Use of Semantic Wiki Tool to Build a Repository of Reusable Information Objects in Agricultural Education and Extension: results from a preliminary study

Pritpal Kaur[1], Shelly Patwar[1], Asil Gerard Sylvester[1], Venkataraman Balaji*[1]

* Corresponding author: v.balaji@cgiar.org

[1] Knowledge Management and Sharing Group
International Crops Research Institute for the Semi-Arid Tropics
Patancheru – 502324, AndhraPradesh. INDIA.
Ph : 91-40-30713205
Fax : 91-40-30713074

Abstract

There has been significant interest in applying the practices of semantic web to build an online repository of agricultural information (example: FAO-India efforts in AGROPEDIA). We will describe the results from a preliminary investigation into the use of a semantic tool for Mediawiki with content from the agricultural domain as the substrate.

The English AGROVOC have been used to generate 17 broad categories of terms to harvest entries from the Wikipedia and to form the relationships. These can be browsed at http://vasatwiki.icrisat.org/index.php/. A total of 1000 articles were downloaded and posted onto a local server running MediaWiki software. A semantic tool (http://ontoworld.org/wiki/Semantic_MediaWiki) was used to create additional tags in the hosted entries and links were formed.

The searches inside the semantically re-constructed collection are more specific compared to searches on the downloaded collection. We are conscious that this may not be a rigorous comparison. An even more interesting possibility is the ease with which the entries can be constituted into information objects that can be reused because their relationships with other objects in this re-constituted domain are rich and complex. With suitable filters the information objects can be exported to various interfaces (re-purposing). An area of interest is the use of this approach to aggregate content for use in education and in extension.

Keywords: collaborative authoring, semantic web, reusability of learning resource

Introduction

The emerging Web 2.0 paradigm has much promise for the agricultural knowledge domain. In the earlier phase, institutionalized agricultural research, education and extension practices could not take full advantage of the power of the web as a medium of communication and instruction. However, Wiki-like content management systems, supplemented by blogs and online social networks, are likely to make significant contribution to the way agricultural research, education and extension are practised in the developing countries. The perceived challenges to making contributions to online content or to discussions are much less now because of the participatory nature of Web 2.0.

The emerging paradigm of Semantic Web (Berners-Lee, 2001) offers unprecedented new opportunities in agricultural information organization. Tools and approaches of the semantic web, combined with Web 2.0 technologies, can offer novel possibilities of highly targeted, power searches and more effective access to information using map-like interfaces. In this paper, we shall report the results from a study in the agriculture domain that combines the use of Mediawiki software (that runs the well-known Wikipedia) with a semantic tool developed in the Open Source mode.

Why semantic web tools?

A key challenge in information management in agriculture, especially in extension, is to facilitate de-linking of content from the medium of presentation. This requires a wide array of tools and technologies when attempted in the non-semantic web paradigm. The new paradigm offers a number of tools and techniques that allow for more integrated view of content, specialized navigation and improved representation. According to Paul Warren," In the current WWW the search process lacks precision because it's based on a search for matching text strings. In the Semantic Web, most of the information will be semantically marked up" (Warren, 2006).

A significant new advantage is the opportunity for information and content aggregation from information objects that may be in one or more inter-linked online repositories. Such objects have the advantage of supporting rapid re-purposing, which is a key requirement in extension communication and in flexible learning in agriculture. In the following, we present results from a set of experiments carried out with a semantic wiki tool with agricultural content derived from the Wikipedia in English, and with content from the learning resources collection maintained by the Virtual Academy for the Semi-Arid Tropics (VASAT www.vasat.org). We are aware of the DBpedia.org, which is an effort to apply semantic web tools for carrying out sophisticated searches on the Wikipedia in multiple languages.

In our experiment, we have attempted to use the FAO's AGROVOC in English as the equivalent of ontology, in the absence of a widely accepted model for knowledge representation in the agricultural domain.

Our experiment

The platform

VASATWiki (http://vasatwiki.icrisat.org/) is a content management system that runs on MediaWiki software and has an added semantic extension (http://wiki.ontoworld.org/wiki/Semantic_MediaWiki) to it. The FAO's AGROVOC (http://www.fao.org/aims/ag_intro.htm), the multi-lingual thesaurus, was used as a guiding ontology for domain-specific information categorization in this experiment.

Method

Articles in the agricultural domain were harvested from the English version of the Wikipedia, which were uploaded onto the VASATWiki. These articles included various sub-domains of agriculture like crops, insect pests, agricultural machinery, aquatic sciences and fisheries, horticulture, agronomy and many others. VASAT's learning modules (http://www.vasat.org/learning_resources) were also made available on the VASATWiki.

