

Innovative ICT Tools for Information Provision in Agricultural Extension (December 2005)

Krithi Ramamritham, Anil Bahuman, Subhasri Duttagupta, Chaitra Bahuman, Srividya Balasundaram
Developmental Informatics Lab, Indian Institute of Technology -Bombay

Abstract—aAQUA is an online multilingual, multimedia Agricultural portal for disseminating information from and to the grassroots of the Indian agricultural community. aAQUA simultaneously addresses two major challenges in farmer outreach programs - geographic reach and customized delivery. It answers farmers queries based on the location, season, crop and other information provided by farmers. aAQUA makes use of novel database systems and information retrieval techniques like intelligent caching, offline access with intermittent synchronization, semantic-based search, etc. Agricultural content repositories (Digital Library), Agri-price information (Bhav Puchiye), farmer schemes and various operations support databases (aAQUA-QoS) have also emerged from the experience of aAQUA deployments. aAQUA's large scale deployment provides avenues for researchers to contribute in the areas of knowledge management, cross-lingual information retrieval, and providing accessible content for rural populations. Apart from Agriculture, aAQUA can be configured and customized for Expert advice over mobile networks and the internet in Education, Healthcare and other domains of interest to a developing population. This paper will showcase the utility of various component databases built into aAQUA to enhance the QoS delivered to rural populations.

Index Terms— Agriculture, Caching, Databases, Database synchronization, Farmer Outreach, Human Computer Interaction, Internet and Communication Technologies for Development, Digital Library, Knowledge Management, Low-bandwidth Application, Multilingual Information Retrieval, Offline Access, Price Information.

I. INTRODUCTION

AGRICULTURE extension and farmer-outreach programs face three major challenges – cost-effective outreach, solutions tailored to needs of individual farmers and an image that is farmer-friendly. The internet and mobile networks have the potential to provide agro-information services that are (i) affordable, (ii) relevant (timely and customized), (iii) searchable and (iv) up to date. Large sections of the farming community, particularly the rural folk, do not have access to the huge knowledge base acquired by agricultural universities, extension-centers and businesses. While telecenters are beginning to dot the Indian rural landscape [1], one of the big barriers remains the lack of agro-content that (i) is in the language of the

farmers (ii) is relevant to their needs and (iii) is delivered in a form that is of immediate use to them.

Keeping these factors and the needs of Indian farmers in mind, the Developmental Informatics Lab, at IIT Bombay has developed and deployed innovative ICT tools for dissemination of agricultural information over the internet. This paper discusses three database-backed tools - aAQUA, Bhav Puchiye and a digital library of Agricultural Documents that are all accessible at the www.aaqua.org portal.

aAQUA - which stands for *almost All QUestions Answered*- is a farmer-expert Q&A database supporting Indian languages (Figure 1).

Bhav Puchiye (meaning, "ask for the price", in Hindi) is a web-based application for viewing the price and price-history information of agrarian products at the nearby wholesale markets (called *mandis*).

The Crops Library project was initiated to build a repository of agricultural documents. The content has been

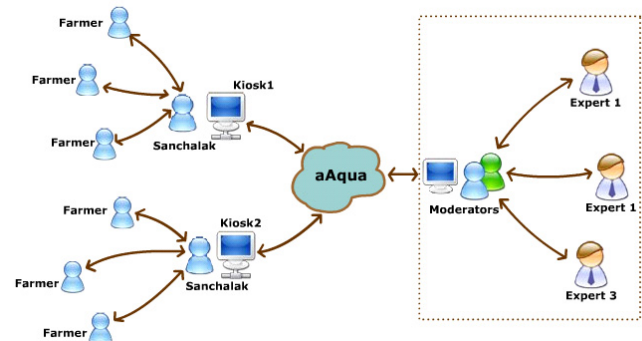


Fig.1. Flow of information in aAQUA system

provided by Krishi Vigyan Kendra, Baramati, who have collected recommendations from agricultural universities, residential experts, journal articles, disease forecast reports etc.

These research artifacts aim to incorporate innovations from the perspective of database query optimization and caching, cross-lingual multimedia information storage and retrieval, human-computer interaction and new ways of providing expert help to redress farmers' problems.

II. USAGE PERSPECTIVE

A. aAQUA:

aAqua [2] provides a communication framework for rural users to get their problems solved. Solutions are shared by a larger community and are provided by either person(s) having similar experiences or by agricultural expert(s). aAQUA is available currently on the internet and also through a number of computer operator mediated internet kiosks that are being deployed in India by both commercial and government funded ventures. The kiosk operators help in making the computer kiosks usable by users who are unfamiliar to computers and often not literate.

