

Agricultural Knowledge-Based Systems at the Age of Semantic Technologies

E. Jahanshiri and S. Walker

Abstract—Knowledge has the potential to become data in a sense that it can be used for deriving new ideas, relating new disciplines and more importantly expressing facts in ways that have only been possible through publication of documents. Using methods for quantifying knowledge about agriculture, the underutilised crop knowledge-based system will make the promotion and research on these neglected crops for food and non-food purposes more widespread. The knowledge database will store facts of interest about unknown crops from the peer reviewed and verified sources. The database design will follow standard vocabularies and ontologies that are developed through years by the agricultural community. These standard ontologies will provide ground for the development of semantic products like ‘query system’ and ‘answer engines’ that can provide direct answers to direct questions. The actual information will also be linked to real experimental data created by research groups through the other component that is called collaborative research environment (CRE) that is aimed to provide an online environment for research data curatorship and analysis. In this context standard vocabularies and ontology languages will be used for metadata storage and retrieval as well as relating research data to knowledge.

Index Terms—Knowledge-based system, agriculture, answer engine, semantic product.

I. INTRODUCTION

Abundance of information in any field and discipline provides an opportunity for local and international organizations to build knowledge products that suit the needs of public and researchers. The web interface for information database in these knowledge products serve as one stop query centre for general and specific questions. Researchers can use these knowledge products to search for data/information and define new research projects based on knowledge gaps, new reasoning and new concepts that arise from relating information from different disciplines. Agricultural research has long history and most of these researches are available online in the form of human-readable documents. One challenge regarding the agricultural knowledge is that it is scattered and is not systematically organised in a knowledge database or an ontology/semantic related system. This slows the process of knowledge acquisition by growers and managers that often need direct answers to direct questions.

Manuscript received December 3, 2014; revised February 5, 2015.

E. Jahanshiri is with the University of Nottingham Malaysia, Campus Jalan Broga, 43500 Semenyih Selangor Darul Ehsan, Malaysia (e-mail: ebrahim.jahanshiri@cffresearch.org).

S. Walker is with the University of the Free State, Bloemfontein, South Africa. She is also with the University of Nottingham Malaysia Campus Jalan Broga, 43500 Semenyih Selangor Darul Ehsan, Malaysia (e-mail: sue.walker@cffresearch.org).

Diversification of agriculture through research on underutilised and neglected crops is of no exception to the knowledge dissemination challenge in agriculture domain [1]. Crops for the Future’s Underutilised Crops Knowledge-Based System (CFF UCKB) aims to provide a platform for general public, researchers, agricultural organizations and research funding agencies to promote research and education on these crops. The design consists of a knowledge portal (web user interface), knowledge database, and knowledge toolset including semantic products for knowledge mining and identification of gaps and generation of ideas through a social networking system for the underutilised crops. It will accompany CFF’s Collaborative Research Environment (CRE) to provide an environment encompassing all the research areas, metadata cataloguing/linked data and online analysis tools for the benefit of researchers and scientists working on these underutilised crops.

II. KNOWLEDGE IN AGRICULTURE

Agricultural knowledge has a wide meaning to different players and sectors; farmers refer to it as experience; indigenous and tacit facts, extension and research organizations recognise it as proven good practices that maximises the crop yield, conserves environment etc. Currently the most common form of knowledge exchange in agriculture is knowledge bases and knowledge management tools for document creation and sharing, support/expert systems and information systems [1]. The term knowledge base specifically in agriculture might refer to a document repository system or centre for reports, scientific papers white papers, a forum or a social networking system for knowledge dissemination and publication through print, world wide web, books and so on. For instance E-agriculture, e-agriculture.org, is a global community for exchange of information, My Agriculture Information Bank, agriinfo. in, is a web-based information dissemination created by experts in agriculture and AGRIS database, agris.fao.org, is a document management system that contains augmented bibliographically indexed information in the form of documents. Information is often augmented by addition of datasets and statistics and maps related to the user search query.