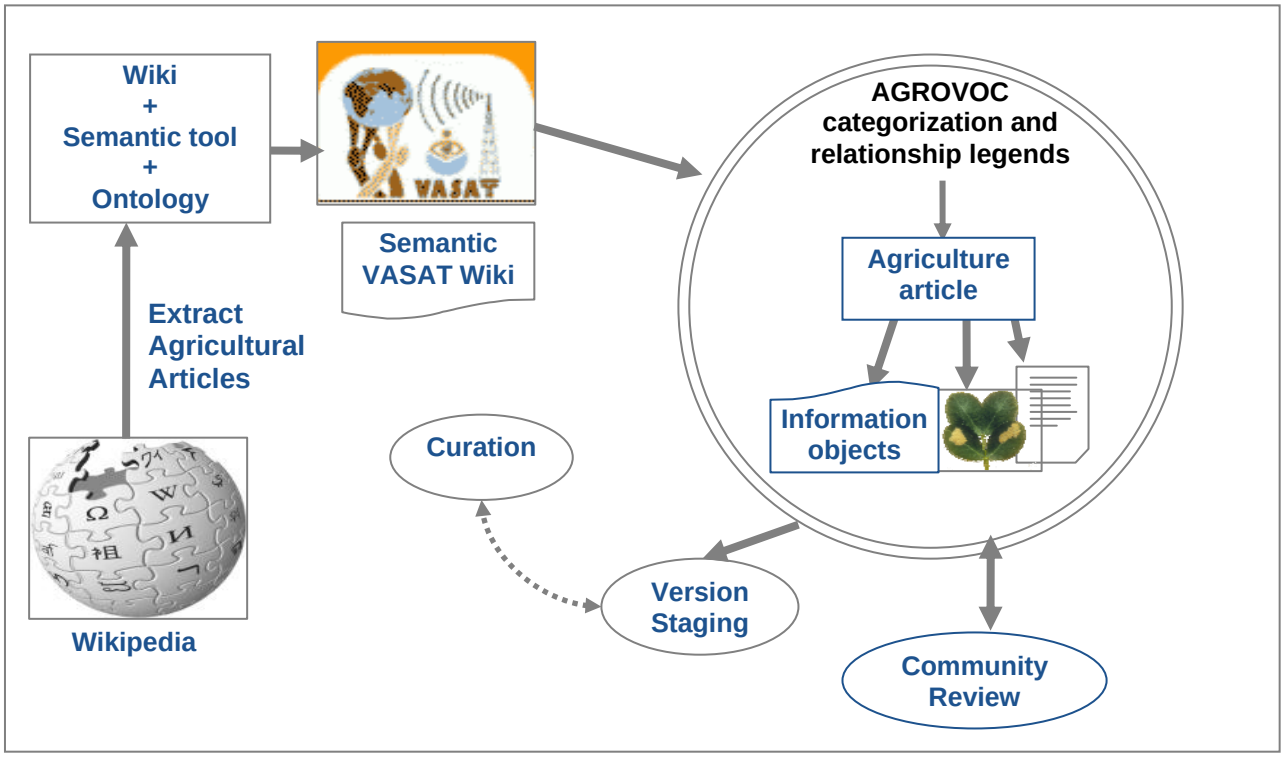

Figure 1: Workflow for creating 'semantically' linked content on VASATWiki

These articles and learning material were categorized according to the FAO AGRIS/CARIS classification scheme (http://www.fao.org/aims/ag_classifschemes.jsp) into 17 broad categories and further sub-categories *(Figure 1)*. The article title is taken as the basis for categorization. Some articles that do not fall under any of the classification schemes of AGRIS/CARIS were added to the closest relevant category in the VASATWiki.

The articles were then divided into information objects that could be reused based on the sections in the original article in Wikipedia *(Figure 2a)*. The information objects were

then semantically linked among themselves using the relationship legends from AGROVOC (*Figures 2b)*, which is the equivalent of ontology for this experiment. A comprehensive and structured metadata set about every article on VASATWiki has been created using the AGROVOC's classification and relationship terms. This is available as Resource Description Framework (RDF) feed *(Figure 2c)*.

- For example, the pigeonpea article was extracted from English version of the Wikipedia (http://en.wikipedia.org/wiki/Pigeonpea), (*Figure 2a)* and this was uploaded onto the VASATWiki. The pigeonpea article was then divided into information objects based on the sections in the original Wikipedia article like uses, cultivation etc.,

Figure 2a: The pigeonpea article was extracted from Wikipedia

- Tags to enable semantics, as provided by the Semantic tool extension, were coded into the granularized pigeonpea article in VASATWiki *(Figure 2b)*.

  For example, [[pest::Maconellicoccus hirsutus]] would establish a semantic link between the information object pigeonpea and *Maconellicoccus hirsutus* with a relationship.

