AgroKnowledgeBase (AKB) for plant diseases: Poppy plant use case

Andrew Terhorst, and Ahsan Morshed

Intelligent Sensing and Systems Laboratory, CSIRO, Hobart, Australia
{andrew.terhorst, ahsan.morshed}@csiro.au

Abstract. World's economy drives on crop production. Currently, most of the countries are facing food shortage in each year. Farmers are trying to increase their productivity but they need specific information so that they can take right decision in the right time. One of particular challenge facing farmers is plant disease, which can be defined as deviation from normal physiological functioning that harmful to a plant. In this paper, we proposed a knowledge based prototype called AKB that help farmers to get a better decision.

Keywords: Ontology, Linked Open Data, Plant, Knoweldgebase

1 Introduction

Each year 80 % of the world's economy depends on crop production [1, 2, 9, 10, and 15]. Farmers cultivating crops need good information for increasing productivity. Thanks to modern technology, it is now possible to help farmers take the right decision in the right time. One particular challenge facing farmers is plant disease, which can be defined as a deviation from normal physiological functioning that is harmful to a plant [4, 6, and 9]. Disease can impact negatively on crop productivity. Sensor technology can warn farmers about potential outbreaks of plant diseases. Most of disease information is kept in the relational databases which are not easily shared with others through the world-wide web. It is necessary to break these information silos so that everybody can access this important information.

The principle[8,7] of open data is that data has to convert into the resource description frame work (RDF) and data can be accessed through the uniform resource identifier (URI). RDF data can be understood by human and machine. The RDF is presented in triple format (subject, predicate and object). Each triple is considered as a resource having an URI and data can be downloaded on the fly by suing the URI. Then, this RDF is stored in the triplestore which is also called knowledgebase.

We propose a prototype knowledge sharing system called AgroKnowledgeBase (AKB) where we combine the disease information with dynamic information about environmental factors that contribute to heightened disease risk. This knowledgebase

can be used for populating the disease information in distributed manner; extract knowledge for educating the farmer and facilitate the knowledge for decision support engine so that a farmer can take decision about crops; AKB also gives the facilities to context based awareness services.

This paper starts with the related studies. In section 3, we present the prototype AKB. In section 4, we describe a use case for poppy plant. Finally, we conclude that the prototype can be used for providing knowledge about the plant diseases in Tasmania.

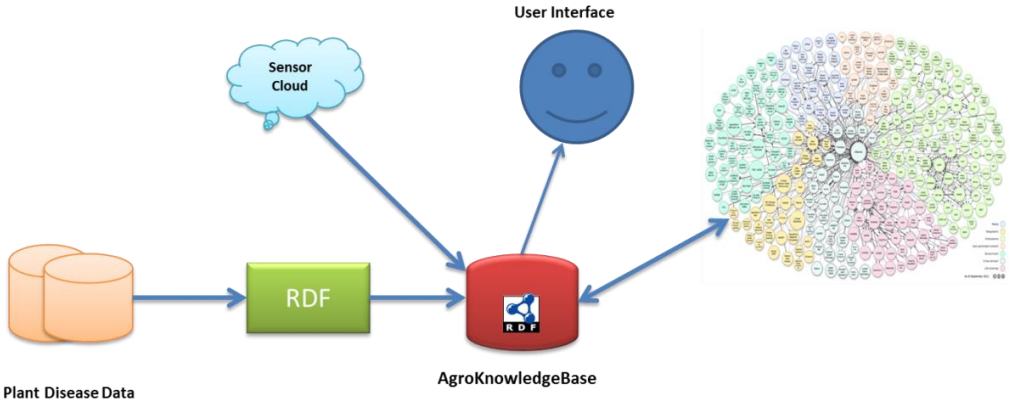
2 Related Studies

The approach is specially-tailored for agriculture science data and more plant data and associated environmental factors. Similar work has been done in the context of life sciences, specifically in the medical domain such as Linked Open Drug Data, Yeast Hub, BioDash and BioGateWay [4, 6]. In the agricultural domain there has been some work developing expert systems around phytopathology for plant diseases [11, 13]. For example, botrytis decision support system[1] is tool for generating alert for botrytis diseases'. None of this work has considered the concept of an open knowledgebase system [12, 17].