In this section, we present various stages a typical aAqua question passes through in its life cycle.

An aAqua question is posted either by a registered user directly or through a telecenter/kiosk operator who has an account in aAqua. Usually the question is from a farmer whose profile information provides details such as crop, farm size, pesticides and fertilizers used, dosage etc. This provides an appropriate context for the question. If the question is clear and complete, the Agri-experts provide an answer to the question. Otherwise they ask the original user for further details about the problem. aAQUA Experts response times averaged around 36 hours in the year (Jan – November 2005). A local help desk operator follows up with the farmer over a phone, a week after the answer has been given, to check if the answer solved the farmer's problem. In case the farmer found the solution useful he documents the impact. In case the solution was found to be ineffective the expert follows up on the phone and captures the limitations of the solution. In case of power, connectivity, or PC failure, farmers call the local center.

From a user perspective, the success of aAqua and the Crops Library lie in its incorporation of effective solutions that:

1. Provide an interface suitable for novice internet users.
2. Allow users to browse content by forum name as well as agricultural keywords and key phrases presented to the user.
3. Allow users access to content even over intermittent or low bandwidth connectivity.
4. Provide solution to a user in his/her own language.
5. Allow experts to recycle answers from the continuously growing knowledge base of aAqua for replying to recurring questions.

B. Bhav Puchiye:

Bhav Puchiye incorporates innovations from the perspective of interface design and data provisioning. An iconic interface is presented to the user (Figure 2) with a choice of commodities, the available varieties (from which the user chooses the commodity and variety of interest) and a

calendar (from which the user selects the date of interest). The prices are displayed spatially over a map. The user can decide where to sell his produce to get the maximum profit, depending on the prices and the distance of the markets. The user can create a login, store profiles of commodities and locations and create and receive e-mail alerts when prices of certain commodities of interest change in the markets of interest.

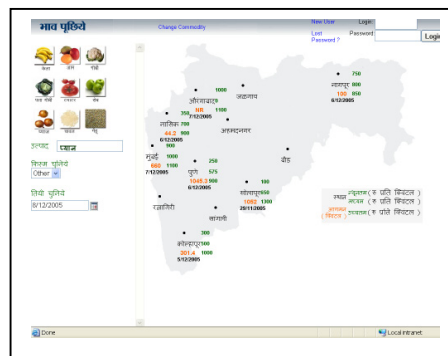


Fig.2. Bhav Puchiye: Market prices are shown on a map of Maharashtra state, commodities, varieties and dates can be chosen on the panel on the left.

C. Crops Library:

Crops Library consists of collections of crop diseases (Crip Doctor), crop recommendations and translated aAQUA threads. The collections are built centrally using open source software called Greenstone. A comparison of 3 open source digital library softwares [Bahuman et al.] – DSpace, EPrints, and Greenstone is shown in the table below.

Features	DSpace	GreenStone	EPrints
Open Source	Yes	Yes	Yes
Unicode Compliant	Yes	Yes	Yes
Hindi Interface (Existing)	No	Yes	No
Built in HTML viewers	No	Yes	No
Size	6 MB	31 MB	2 MB
Server	Unix Only	Unix/Windows	Unix Only
Offline viewing/Browsing	No	Yes	No

A user of Crop Doctor browses through a gallery of images classified by Crop and Disease names. He/she can choose a matching image and read the symptoms, causes, prevention and control of the disease.

A user of Crop recommendations chooses a crop of interest classified as vegetables, fruits, cereals and pulses, flowers and

“Others” and reads through tested recommendations collected from agricultural universities.

A user of aAQUA Translations collection can choose to view the questions and answers in English, Hindi or Marathi. The question-answers are also classified by crop names.

III. TECHNICAL PERSPECTIVE

Figure 3 explains pictorially how aAQUA and Bhav Puchiye [4] employ the three tier web architecture using Java technology (Java Server Pages/Servlets) and Oracle (aAQUA) or MySQL (Bhav Puchiye) databases.

Based on the standard MVC (Model View Controller) architecture, they are compatible with any Servlet container which supports JSP 1.2 and Servlet 2.3. They are currently being deployed within the Servlet container (Catalina) of the Tomcat Web Server. The systems use Unicode UTF-8 compliant databases. The aAQUA database comprises mainly of tables that include attributes for Member, Farmer, Expert, Moderator, Category, Forum, Posts, Thread, Permissions and Attachments. The Bhav Puchiye database comprises of tables that include attributes for Member, APMC user, Commodity names, Categories, Variety names, Market names, Arrivals and Prices.