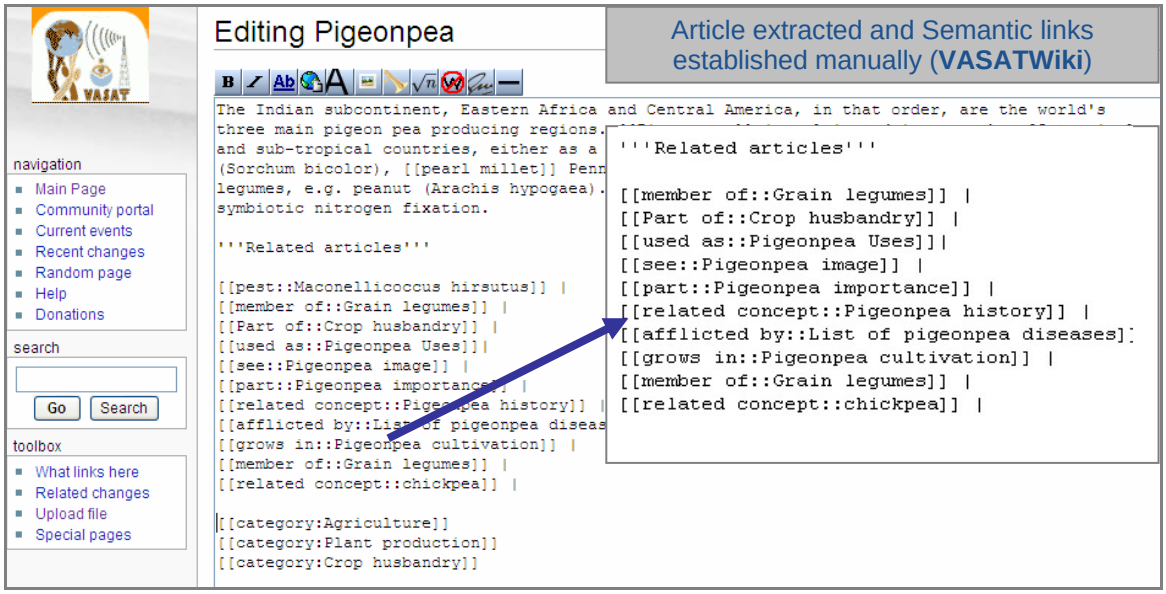

Figure 2b: Semantic annotations in pigeonpea article on VASATWiki

- The resultant pigeonpea information object found in VASATWiki (http://vasatwiki.icrisat.org/Pigeonpea) now contains semantic links to other related articles (*Figure 2c*). This could be exported to an RDF feed for machine-reading.

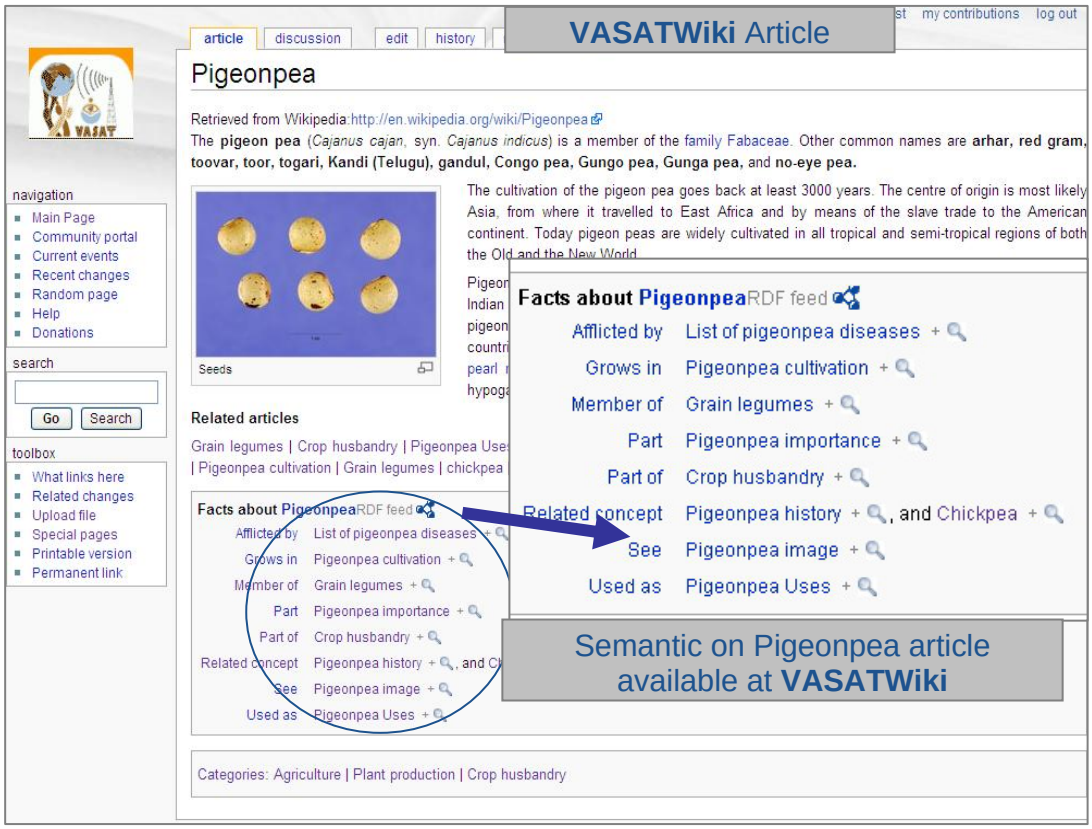

Figure 2c: Semantic on pigeonpea article available on VASATWiki

As illustrated in the above screenshots, the pigeonpea article from the Wikipedia is harvested, split into smaller information objects and is semantically linked together with

other objects in the repository using tags provided by the semantic extension to MediaWiki.