3 AgroKnowledgeBase(AKB)

Fig.1. shows a conceptual sketchy view of the AKB prototype for plant diseases:

- **Data** : we collect the legacy plant data and convert them into RDF by using the java programming wrapper and store the data into the triplestore called "AgroKnowledgeBase"



[1] http://www.botrytis.co.nz/

Fig. 1. AgroKnowledgeBase for plant diseases

- **Sensor Cloud:** CSIRO has a sensor cloud platform where all the environmental data are stored. Sensor data comes in different format such as .txt, .shp, .netCDF etc. The CSIRO Sensor Cloud provides an API for accessing the data in JavaScript Object Notation (JSON) format. This format is converted into RDF format using a java wrapper.

- **AKB**: is built on a triplestore where data are store in triple (subject, predicate and object) format. It contains plant disease and environmental data.

- **User Interface**: A user interface shows the RDF data as human readable format.

- **Knowledge cloud**: Linked Open Data Cloud (LOD) is a data hub. Providers have been published their data in an open lenience format in RDF format. Most of the domain data can be found on the cloud. But we considered the Agriculture related data. These dataset can be accessed via SPARQLEndpoint. The SPARQL is a query language for RDF data Set.

The workflow of AKB collects heterogeneous plant disease data from the different sources, converts them in the RDF format and stores theses in AKB. Dynamic data such as temperature, rainfall, and humidity from sensor cloud is also converted and stored in AKB. This dynamic data is used to predict disease risk at run time. A userface has been built on the top of the knowledgebase for showing the data in human-readable format. Finally, we integrate concepts from AGROVOC; the largest Agriculture linked Open Data in the world, and DBpedia. This integration allows large amounts of information to be navigated. All these data can provide the better input for a decision support system [5, 7, 12, 17, 13, and 21].

4 Use case: Poppy

We use poppies as case study because Tasmania is the world's largest producer of opium alkaloid for the pharmaceutical market [3, 16]. About 20 000 ha is under poppy cultivation and the industry is one of the largest employees in the state. The poppy industry is experiencing major disease issues. To solve the problem, we developed an AKB prototype to provide poppy farmers with useful, contextualized, information.

4.1 Plant disease Ontology in SKOS format

Our proposed case study has been tested with data from the "*Plants For A Future*" (PFAF). The PFAF MySQL plant database consists of approximately 7000 species of plants hazard and dynamic information [2]. We convert this legacy format to unified

"Simple Knowledge Organization System" (SKOS) format [19]. SKOS is a W3C standard mostly used by the library community. We use SKOS because it gives flexibility of semantics than RDF/OWL model. Furthermore, SKOS is a relatively easy platform to use.