The Crop library uses Dublin Core standard of metadata storage [5]. Documents are uploaded to the server and indexed by meta-data provided by a librarian. Alternatively, meta-data may also be extracted from the documents.

A. Enhancing Usability

The target users are predominantly non-English speakers, semi-literate and new to the internet. Our tools provide a simple, yet rich interface suitable for new internet users. A web-based keyboard is also available to assist users. The questions on aAQUA are answered by agricultural experts who provide answers in the local language (or a combination of languages), paying special attention to the language used. Use of technical jargon is generally avoided, for e.g., instead of prescribing quantities measured in parts per million (ppm) or grams, common measures such as the teaspoon are made.

Some agricultural and veterinary problems are better addressed by photographs or audio and video files which provide details to the expert. aAQUA allows attaching images taken by a digital camera or scanner and experts can zoom to specific portions of the images. As shown in Figure 4, users can add images and also attach audio and video files which are automatically played back on viewing.

There are many ways of organizing aAQUA content and it remains an open research issue especially when catering to different sets of users- farmers, experts, agri-businesses - both from literacy as well as expertise point of view. The two approaches in use so far are organizing (a) by question category – crops, animals, market price and others and (b) by agricultural keywords (in categories) chosen by experts.

The interface design of Bhav Puchiye employs the “Inverted Pyramid Approach”[4], which aims at maximizing the relevant results with minimal inputs from the users. The principle involves providing some results, by assuming default options, as soon as the web page loads. The user can change the options (usually provided on the same page) to view other results. Every mouse click in Bhav Puchiye refines the results.

The Crop Library interface is also designed using similar principles. In addition, the retrieval of documents is near instantaneous since the documents are converted into XML after uploading and displayed as HTML within the browser at runtime. This also eliminates the need for installing proprietary editor and viewer applications.

B. Offline Access, Database Synchronization and Caching

The underlying assumption in almost all web based applications is the availability of a continuous connection to the server over the Internet. However, aAQUA caters to users who connect to the internet on unreliable dial-up connections where the bandwidth is low or intermittent and the user is usually charged by the number of hours online. aAQUA pages are designed to have a lower payload and can also be installed as a standalone application which connects to the internet whenever available. The offline version of aAQUA is created for such users and can be personalized based on individual or group profiles and access patterns. It incorporates a store-and-forward mode, delaying authentication and allowing users to login and ask questions.

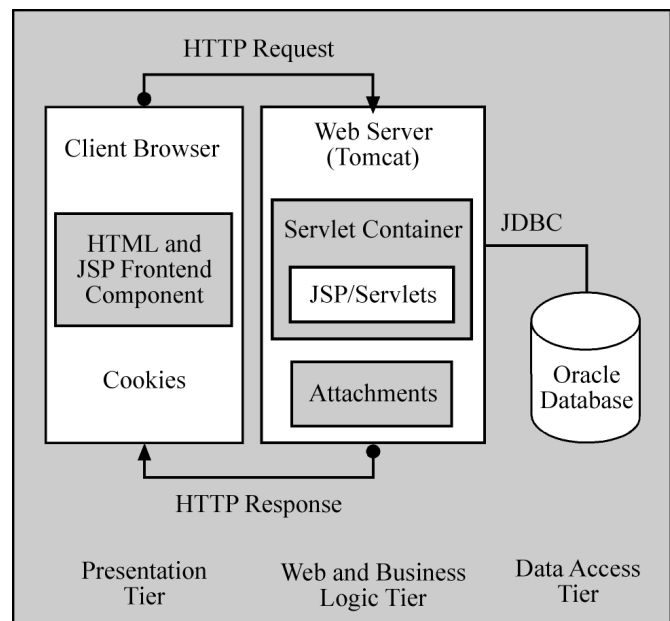


Fig.3. Three Tier Web Architecture of the aAQUA system

The application periodically updates the main server and also receives the latest updates. The offline version can also be searched using the keyword search interface described in Section D without connecting to the network.



Fig 4: Pictures are worth a thousand words for experts answering questions on aAQUA

In order to improve the response time and robustness in delivery of content, we also have deployed aAQUA mirrors closer to the users. These mirror sites periodically synchronize with the main aAQUA server with the help of our synchronization tools for databases of different vendors (Oracle and MySQL). Query logs are stored at either end and conflicts are resolved at the time of synchronization. The synchronization tool is designed to be tolerant to network and server failures during synchronization so that over time the synchronization process will complete.

The Crop Library can be configured for offline browsing and offline search and can be distributed in Compact Discs.