Results

Improved content navigation

The ability to conceptually navigate through an information maze facilitated by a semantically linked article provided a greater level of flexibility in navigation and knowledge acquisition.

- As seen from the figure 3a, the groundnut article found on VASATWiki (http://vasatwiki.icrisat.org/index.php/Groundnut) contains semantic links to its related concepts in the agriculture domain.

Figure 3a: Groundnut article with semantic links on VASATWiki

- *Figure 3b* shows an extract of the *"Facts about Groundnut"* that is generated by the tags encoded in the article.

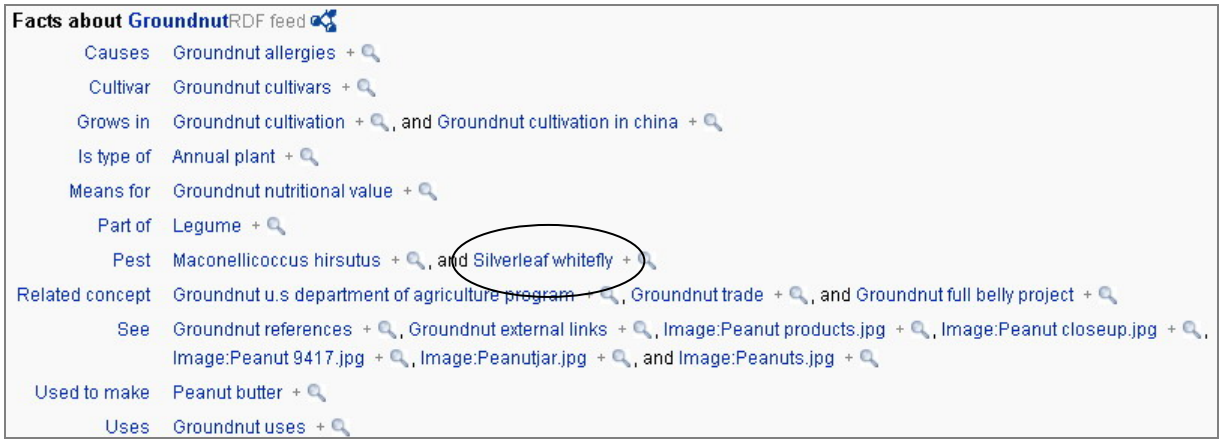

Figure 3b: Facts about groundnut on VASATWiki (generated by semantics in the article)

- These tags make for easy navigation based on relationships among related articles and also set the context for navigation as is evident from *Figure 3c* which is reachable from the relationships of groundnut article shown in figure 3b.

Figure 3c: Silverleaf whitefly (a pest of Groundnut) reachable through the semantics on the groundnut article on VASATWiki

VASATWiki also facilitates the ability to extract RDF feed of articles which could then be consumed by other applications. The feeds can be converted to an OWL-DL (Web Ontology Language) format, which could then be run through an OWL reasoner, such as Pellet (http://pellet.owldl.com/), for further knowledge discovery.

Improved content presentation

The semantic annotations in the Wiki articles enhance navigation as the relationships used gives meaning to the linkages build between the pages. Context based and context-sensitive searches together with searches on relationships, attributes or a free form semantic search could be executed on VASATWiki (*Figures 4a)*.