These knowledge bases or knowledge management tools have an important role in knowledge exchange around the world. Their output and form of the knowledge exchange however, rely solely on document preparation and sharing. As a result, majority of international research and collaboration results are published documents that are not necessarily useful

for direct assessment of research and immediate knowledge related queries. As an example, International Assessment of Agricultural Knowledge, Science and Technology for Development [2] was an international project to provide available knowledge to stakeholders and the outcome was a report in terms current situation and future “what if” scenario outcomes. The project was finalised on 2008 and produced variety of useful insights. This information however could have been annotated for usage in computer analysis. The research data that was used for the report could have been linked with the knowledge to collate, compare and infer new knowledge. Linked data could provide a simple way to run the analysis again at a later time with new data for the consecutive years.

Another related technology in this area is knowledge management system (KMS) that is a streamlined way to disseminate research findings through articles and papers. Investing on the development of these systems nevertheless is questionable nowadays with the availability of powerful search systems like Google scholar that are capable of indexing thousands of scientific articles and information pages in the shortest time after their publication.

The expert and decision support systems (that are now categorised under knowledge-based systems [3]) are useful tools for aiding the growers in providing “if-then” question and answers and many web-based versions of them are available right now in the agriculture sector for variety of purposes [4], [5].

The combination of the abovementioned tools comprises the current IT infrastructure for knowledge sharing in agriculture. The effectiveness of this infrastructure can be increased with the addition of new software that is available both in terms of knowledge databases and semantic products.

III. KNOWLEDGE-BASED SYSTEMS

Knowledge bases over the internet are becoming increasingly popular. As mentioned before, the dominant form of text contents and articles are only human readable. A ‘Knowledge-Based System’ (KBS) [6] aims to provide query results on the ‘information’ or ‘verified knowledge’ from structured information based on database design or semantic annotations (e.g. meta-data, ontologies). KBS provides knowledge processing and representation tools that can be used effectively for problem solving [7]. The beginning of KBS was development of expert systems and artificial intelligence [3]. These systems however could not provide the type of functionality that is expected from a biological system like human brain that uses complex pattern recognition and information recovery/linkage to derive knowledge and make decisions. Despite these drawbacks, rule-based systems that can perform “if-then” questions can be widely used in many areas like medicine to aid specialists that deal with complex symptoms. In farming, some of these techniques are embed into the decision support [8].

Appearance of knowledge mining and other automated knowledge extraction tools however opens new possibilities to build knowledge systems that extend beyond rule-based decision aiding systems; perform query on stored knowledge

in any form (natural language or formal database with predetermined queries) and relates knowledge concepts to the areas that the researcher is not aware. A KBS can logically consist of a database and a query mechanism. A KBS can be built using variety of techniques from object oriented programming to hierarchical structures. A KBS product can be used in conjunction with an inference engine to provide new knowledge, knowledge gaps and knowledge-data analyses. Research on KBS has recently accelerated due to availability of classification and analysis techniques for information extraction and decision techniques [9].

IV. KNOWLEDGE DATABASES AND STRUCTURED INFORMATION

A Knowledge-Based System is different than a Knowledge base since it contains formal database of information or semantic annotations. The query can be as simple as a SQL routine for retrieving information or by formal questions i.e. “what temperature range is suitable for Moringa plant to grow?” that can be analysed by more complex analysis systems. These formal queries are derived from an annotation/machine-readable database that is populated by information from variety of sources (peer reviewed journals, white papers and informal, verified information from people). The inference engine will provide analysis on the type of question that is asked with older natural language processing techniques [10] to semantic similarity and lookup table analysis [11] on each word and a combination of keywords in a sentence and then perform search on the available information in the database table. The system will then provide output in the form of texts, labels, charts, photos, videos and other available information. Result can be exact or implied information. It can also be augmented with other information/media that might be of interest to the user. Design of this database requires standard definitions or vocabularies in the knowledge domain.