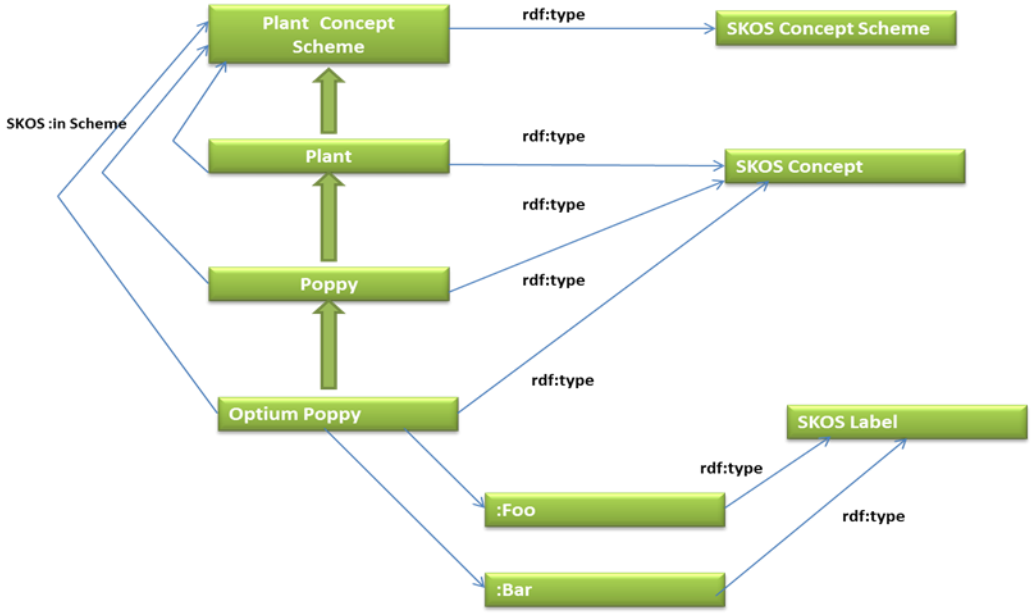


Fig. 2. Concept model for Poppy plant

Fig.2. describes the plant concept scheme where plant is abstract concept, narrow concept poppy and optium poppy. Optium poppy has some characteristics "hasHazard", "hasDemography" etc. More properties are stored in the triple store. For example, "hasHazard" can be used for providing alter to the users and also "hasDemography" can provide the information about the location where this plant can be found.

4.2 Linked Open Data for Plant Diseases

One of principles of LOD is that it needs to provide resource information in a human readable format. We create a simple interface (see Fig.3.) where each page represents as a resource. We use the Pubby tool [8, 18] for Hypertext Markup Language (HTML) presentation and content negotiation by using Hypertext Transfer Protocol (HTTP) access. It also provides download functionality for getting the RDF and number of triple (NT) files.

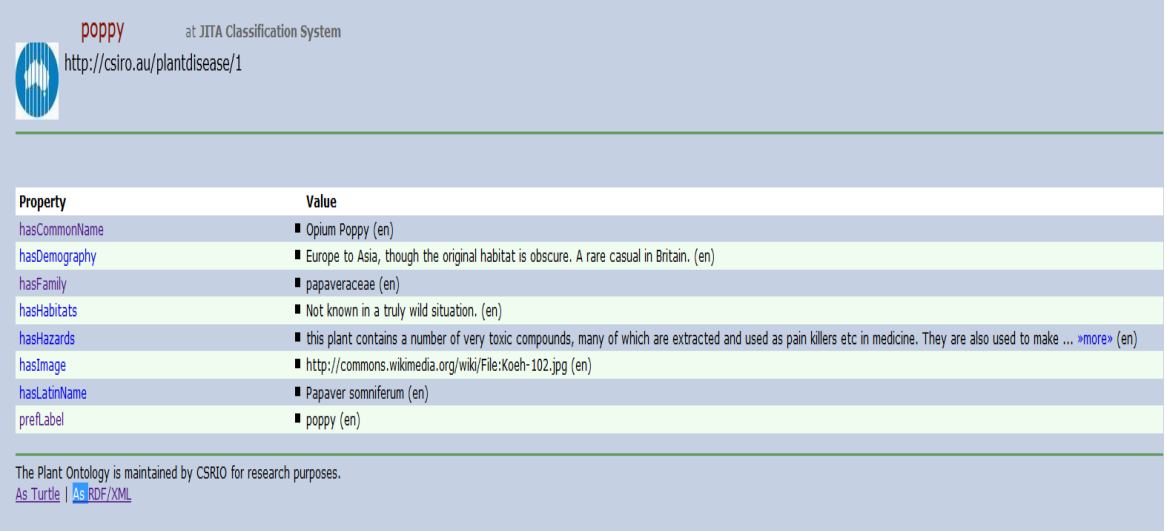


Fig. 3. User interface for AKB

Furthermore, we also provide the SPARQL endpoint by using the Sesame triple store where a user can query the particular information.