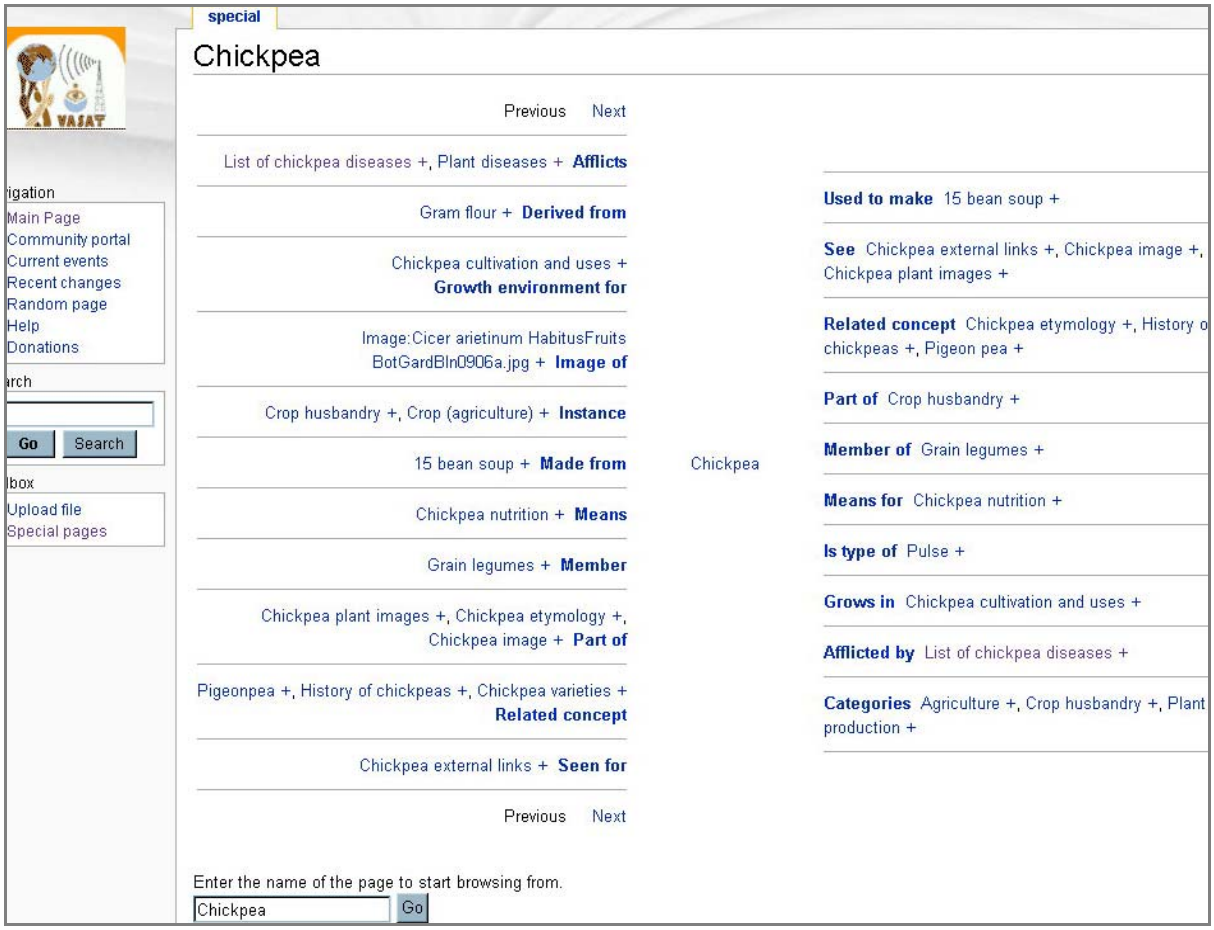

Figure 4a: A semantic search on relations on VASATWiki

Rapid content aggregation and repackaging using the eXe application

As a part of the experiment, we attempted to rapidly combine pieces of information and export the new information objects in various formats. The eXe (http://exelearning.org/) is a tool that provides professional web-publishing capabilities that can be easily referenced or imported by a learning management system. The eLearning XHTML editor (eXe) is an authoring environment to assist teachers and academics in the design, development and publishing of web-based learning and teaching materials without the need to become proficient in HTML or complicated web-publishing.

We have been experimenting with eXe editor for quick reuse of the content in different formats for potential delivery in different media.

In this experiment, we have tested out a wide range of information object combinations to generate new content in a relatively short time as shown in the workflow described in the figures below. The eXe provides for repackaging content from any Wiki (*Figure 5a)* and from websites *(Figure 5b)*.