For the agriculture domain, there have been formal attempts to define standard vocabularies by Food and Agriculture Organization of the United Nations (FAO). AGROVOC, aims.fao.org/agrovoc, is a multilingual standard vocabulary that covers all of the areas related to agriculture that are of interest to FAO. It currently covers 32,000 concepts in 20 languages in RDF/SKOS-XL as well as linked-data formats and can be used in variety of ontology related products like building standard databases and semantic products. The current version of AGROVOC contains standard concepts however these tables should be expanded to encompass sub-domain specific vocabularies that might be related to different sets of knowledge e.g. crops that are underutilised in a geographic area. Expansion of this database is done automatically using context driven ontology extraction techniques from available documents [12]. United States Department of Agriculture (USDA) maintains taxonomic portion of Germplasm Resources Information Network (GRIN) that provides the classification and nomenclature for these genetic resources and many other economic plants on a worldwide basis. Included in GRIN TAXONOMY, ars-grin.gov, are scientific names for 26,975

genera (14,281 accepted) and 1,375 infragenera (1,316 accepted) and 103,339 species or infraspecies (61,314 accepted) with common names, geographical distributions, literature references, and economic impacts. GRIN database provides especial data extraction for 9 tables that contains the family, genus, or taxon name, necessary to set up links to GRIN family, genus, and species reports. The combination of the standard vocabulary and standard names for plants and crops will provide ground for the design of two dimensional knowledge databases.

From the semantic standpoint, this knowledge system however is preliminary, considering the wealth of information that is available and also constrains that such a design can impose on the information storage and presentation. This design can be extended to include linked data and references to each piece of data/information that is stored in the database.

Available unstructured knowledge that are already gathered and curated over the Internet has been the target of many semantic research products. An example of a knowledge product that performs knowledge analysis is DBpedia, wiki.dbpedia.org, that aims to extract structured information from the Wikipedia.com articles. The main point here is to perform sophisticated queries against Wikipedia content and link data and information using resource description framework, metadata schema or other ontology related computer readable structures. The other aim of DBpedia is to use the vast amount of information in Wikipedia in new and interesting ways. Another intermediary product on the wealth of Wikipedia information is SemanticMedia Wiki, semantic-mediawiki.org, that provides semantic tools for exporting textural knowledge into a series of annotated products that will make the queries more extensible and robust.

Wolfram-Alpha, wolframalpha.com, created by Stephan Wolfram is computational knowledge engine that performs queries against its internal peer reviewed knowledge sources in order to provide direct answers to the direct questions. The main goal is to provide the exact information that the user has asked rather than an article that 'might contain an answer'. The focus of this product however is on mathematical problems rather than other domain questions like agriculture, although it provides suitable answers on the crop statistics that are already available by agencies like FAO.

Underutilised Crops Knowledge-Based system aims to provide a database of information from unstructured data available on the underutilised crops. This data are mostly in the form of documents. Users will utilise this database to find available knowledge and knowledge gaps related to these crops. Complex queries and knowledge extraction tools will accompany this system to provide a knowledge platform. The database is developed and currently being populated with the information from peer reviewed sources.

Building tools for information insertion and extraction has already begun based on simple web based information insertion tools and also technologies like Saffron, saffron.insight-centre.org, for automatic text mining based on standard vocabularies (AGROVOC). With the availability of the knowledge computational engines and their free and accessible application programming languages (APIs), developing the semantic products for the betterment of

agricultural knowledge and practice is not a farfetched idea anymore.

V. LINKED DATA AND CRE

Linked-data is a method of publishing data online so that it can be of more use than the conventional data. Using metadata standards [6], it is possible to create templates for data publication and sharing. The system will use the same set of data standards to link knowledge data to the knowledge systems. This will allow scientists to find datasets related to specific questions and examine the data. Also this will allow interrelationships between datasets and knowledge to be determined for defining new research projects. There are initiatives around the world and especially in Europe (EU Open Data Portal) to provide linked open data services using semantic web concepts. Linked data using metadata standards make the process of linking knowledge and data much easier and coherent. linkeddata.org provides guidelines, regulations and tutorials and RDF data models for linked data. The CFF knowledge platform uses the same set of standard and specific metadata to link CFF CRE to public datasets as well as datasets created by CFF researchers. The system will also be able to list all other related information like documents and media files based on the query that is performed on the knowledge database. Ready-to-customize templates from Drupal.org and other sources are available now to build custom made content management systems. AgriDrupal is an open source extension of Drupal by FAO AIMS program for agricultural information management. The web based system is using some of these open source initiatives to produce research environments for the benefit of researchers.