Web caching and database query-optimization strategies are also being exploited to help improve response. Frequently accessed fragments of pages as well as query responses are cached in the (i) browser (ii) one or more application server tiers.

C. Language Independent Semantics-driven Retrieval

We are converting the knowledge contained in the Questions and Answers in the forum to a language independent form – Universal Networking Language (UNL) – so that Semantic Based Search techniques can be used for multi-lingual retrieval. This performs better than keyword-based approaches [2]. When a user types a query, the system converts it to a UNL query graph and looks up the UNL document base to find matches. It responds with the answers (also in UNL) and translates them into the language chosen by the user in the AgroExplorer [3] tool.

In addition to enhance multi-lingual retrieval, we are currently integrating functionality to (a) automatically pick agricultural keywords and (b) perform a multilingual keyword based search on the database. This is especially useful for users fluent in two or more languages (which is quite common in India). The original search query is expanded with their counterparts in each language. It allows users to

search in their own language and retrieve content in other languages.

Since aAQUA serves many novice users, a popular search feature is a categorized list of hyperlinked keywords (also called Virtual fora) that appears on the home page. On clicking these hyperlinks, aAQUA performs a search on the corresponding keyword (e.g. Onion invokes all Onion related threads – independent of language).

Crop Library and Bhav Puchiye have also been localized in Hindi and Marathi. The success in organizing any digital library is in identifying good metadata. However, encouraging the farmer (or kiosk operator on his behalf) to provide relevant keywords and meta-text would not yield good results. The challenges of organizing and harvesting meta-data from aAQUA content are many. They arise due to the heterogeneous data in aAQUA (both textual and non-textual elements), improper grammar and spelling (requiring corrections) and at times manual choice of metadata. The effort would need separate resources like agri-ontologies and corpora for cleaning, organizing and identifying good metadata.

D. Reuse of knowledge in previous answers

Unlike most online discussion fora and web portals, a review of our search query logs has indicated that our users infrequently use the Search feature. The virtual fora feature was an attempt to encourage our users to perform search on the database. The “Browse by Agricultural keywords” was another feature motivated by the same observation. When a question is asked, the expert uses this tool to find previous answers that can be reused. e.g. If the expert is searching for “Powdery mildew” disease on “tomato” he could look under “Crops” and refine his search to “Powdery mildew” (Figure 5).

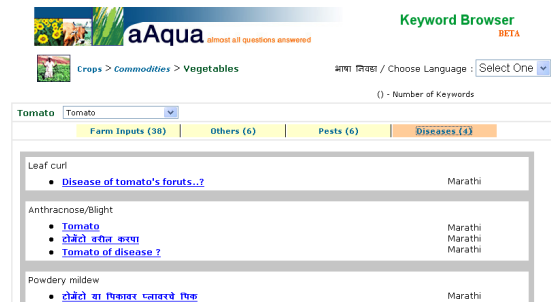


Figure 5: Refining search by crop and disease name

The agricultural keywords are in the noun form and are tagged to enable classification. Retrieval of documents (Recall) by these keywords is being improved by incorporating stemmers and spell-checkers [6]. The limitation is that the precision may be compromised. E.g. If “powdery” is stemmed to “powder” the number of hits increase several times.

A feature has also been provided for experts to help them give quick answers while they formulate more detailed answers. The quick answer contains a hyperlink of keywords

chosen by the expert allowing the user to click and browse through the results while he is waiting for the answer. Although technically possible, we have avoided giving automated answers. There are a number of reasons for this. A review of the question-answer database has shown that while farmers' questions are imprecise (short, sometimes erroneous, wrong grammar, spelling, local language typed in English alphabet, question captured as scanned text, as pictures, audio or video), they expect precise, descriptive answers. Another issue is the temporal validity of historical answers in the database. Yet another challenge is the number of variables that makes each farmer's question unique even though they may be expressed in similar ways on aAQUA. Thus, techniques those assist experts in reviewing past Q&A while answering new questions are more useful and are being incorporated.

E. Integration with Related Content Repositories

aAQUA has been integrated with other sources of information including government schemes for farmers, monthly forecasts and cropping suggestions, market price information (Bhav Puchiye), crop diagnostics (crops library), crop recommendations (crops library) and a glossary of agricultural terms.

The challenge with providing many repositories (although related) is that of organizing and updating the content. We track usage history to decide which repositories need frequent updates. Bhav Puchiye's market price information is updated daily. aAQUA questions are replied in 24-36 hours.