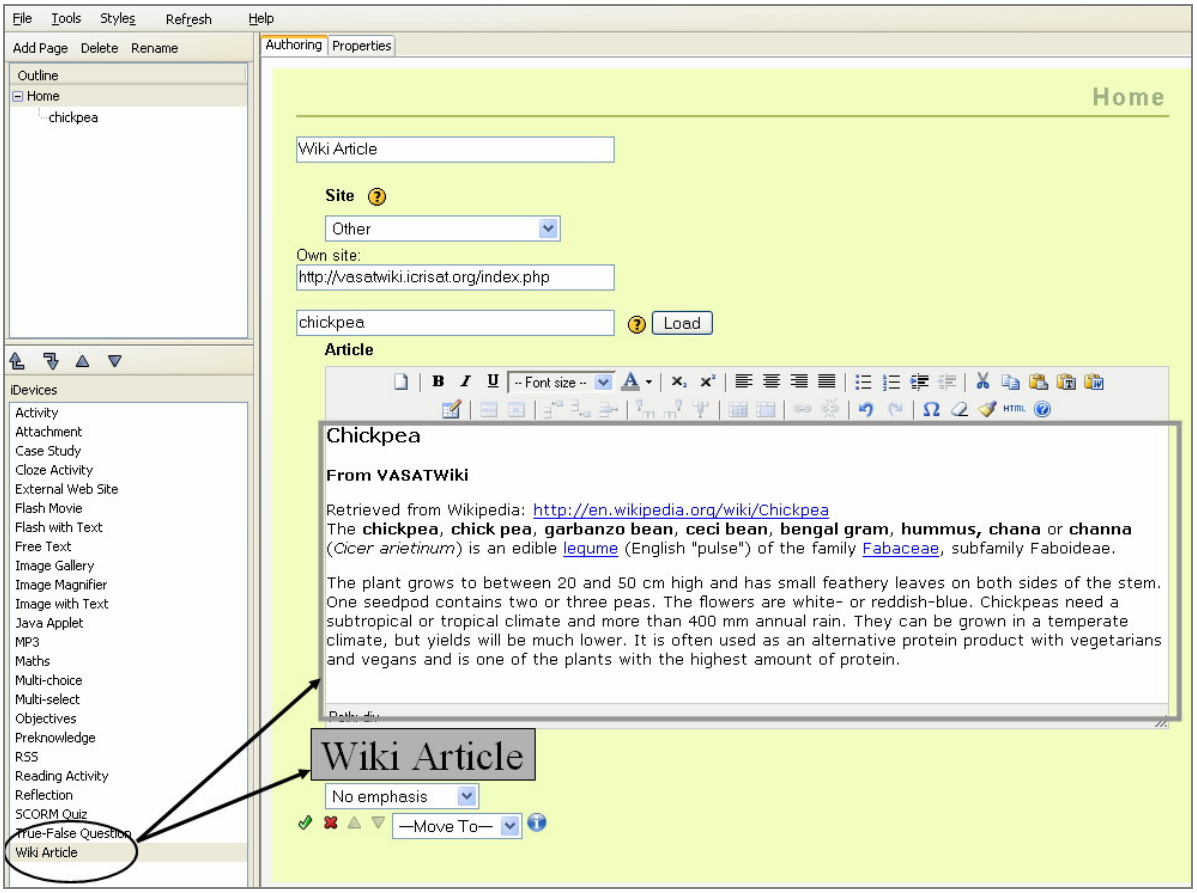

Figure 5a: Content being aggregated from VASATWiki

Other type of resources like Flash movies, multimedia content, Java applets as well as activity-based interactive content could be added. This repurposed content could then be packaged and exported to various formats such as plain text files, HTML pages, SCORM or IMS-compatible packages for use in an LMS.

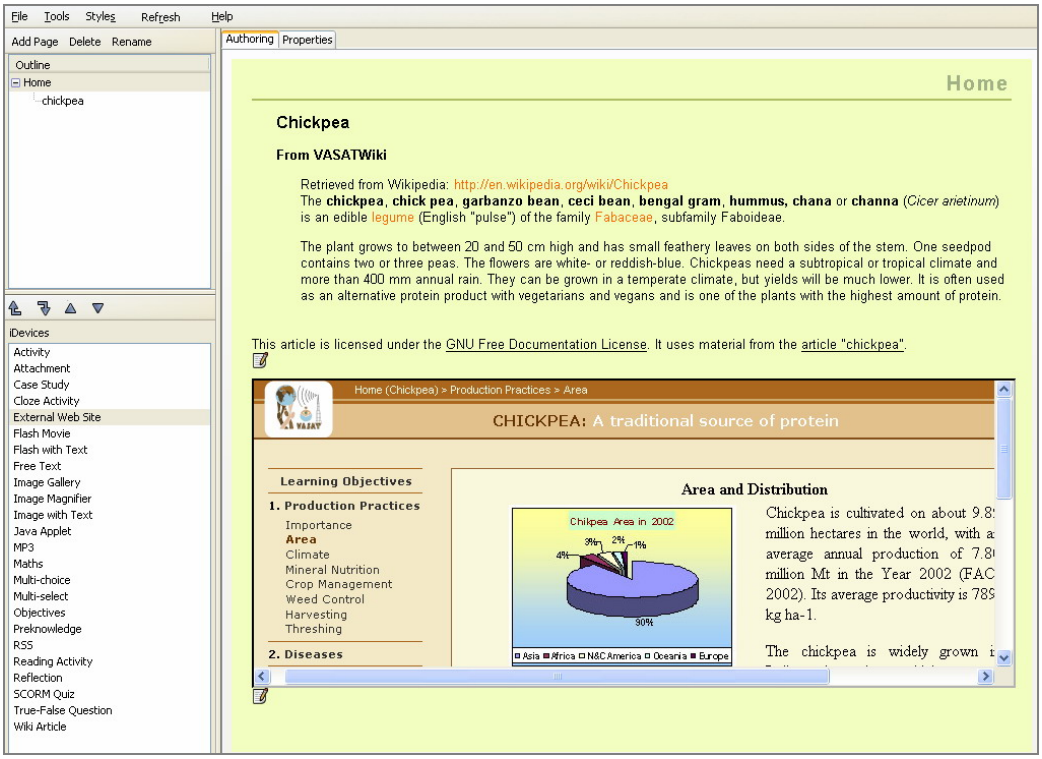

Figure 5b: Content being repurposed on eXe XHTML editor

The content so generated from the new eXe based workflow *(Figure 5c)* has also been tested for delivery in different digital media like iPod and MP4 players. The easier navigation on the VASATWiki and more pointed searches made it easy to locate the most relevant objects for re-combination.