VI. SOCIAL NETWORKING AND CROWD SOURCING

Domain specific social networks have recently begun to rise due to the needs for domain specific knowledge dissemination and networking needs. Examples of such networks are mundogeo.com for geographic information, researchgate.net, academia.edu and linkedin.com for citations and discussion, sciencestage.com for knowledge and multimedia sharing among scientists and recent by-invitation-only Elsevier Innovation Explorers social network for specific groups of scientists that might use Elsevier's services. These domain specific social networks tend to be more successful in bringing experts together than the general purpose websites like Facebook. These social networking platforms along with crowd sourcing apps and software can dramatically improve the knowledge on any subject matter. Features like activity stream, user authentication, invitation and recommendation, messaging, status update, location and map APIs are embedded in the new social networking platforms. Ready-to-customize templates like rigg-code.org and pligg.com (both based on Drupal concept) are available for building social networks.

The underutilised crops social networking system as part of the CFF knowledge platform will bring together people involved from normal farmers to scientists. Questions can be easily discussed and solutions can be found for simple and

complex issues. Later semantic analysis on the discussed data will also provide another source of information for the underutilised crops knowledge platform.

VII. REGULATION OF INFORMATION

As important as is the knowledge dissemination, validation of the recorded knowledge and information is also important. Regulatory mechanisms on a knowledge platform can be designed in the form of restrictions on the type of information that can enter in the knowledge bank (only peer reviewed journals etc.) and also sanity checking on answers that the system provides for the users upon asking the questions from the system. If manual control over flow of knowledge and information is needed, a specific user interface will be provided for domain experts to verify information. Regulatory systems such as web based login for experts will be created for the CFF knowledge platform so that automatic and manual checking of the answers can be implemented.

VIII. CONCLUSION

Technologies for the development of knowledge bases that are not in the form of texts and documents have already been developed. A knowledge database is under development for more effective agricultural knowledge dissemination at Crops For the Future. This conventional database design will accommodate information and facts about the underutilised and neglected crops information using the standard vocabularies and other variables of interest about these crops. These standard vocabularies that are in the form of ontologies will provides bases for developing semantic products that will be used to both create natural language answer engines and also semantic discovery and inclusion of new information into the knowledge database. These semantic products are under development following variety of technologies that are already accessible to the developers team. The other aim of using semantic technologies is to use linked data to relate knowledge to experimental data and in this way create new horizons for research at the other component of the system that is called collaborative research environment (CRE) for underutilised crops. Semantic technologies are also being used for preparation of new generation articles and documents that are both human and machine readable.

REFERENCES

- [1] A. Rafea, *Managing Agriculture Knowledge: Role of Information and Communication Technology*, FAO Publications, 2009.
- [2] *IAASTD*, Wikipedia, Dec. 11, 2014.
- [3] *Knowledge-Based Systems*, Wikipedia, the free encyclopedia, Nov. 24, 2014.
- [4] C. P. Yialouris and A. B. Sideridis, "An expert system for tomato diseases," *Comput. Electron. Agric.*, vol. 14, no. 1, pp. 61–76, 1996.
- [5] F. S. Khan, S. Razzaq, K. Irfan, F. Maqbool, A. Farid, I. Illahi, and T. Ul Amin, "Dr. wheat: A web-based expert system for diagnosis of diseases and pests in Pakistani wheat," in *Proc. the World Congress on Engineering*, 2008, vol. 1, pp. 2–4.