5 Conclusion

80% of the world's economy is dependent on agricultural production. Effective management of plant disease remains a significant challenge to farmers. Recent advances in sensor and knowledge management technologies can help farmers mitigate the negative impact of plant disease. AgroKnowledgeBase is an example of an open knowledge management system. This system provides farmers with continuous disease risk assessments by integrating real-time sensor information stored in the cloud with plant disease information. We have successfully demonstrated how AgroKnowledgeBase can be applied in poppy farming. The system can be easily extended to other crop types.

References

1. Coakley, Stella Melugin, Harald Scherm, and Sukumar Chakraborty. "Climate change and plant disease management." Annual review of phytopathology 37.1 (1999): 399-426.
2. Plants for a future. See at http://www.pfaf.org/user/default.aspx
3. Norn, S., Kruse, P. R., & Kruse, E. (2005). History of opium poppy and morphine]. *Dansk medicinhistorisk årbog*, *33*, 171.
4. Walls, R., Smith, B., Elser, J., Goldfain, A., Stevenson, D. W., & Jaiswal, P. (2012). A plant disease extension of the Infectious Disease Ontology. In *ICBO*.
5. Morshed, A., Caracciolo, C., Johannsen, G., & Keizer, J. (2011, September). Thesaurus alignment for linked data publishing. In *International Conference on Dublin Core and Metadata Applications* (pp. 37-46).
6. McRoberts, N., Hall, C., Madden, L. V., & Hughes, G. (2011). Perceptions of disease risk: From social construction of subjective judgments to rational decision making. *Phytopathology*, *101*(6), 654-665.
7. Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *5*(3), 1-22.
8. Tim Berners-Lee. See at http://www.w3.org/DesignIssues/LinkedData.html/.
9. Chet, I. (1987). *Innovative approaches to plant disease control*. Wiley
10. Chiarappa, L. (1971). Crop loss assessment methods. FAO manual on the evaluation and prevention of losses by pests, disease and weeds. Farnham Royal, UK: Commonwealth Agricultural Bureaux [by arrangement with the Food and Agriculture Organization of the United Nations
11. Yialouris, C. P., & Sideridis, A. B. (1996). An expert system for tomato diseases. *Computers and electronics in agriculture*, *14*(1), 61-76
12. Plant, R. E., & Stone, N. D. (1991). *Knowledge-based systems in agriculture*. McGraw-Hill, Inc
13. Yuen, J. E., & Hughes, G. (2002). Bayesian analysis of plant disease prediction. *Plant Pathology*, *51*(4), 407-412.
14. Schriml, L. M., Arze, C., Nadendla, S., Chang, Y. W. W., Mazaitis, M., Felix, V., & Kibbe, W. A. (2012). Disease Ontology: a backbone for disease semantic integration. *Nucleic acids research*, *40*(D1), D940-D946.
15. Crop Production. See at http://www.fao.org/docrep/004/y3557e/y3557e08.htm
16. Fist, A. J. (2000). *The Tasmanian poppy industry: A case study of the application of science and technology*. Tasmanian Alkaloids Pty Limited.
17. Pavan, W., Fraisse, C. W., & Peres, N. A. (2010). A Web-based decision support tool for timing fungicide applications in strawberry
18. Pubby. See at http://wifo5-03.informatik.uni-mannheim.de/pubby/
19. SKOS. See at http://www.w3.org/2004/02/skos/
20. DBpedia. See at http://dbpedia.org/About
21. Caracciolo, C., Stellato, A., Morshed, A., Johannsen, G., Rajbhandari, S., Jaques, Y., & Keizer, J. (2012). The AGROVOC Linked Dataset. *Semantic Web*.