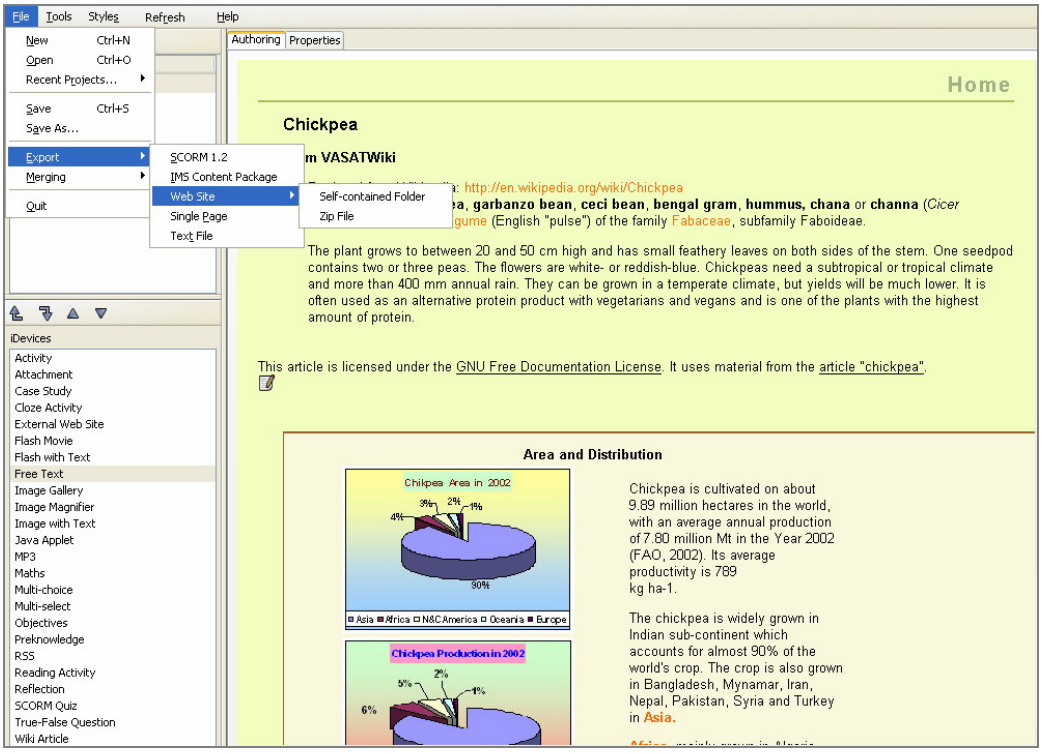

Figure 5c: Content packaging from different sources using eXe editor

Use of Topic Maps

Topic Maps enhance site cohesion, context-based navigation and easier searches for agriculture information. Topic Map is a new semantic web technology, which separates semantic connectivity from content and can be applied in conceptual navigation systems (Rao et al, 2005).

For experimental purposes, all the terms related to ICRISAT's five mandate crops were collected, classified and incorporated into a Topic Map using an open source software, TM4L (http://compsci.wssu.edu/iis/nsdl/download.html). The ontological relationships between different topics are defined using the AGROVOC's relationship legends. The topics are linked on resources on VASAT's learning resources repository. The Topic Map-driven website is currently under test at http://test2.icrisat.org. Each map is integrated with the top layer of the VASAT web site *(Figure 6)*.