- [6] L. Adelman and S. L. Riedel, *Handbook for Evaluating Knowledge-Based Systems: Conceptual Framework and Compendium of Methods*, Springer Publishing Company, Incorporated, 2012, pp. 63–92.
- [7] G. J. Nalepa, "Collective knowledge engineering with semantic wikis," *J. Univers. Comput. Sci.*, vol. 16, no. 7, pp. 1006–1023, 2010.
- [8] A. D. Moore, D. P. Holzworth, N. I. Herrmann, H. E. Brown, P. G. de Voil, V. O. Snow, E. J. Zurcher, and N. I. Huth, "Modelling the manager: Representing rule-based management in farming systems simulation models," *Environ. Model. Softw.*, vol. 62, pp. 399–410, Dec. 2014.
- [9] R. Akerkar and P. Sajja, *Knowledge-Based Systems*, 1st ed., Sudbury, Mass: Jones & Bartlett Learning, 2009, pp. 31–50.
- [10] R. J. Passonneau, C. Weir, T. W. Finin, and M. S. Palmer, "Integrating natural language processing and knowledge based processing," *Advancement of Artificial Intelligence*, 1990, pp. 976–983.
- [11] S. Walter, C. Unger, P. Cimiano, and D. Bär, "Evaluation of a layered approach to question answering over linked data," in *The Semantic Web – ISWC 2012*, P. Cudr6-Mauroux, J. Heflin, E. Sirin, T. Tudorache, J. Euzenat, M. Hauswirth, J. X. Parreira, J. Hendler, G. Schreiber, A. Bernstein, and E. Blomqvist, Eds., Berlin Heidelberg: Springer, 2012, pp. 362–374.
- [12] M. Hazman, S. R. El-Beltagy, and A. Rafea, "Ontology learning from domain specific web documents," *Int. J. Metadata Semant. Ontol.*, vol. 4, no. 1, pp. 24–33, 2009.



Ebrahim Jahanshiri is a PhD in GIS and geomatics engineering, the masters in precision farming from Universiti Putra Malaysia. His BSc was on agronomy and plant breeding from Ferdowsi University of Mashad, Iran.

During his candidacy, he lectured and supervised agriculture and civil engineering students and conducted GIS training for several government and corporate bodies. He has spent over five years as a research assistant, research fellow and research collaborator in universities and agriculture industry in Malaysia, Singapore and Iran. His main focus of research is crop performance monitoring and prediction, web based systems for data analysis and knowledge processing. He is currently research programme coordinator for CropBASE at Crops For the Future, The University of Nottingham Malaysia Campus.

Dr. Jahanshiri is a member of Foundation for Open Access Statistics (FOAS), Pedometrics International Union of Soil Scientists, American Society of Agricultural and Biological Engineers and the Free and Open Source Software and Open Source Geospatial Foundation (OSGeo).



Sue Walker has wide experience in agrometeorological projects across Southern and East Africa, which included projects on crop-climate modelling and agromet services (including seasonal forecasts) for farmers in a wide range of farming systems, focusing on semi-arid conditions. She is a professor emerita for agricultural meteorology at the University of the Free State, in Bloemfontein, South Africa where she has supervised 15 PhD students over the past 15 years. Prof. Walker was awarded a Full bright Scholarship to study for her PhD in Plant Physiology at the University of California, Davis, USA from 1983 to 1988, where she worked under Dr Ted Hsiao on the Spatial Pattern of Leaf Growth of Sorghum as affected by Water Stress and Implications for Canopy Development.

Prof. Walker is the chairperson of "Agrometeorological Services for Agricultural Production" World Meteorological Open Programme Group and has been the South African representative to WMO-Commission of Agrometeorology since 1998 and member of CAgM Management Group, WMO from 2006 to 2014. She is also a member of the International Commission for Irrigation and Drainage (ICID) Working Group on Global Climate Change and Agricultural Water Management (WG-CLIMATE), as representative from South Africa. She was made a Fellow of South African Society of Crop Production, South Africa, in January 2010 for the good service provided to agriculture.

Prof. Walker is the chairperson of "Agrometeorological Services for Agricultural Production" World Meteorological Open Programme Group and has been the South African representative to WMO-Commission of Agrometeorology since 1998 and member of CAgM Management Group, WMO from 2006 to 2014. She is also a member of the International Commission for Irrigation and Drainage (ICID) Working Group on Global Climate Change and Agricultural Water Management (WG-CLIMATE), as representative from South Africa. She was made a Fellow of South African Society of Crop Production, South Africa, in January 2010 for the good service provided to agriculture.