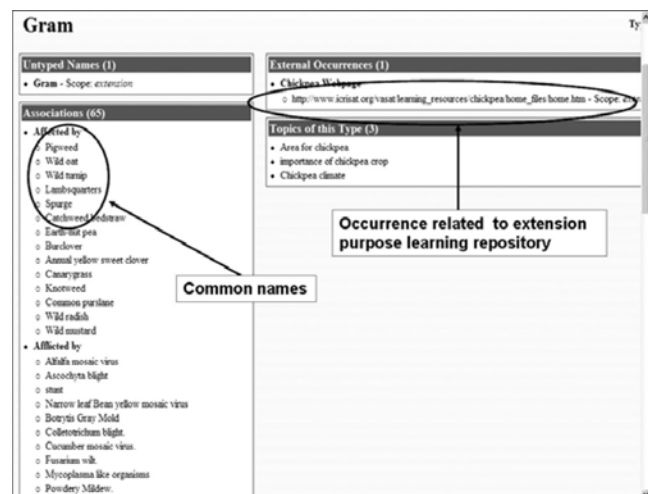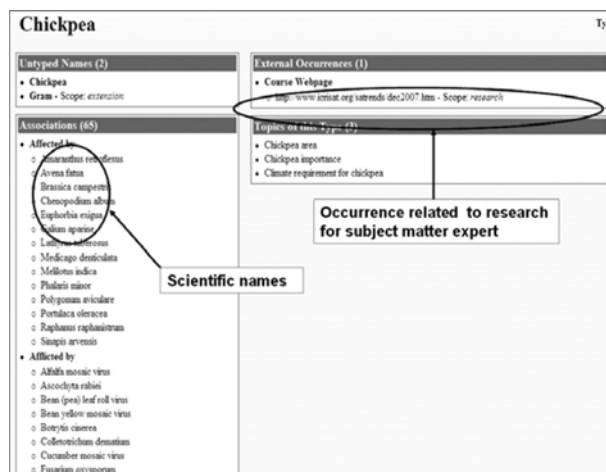**Figure 5** Chickpea topic map interface for an extension worker

Figure 6 Chickpea topic map interface for a subject matter expert

4 Discussion

The TM approach to information management is fast finding a range of practical applications. Complete portals are being run on topic map technology where the website's content, resources and site structure are controlled by TM and the layout and style are handled by the use of Extensible Stylesheet Language Transformation (XSLT).

An advantage in our approach is in the ability of a TM-based crop-knowledge model to integrate resources from various information repositories under one portal.

By separating the content repository from the site structure, it becomes easier to build sophisticated applications on top of a topic map driven portal. A previous effort at TM-driven portal, based on less specific, non-AGROVOC ontology can be accessed on the web (http://opaals.iitk. ac.in:9000/tmportal/index.html).

The meta-structure in this study was developed using the AGROVOC thesaurus. It is thus possible to translate it in to other languages supported by the AGROVOC, through mapping it with equivalent terms from the thesaurus (Figure 7). Experiments are in progress to achieve certain level of instance translation/mapping, and these will be communicated separately. With ontology mapping techniques and the use of shared ontologies, instance transformation is achievable to some extent.

Figure 7   Crop knowledge model terms automatically translated into French using AGROVOC

5   Conclusion

Use of controlled vocabulary is an important feature in creating knowledge models for information repositories. The knowledge models created serve as templates that could be reused over other information repositories with modifications on the URI to map to the underlying information objects of the intended repository. Knowledge models thus generated could be combined to generate newer knowledge models which overlay other information repositories. The models so created are not specific to a particular application. A collection of such reusable knowledge models, although heterogeneous, would provide the base for many complex information systems, in effect contributing to a new way to organise information in the domain of agriculture.

A similar experiment using a Wiki-based content management system and semantic extensions has been reported earlier (Kaur et al., 2007). There is an ongoing effort to construct an Agropedia indica (http://www. agropedia.net/), a novel content organisation with specific reference to agriculture in India using a similar technology namely concept maps (Novak and Canas, 2006). Ours is an effort to support this and similar online content organisations.

References

Ahmed, K. (2000) *Topic Maps for Repositories*, Available at: http://www.gca.org/papers/xmleurope2000/pdf/s29-04.pdf Last visited: September 2008.

Berners-Lee, T. (1999) *Weaving the Web: The Original Design and the Ultimate Destiny of the World Wide Web*, Harper San Francisco, Available at: http://www.w3.org/People/ Berners-Lee/Weaving/ Last visited: August 2008.

Biezunski, M. and Newcomb, S.R. (2001) 'XML Topic Maps: Finding Aids for the Web', *IEEE MultiMedia*, Vol. 08, No. 2, pp.104–108, Available at: http://www.coolheads.com/ SRNPUBS/ieee-mm-topicmaps-article.pdf Last visited: March 2008.

Dicheva, D. and Dichev C. (2006) 'TM4L: creating and browsing educational topic maps', *British Journal of Educational Technology (BJET)*, pp.391–404, Available at: http://compsci. wssu.edu/iis/nsdl/Publications/BJET06-Dicheva-Dichev-Submitted.pdf Last visited: September 2008.

Kaur, P., Patwar, S., Sylvester, A.G. and Balaji, V. (2007) 'Use of semantic Wiki tool to build a repository of reusable information objects in agricultural education and extension: results from a preliminary study', Paper presented at the *Web2fordev Conference*, Available at: http://www.web2fordev.net/fileadmin/user_upload/web2forde v/web24dev-ICRISAT-2.ppt, Last visited: September 2008.

Novak, J.D. and Canas, A.J. (2006) *The Theory Underlying Concept Maps and How to Construct and Use Them*, Available at: http://cmap.ihmc.us/Publications/Research Papers/TheoryUnderlyingConceptMaps.pdf Last visited: September 2008.

Pepper, S. (2000) *The TAO of Topic Maps – Finding the Way in the Age of Info Glut,* Available at: http://www.ontopia.net/ topicmaps/materials/tao.html Last visited: August 2008.

Rao, J., Prabhakar, T.V., Chattarjee J. and Singh, M.D. (2006) 'A topic map driven portal for agriculture information', *17th International Conference on IRMA (IRMA 2006)*, Washington D.C, USA. Available at: http://opaals.iitk.ac.in/ deal/other/topic_map.pdf, Last visited: September 2008.

Rath, H.H. (2000) *Making Topic Maps More Colourful*, STEP Electronic Publishing Solution GmbH, XML Europe, Available at:http://www.gca.org/papers/xmleurope2000/ author/../papers/../pdf/s29-01.pdf Last visited on September 2008.

Websites

Agropedia indica portal: http://www.agropedia.net/

AGROVOC's AGRIS/CARIS classification scheme: http://www.fao.org/aims/ag_classifschemes.jsp

Food and Agricultural Organization's Multilingual structured thesaurus, Agricultural Vocabulary (AGROVOC) http://www.fao.org/aims/ag_intro.htm

International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), http://www.icrisat.org/

TM-driven portal developed by Indian Institute of Technology – Kanpur. http://opaals.iitk.ac.in:9000/tmportal/index.html

Virtual Academy for the Semi-Arid Tropics: http://www.icrisat.org/vasat

XML Topic Maps: http://xml.coverpages.org/XTM-SC34-N0328.html

Queries

AUTHOR PLEASE SUPPLY CITATIONS FOR FIGURES 1–6 IN TEXT.

AUTHOR PLEASE CHECK IF THE CHANGE MADE IS OK IN PAGE 4.