The quality of aAQUA Q&A is measured in the aAQUA-QoS module. The time taken to answer the question is tracked for every question and the minimum, average and maximum time for every month is computed. This can be viewed as an average for all aAQUA experts or can be viewed by individual experts. As each question is answered, our local team follows up with the farmer over a phone to check if he found the answer useful. The feedback is stored as part of the Q&A and a flag is also set to indicate the usefulness. aAQUA-QoS reports allow us to track usefulness and response time of answers by farmer, kiosk and expert. This is possible by maintaining separate registration for kiosks, farmers (done by kiosk operator), experts and other users.

IV. CURRENT STATUS AND THE FUTURE

aAQUA provides a language independent Agri-portal and question-answer forum with textual, graphic and audio-visual elements to provide a simple, yet rich interface accommodating novice users. The users can formulate questions in a natural manner. Keywords are suggested to guide the user to similar questions asked in addition to answers given by Agricultural Experts (many of whom themselves search the repository.) Keywords are stored in a language independent fashion. This allows multi-lingual users to ask questions in their own language and search

content in other languages.

aAQUA integrates multiple databases – farmer, kiosk operator, expert and other user database, Q&A database, farmer feedback data, farmer schemes data, Price Information database (Bhav Puchiye), Crop Recommendations database, Crop Disease Images database and Q&A translations database. The usage of each repository drives the content creation effort.

Currently aAQUA is delivered over kiosk networks, cyber centers and PCs via the Internet and also via handheld devices. aAQUA takes services to people even where they are; leveraging existing skill sets and minimizing the need of a separate expert team by partnering with agri-expert organizations. aAQUA is evolving with the needs of the rural users in mind and is a live corpus for researchers interested in the areas of knowledge management, cross-lingual information retrieval, textual, multimedia and web data mining and data dissemination, web and database caching and prefetching to counter the need for always-on connectivity.

aAQUA content can also be delivered over other media that is more easily accessible to the farmer. We are experimenting with telephony extensions, FM radio broadcasts and mobile Short Messaging Service (SMS) providers. Farmers and traders using the Bhav Puchiye site register for alerts sent over mobile phones and email helping them track prices of certain commodities. Another feature allows the collection of price information from the community.

aAQUA is being spread geographically by building strategic partnerships with the Maharashtra state government (aAQUA is available at www.maharashtra.gov.in), kiosk network providers and Agricultural expert organizations. Additional services being considered for aAQUA as part of the scaling up effort include (i) weather reports (ii) database of populous villages and their location information, (iii) quality standards for exports. Apart from Agriculture, aAQUA, Bhav Puchiye and the Digital Library can be configured and customized for Expert advice and content in Education, Healthcare, eGovernance esp. by organizations who are working in connecting laymen with experts. aAQUA was customized for the dissemination and signature campaign of petitions by Mumbai dwellers for the consideration of the state government.

ACKNOWLEDGMENT

This research has been sponsored in part by Media Lab Asia, Developmental Gateway Foundation and PanAsia ICT Grant. Vigyan Ashram and KVK Baramati are partners in running aAQUA deployments. aAQUA content is free for use by academia. Related Publications: <http://www.dil.iitb.ac.in/pubs.htm>

REFERENCES

- [1] A Bahuman, S Inamdar, R Swami and K Ramamritham, "Robust Network for Rural Areas: study of two of Nlogue's ICT projects (in Maharashtra) and a compilation of the weakest links in their services and IIT Bombay's efforts in addressing them."
<http://www.dil.iitb.ac.in/docs/Interim%20Report-Feb%202005-IIT-Bombay.pdf>
- [2] www.aaqua.org
- [3] S Kagathara, M Deolalkar, P Bhattacharyya, "A Multi Stage Fall-back Search Strategy for Cross-Lingual Information Retrieval", Symposium on Indian Morphology, Phonology and Language Engineering, Kharagpur, February 2005, <http://agro.mlasia.iitb.ac.in>
- [4] K Ramamritham, S Duttagupta, A Joshi, G Mathur and T Vilankar, "An Interface-Driven Approach to Information Provision for Wired and Wireless Customers", Mobile Commerce Proc Int'l Workshop, pp. 102-109, 2003,
http://www.projects.mlasia.iitb.ac.in/bpiconic-cust/index_en.html
- [5] A. Bahuman, C. Rao, S. Nair, "Building Open Source Digital Libraries", National Seminar on empowering the masses through technology application and skill training, <http://www.dil.iitb.ac.in/dil.htm>
- [6] B. Ribiero-Neto and R. Baeza-Yates, Modern Information Retrieval, Addison Wesley, 1999.