Collaboration

Architecture of the Agropedia Platform: Creating Content and Community through Collaboration

Nagaraju Pappu, Canopus Consulting
Runa Sarkar, Indian Institute of Management, Kolkata
T.V. Prabhakar, Indian Institute of Technology, Kanpur

Due to a lack of precise and exact terms of reference, there is very little content related to agriculture on the Internet. Agropedia, one of the world’s first agricultural knowledge repositories built from semantic, collaborative, and social networking metaphors, bridges this gap via agricultural knowledge models. Creating vibrant and diverse communities requires careful planning as well as an open, yet flexible, protocol. Agropedia, through its reviewed, scientific content contributed by agricultural research institutions and its community-generated interactive folk knowledge, brings together the expert community and the user community. An excerpt of this work appears in IEEE Internet Computing’s September/October 2010 issue.

Most of India’s workforce is tied to the agricultural sector, which faces unprecedented challenges. On one hand, economic growth, the rise of the middle class, and population increases have raised demand for food production by several orders of magnitude, and on the other, changing climatic patterns, depletion of natural resources, overexploitation of farmland, and general deforestation have caused an unprecedented shortfall in agricultural production. Unfortunately, there’s very little public awareness about these problems.

Recognizing the need to create a comprehensive knowledge dissemination platform, a consortium of ICT institutions, state universities, and research organizations have come together to support and promote knowledge exchanges between different stakeholders in the agriculture domain. Agropedia is the result of this ongoing effort. While building Agropedia, we recognized that there isn’t much agricultural content on the Web—even on Wikipedia, of the more than 2 million articles available, only about 3,000 relate to agriculture. Agropedia is an attempt to fill this gap. It’s unique in many respects: first, it’s the result of a consortium of institutions and universities that provide edited and reviewed expert content and articles; second, it’s a way for expert content creators and ultimate target users (farmers) by allowing many different organizations in between to collaborate, participate, and inter-
To accomplish such collaboration at a large scale, Agropedia uses knowledge models (a simplified ontology) to serve as a precise vocabulary of communication, which is very similar to the efforts in fields such as medicine but a first of its kind in the agriculture domain.

Agropedia differs most from the Wikipedia model in its use of inherent semantic indexing capabilities and authorized content. Wikipedia also lacks the mechanisms for the end user community (people interested only in using and reading the content) to participate in the feedback and content enrichment process. Likewise, some of the more common ontology-driven semantic content systems don’t provide for collaboration or community participation.

This article describes the experience we gained while creating this platform. An excerpt appears in print, in the September/October 2010 issue of *IEEE Internet Computing*.

**Organization and Structure of agropedia**

Agropedia is “all things agriculture,” meant for anyone connected with or interested in the agricultural domain, as shown in Figure 1. It aspires to be a one-stop shop for any information related to Indian agriculture—an audio-visual encyclopedia designed to transform the process of digital content creation and organization by making it an enchanting educational experience.

The two most important dimensions of Agropedia are its content and its community, which are intertwined in many ways. Agropedia’s expert community consists of individuals or organizations engaged in agricultural research and knowledge dissemination, such as state and central agricultural universities, research institutions, extension workers, research stations, and so forth. Its user community primarily consists of people at nodal- or village-level government agricultural centers—call center representatives, students, farmers, traders, and so on.

Naturally, Agropedia’s content is also of two types: expert sourced content and community-contributed interactive content. The *Gyandhara*, or expert, content is edited, reviewed, created, and maintained by the participating institutions. Agropedia also supports the creation and distribution of inter-
active content called Janadhara (folk-stream) from anyone in the community through wikis, blogs, forums, questions and answers, reviews, comments on articles, and content tagging.

Agropedia allows content contribution by any member of the national agricultural research system institutions in India or anywhere in the world. It includes a wide range of content, including text, images, and multimedia elements such as audio, video, or animations.

Agropedia also provides a flexible authoring environment and tools for content acquisition, processing, and publishing. It consists of social and collaborative networking features such as blogs, forums, wikis, questions and answers, content tagging by users, commenting, and rating authors and content.

Content and Community Creation through Collaboration

Building a platform like Agropedia goes beyond technology and software: the primary challenge is to enable an environment that allows a community to grow, organize itself, create its own content, and interact and collaborate using the underlying content repository as the primary vehicle of collaboration.

As Clay Shirky noted, a community is very different from an audience (http://shirky.com/writings/broadcast_and_community.html). Audiences can be built, but communities create themselves and grow. However, to develop they need an underpinning of a constitution—a way to govern themselves, facilities to create their own languages of communication and interaction, and methods to recognize and reward contributions by members. At the same time, when the community becomes too large and too diversified, it loses its focus (http://shirky.com/writings/group_enemy.html). The best way to deal with this is to create a platform that serves not only as a community network but would also allow formation of networks of communities.¹

Another challenge to consider is the awareness and capability of the agricultural community in India to absorb, assimilate, and work with a technology-assisted medium. The technology can’t become a barrier to participation—instead, it should enable and encourage it, irrespective of the user’s technical proficiency.