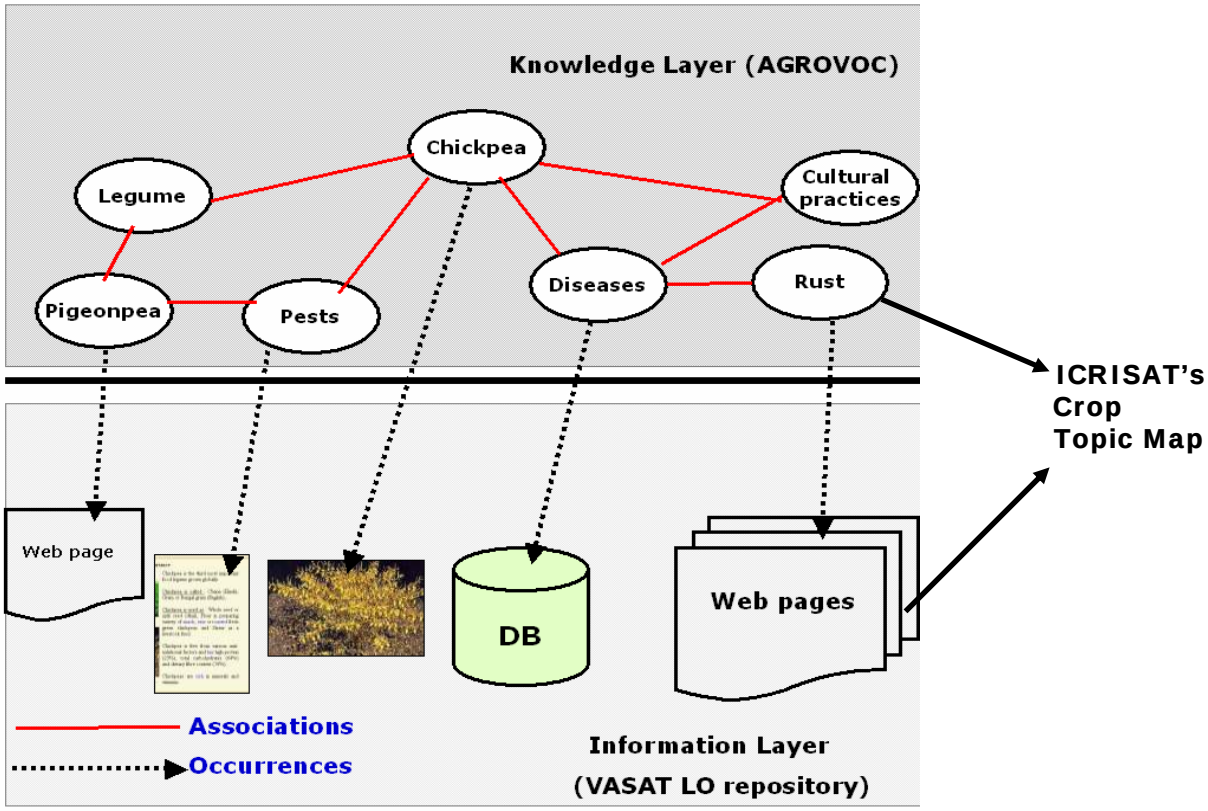

Figure 6: Visualization of the chickpea topic map created over VASAT's learning modules (http://test2.icrisat.org/)

Looking forward

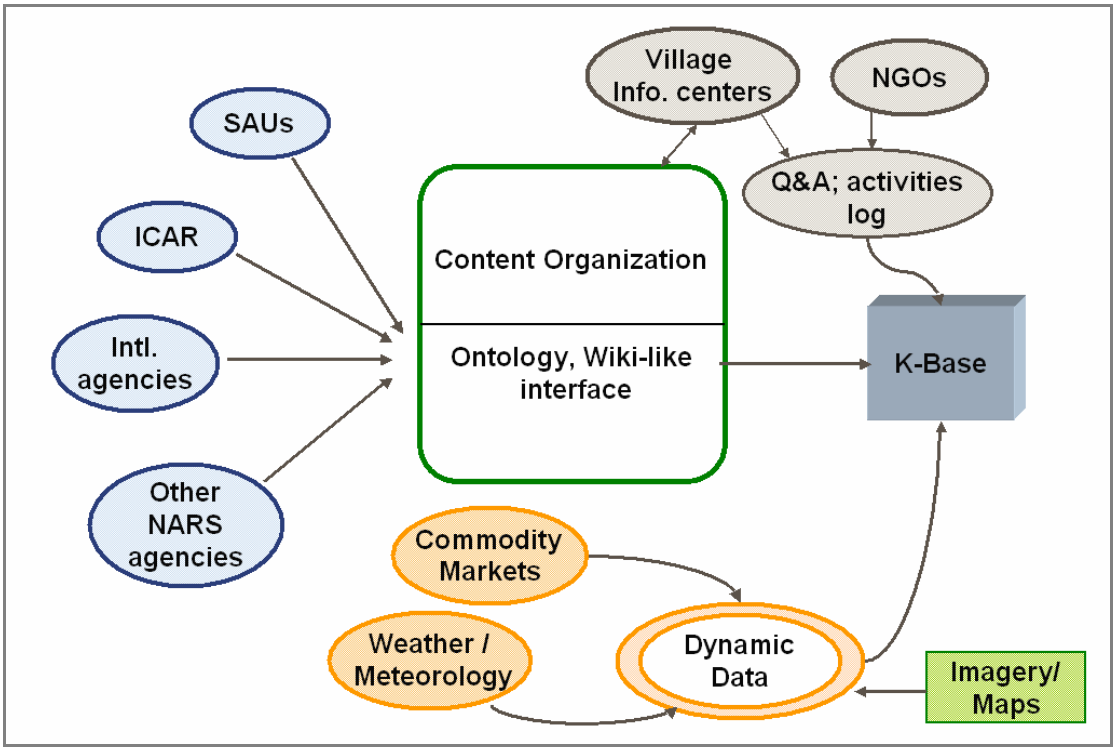

Figure 7: A novel content organization for Indian Agriculture

There is an ongoing effort to construct an AGROPEDIA (http://emandi.mla.iitk.ac.in/deal/). Our experiment can be viewed as an effort to make a contribution to this effort. We have further proposed that this be the basis of a new content organization for agriculture in India and have launched a project for design and implementation involving six national partners from India (http://www.vasat.org/research/agrid.htm). *(Figure 7)*

Acknowledgement

Work on this was supported by ICRISAT through its core funds.

References

[1] Berners-Lee, T., J. Hendler, and O. Lassila, "The Semantic Web," Scientific American, May 2001, pp. 34–43.

[2] Warren, Paul, “Knowledge Management and the Semantic Web: From Scenario to Technology”, IEEE Intelligent Systems, Jan/Feb 2006.

[3] Shneiderman, Ben, “Web science: a provocative invitation to computer science”, Communications of the ACM Volume 50, Number 6 (2007). June 2007, Page 25-27.

[4] Rao, D.V.Janardhan, T.V.Prabhakar, Jayanta Chatterjee and M.D.Singh, “A Topic Map driven Portal for Agriculture Information”, Available at: http://emandi.mla.iitk.ac.in:9000/tmportal/index.html