Function and Role of Knowledge Models

The heart of Agropedia is its knowledge models (http://agropedia.iitk.ac.in/?q=content/knowledge-models), which agricultural scientists and experts from participating institutions created as a set of concept maps (http://cmap.ihmc.us/concept-map.html). These knowledge models are Agropedia’s lingua franca. In the early stages of development, we quickly realized that a large-scale content creation effort for use by a diverse community would require its own language of communication. For example, a paper or article written by a scientist wouldn’t be directly relevant to a farmer—however, it might be useful to a person working in the agricultural research station or an extension worker who needs to provide essential guidance or information to that farmer. Similarly, a user such as an extension worker might want to summarize many articles or papers written by experts, connect such content together, and synthesize it for common, general usage.

Effective collaboration assumes absence of hierarchical authority structures and centralized power centers. One reason why several attempts to build such socially significant systems have failed is because of the naïve assumption that anyone in the chain could produce something for the direct consumption of the final, target end user. The semantics of collaboration means that we must produce for our nearest neighbor. If everyone in the chain works for their nearest neighbor, we end up with a vibrant, participatory community that grows.

The lack of consistent terms of references for the agricultural domain is perhaps the reason for the Internet’s dearth of agricultural content. Precise and exact terms of reference are a fundamental requirement for goal-oriented communication and interaction. Without such terms, it’s extremely difficult for a member of the community to take something from an expert, enhance it, add value to it, and pass it to someone the next level down in the chain. In other words, humanization of knowledge isn’t possible without a shared set of terms of reference.

Large-scale, content-rich systems in specialized domains such as medicine have used semantic networks and specialized ontologies successfully (www.nlm.nih.gov/research/umls/documentation.html). For example, the UMLS’s role in the creation, use, and propagation of medical articles on the Internet makes them accessible not only by experts but also by an ordinary user. Agropedia’s knowledge models take a big step in this direction by contributing to a precise and exact agricultural vocabulary.

IITK and FAO first developed a generic knowledge model for crops, which was then used as a baseline model for crop knowledge models. Figure 2 shows a generic knowledge model and a specific crop ontology.

Knowledge models are used for a wide variety of applications, from content tagging to automatic cross-referencing of content to relationship discovery. V. Balaji and colleagues² described in detail the process of how the agricultural knowledge models were created, their importance, and the various applications built via those knowledge models.

The Function and Role of Folk Knowledge

We expect that the interaction space provided through the Janadhara stream will ultimately make it possible for the community to humanize the “expert” knowledge and make it more accessible for general, common use. In particular, it can help add practical, field-level examples, case studies, and enhancements to the main expert content. It’s also possible for users to link content via their own user-defined tag system, discover content in the repository, use the expert content as a primary reference material to illustrate specific issues for farmers, and keep the content alive and the community vibrant.

Lasting communities make up and transmit their knowl-
edge, culture, and values using folklore, which is the basic idea behind the Janadhara. As Ananda Coomaraswamy pointed out,3 “folk tradition is ‘folk’ only in respect to its transmission, not its origin. Folklore and Philosophia Perinnis spring from a common source.”

Inclusion and User-Centric Design

The Nobel Prize-winning economist Amartya Sen described the economic, social, and cultural value of Swlkriti4 as a culture of openness, tolerance, and inclusion. Swlkriti basically means that everyone, irrespective of his or her capacity, has a place, function, and role. This is one of the central community building principles behind Agropedia. Here, we describe how we used this simple principle to make crucial design choices.

The Knowledge Model Creation Process

The platform and toolset must be designed such that they don’t demand a sharp learning curve by the user community (be it composed of content creators or users). While designing and creating the knowledge models, we introduced simple graphical tools like concept maps (CMAPs), which are easy and fun to use. In fact, the agricultural scientists weren’t even aware that they were making ontologies. The underlying platform was designed in such a way that it converts the CMAPS to internal semantic indexing schemes. We conducted simple workshops and designed a few guidelines5 on what makes good concept maps, which is all that was required to generate the knowledge models.

The Content Acquisition Process

We designed Agropedia’s content acquisition process to be intuitive and nonintrusive. Many agricultural scientists, research workers, and government officers don’t understand the complexities of modern software. Moreover, many of them don’t have a high-speed, always-on connection to the Internet. Therefore, we designed content submission to be extremely easy and nonintrusive. Any authorized member can submit content via email or through the website in any format. A small back-office IT staff accepts this content and runs the content transformation tools. The transformed content is then presented to the original author, who can make any necessary modifications, change links, cross-references, and ontological entities, add or modify the tags, and so on. The original content and all subsequent transformations are preserved so that the entire process can be repeated to generate the final content view at any time.

The Content Transformation Process

The most successful software in the world hides its complexity—for example, just imagine the complexity that Google hides behind one simple text box. Successful software designers intuitively understand that the application and its “features” shouldn’t be conspicuous. The software should present what the user needs and wants—content, community, and collaborative environment—not menus, application buttons, help screens, complex search, and navigational layouts. Thus, the Agropedia interface presents only content to users; its functionality is presented in the form of embedded links and navigation in content elements, much like present-day Web 2.0 applications such as LinkedIn, Orkut, and Facebook.

Agropedia’s underlying architecture decouples content au-

---

Figure 2. Each crop has its own specific knowledge model. All the crop knowledge models use the generic crop model.
thoring from content publishing. As mentioned earlier, the content must be transformed using automated or semiautomated tools so that it becomes useful for the final end user. Figure 3 shows the content transformation model. Submitted content is checked into the main content repository in the user’s original format. Automated tools then convert it into Agropedia canonical format (an XML representation of the content). Then, the content is parsed and the relevant ontological entities and relationships are automatically inserted as navigational links. In the final step, the content is cross-linked in the repository.

Content Access, Community Building, and Web 2.0

The content’s state of readiness, its author’s information, whether it’s reviewed or under review, and its status are shown as part of the content (similar to Wikipedia). Along with this information, the user community can also rate the content—how many people have accessed the content and its author’s standing are all shown as part of the content’s attributes.

This model accomplishes many community building objectives: specifically, it lets the community be aware of the most useful content, and it encourages positive reinforcement and feedback to the authors.

Agropedia Technology

Agropedia is built on open source platforms and is completely open source itself. We realized that its main assets aren’t the technology and software but its content and community. Therefore, our primary architectural goal was to design Agropedia to scale to large amounts of content and a very large and diverse user base. Most importantly, it was crucial that we keep the content’s lifetime high. In general, content outlasts people and even the tools used to create it. This is why we chose an open source platform so that content isn’t locked up in “technology.”

Many applications that use vendor-specific storage formats end up paying recurring license costs just to access content. Locking content formats to any particular representation or choosing binary/proprietary formats aren’t useful in the long run. We adopted an open content representation approach that lets us use a metarepresentation scheme like Tex, SGML, and Docbook to preserve semantic structures and typesetting information. This is equivalent to the content being stored as a program plus data, which enables us to produce documents conforming to various formats and standards at runtime.

To make Agropedia scale to serve a very large end-user base, we decoupled the back end, content authoring, editing, and transformation environment from the content access applications. This allows deployment of the content databases on multiple machines and locations using content delivery network (CDN) strategies.

The entities and relationships in the knowledge models link up the content and index it under all applicable concepts, which serve as basic expert-generated navigational aids. Users can also tag the content and design their own tags to group content together, meaning Agropedia can support both taxonomy-based navigation as well as folksonomy-based navigation.

Figure 4 shows how the content transformation tools take the original content and insert ontological entities and relationships, cross-links to other content, and so on.

The knowledge models, ontological entities, and relationships can be queried by using a simple API. We developed presentation plugins and themes as presentation services on top of the Drupal Environment, which is an open source content management system.
Agropedia is a unique effort in the sense that it’s a collaborative effort between many different institutions and organizations. It’s also unique from a technological and community-enabling point of view. It took us more than two years to create the agricultural knowledge models and the fundamental vocabulary. Currently, we’re in the evangelizing stage. The content transformation and automatic cross-linking environment is still under beta. We plan to convert the knowledge models into a full-fledged ontology such that automatic discovery and inferencing applications can be developed. We plan to use Subversion as the back-end content repository, which allows multiple people to edit, modify, and check-in the content even in an offline mode.

In Agropedia, the intelligent fusion of cutting-edge computing metaphors and collaborative culture is creating a simple yet silent symphony. Its social and economic relevance to a country like India, which faces major food production and distribution challenges in the coming decades, simply can’t be overstated.

Acknowledgments

The Agropedia is the result of an ongoing effort under the umbrella of the National Agriculture Innovation Project (NAIP) of the Indian Council of Agricultural Research (ICAR). IIT-Kanpur is building the overall technology, with ICRISAT being the primary coordinator and GB Pant Agricultural University and University of Agricultural Sciences, Dharwad acting as domain partners. Food and Agriculture Organization (FAO), Rome, participated in building the knowledge models. We’re particularly grateful to V. Balaji, Johannes Keizer, Margherita Sini, Antonella Picarella, and N.T. Yaduraju of these organizations for the help and encouragement they rendered in building Agropedia. OPAALS, a network of excellence project under FP6 of the European Union, also influenced the roadmap for Agropedia through its focus on open knowledge systems.

References


Nagaraju Pappu works as chief technologist at Canopus Consulting, an enterprise architecture consulting services organization. He’s also a visiting faculty member at IIT-Kanpur and IIIT-Hyderabad. Pappu specializes in large-scale enterprise systems architecture and semantic and collaborative environments. Contact him at pnr@canopusconsulting.com, or via www.canopusconsulting.com.

Runa Sarkar is currently a faculty member in the economics group at the Indian Institute of Management, Calcutta. Her research interests are digital ecosystems and the environmental sustainability and social economic impact of ICT on communities. Contact her at runa@iimcal.ac.in or via www.iitk.ac.in/ime/runa/.

T.V. Prabhakar is a professor of computer science at the Indian Institute of Technology, Kanpur. His research interests are databases, software architecture, knowledge modeling, Semantic Web, and collaborative computing. Contact him at tvp@iitk.ac.in or via www.cse.iitk.ac.in/users/tvp/.


Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.