ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Real Time Semantic Search for Agriculture using Ontology and RDF Repository.

Maria Santhia B[1], Saranya R[2], and Nirmalrani V[3]*.

Department of Information Technology, Faculty of Computing, Sathyabama University, Chennai, Tamil Nadu, India.

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 underutilized crop filed 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 , This paper provides solutions to various problems related to weather based Agricultural Census data in these aspects: (1) Storage / Organization of census data using enhanced methods such as ontologies. In this context standard vocabularies and ontology languages will be used for metadata storage and retrieval as well as relating research data to knowledge. By applying NPK values to visualize the clear details from mobile app to the farmers.

Keywords: Agriculture, Ontology, OWL, RDF, Semantic Search.

***Corresponding author**

INTRODUCTION

Since 2003, the Food and Agriculture Organization (FAO) has been concerned with developing a new model for the AGROVOC thesaurus that accounts for semantic and lexical relations in more refined and precise ways, with the objective of building a multilingual repository of concepts in the agricultural domain, the Concept Server (CS). This effort fits in with FAO's overall initiative to establish an Agriculture Ontology Service (AOS) which aims to function as a tool to help structure and standardize agricultural terminology in multiple languages for use by any number of different systems around the world. It will be possible to export the traditional AGROVOC thesaurus, as well as other forms of knowledge organization systems (KOS)1 , from the CS. It will also be possible to extract ontological concepts and use them to build domain specific ontologies. During the research, a number of models and approaches have been studied and proposed. Initially, a relational database was considered an advantageous storage solution because of:

- Its ease of management, scalability, and performance;
- Its similarity to the current format and the ability to ensure backward compatibility;
- The use of RDBs to store other terminologies to be integrated into AGROVOC, such as FAOTERM, FAO Glossary.

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 organized 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.

Diversification of agriculture through research on underutilized and neglected crops is of no exception to the knowledge dissemination challenge in agriculture domain [1]. Crops for the Futures Underutilized 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 underutilized crops. It will accompany CFFs 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 underutilized crops.

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 recognize it as proven good practices that maximizes 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. 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. 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

ISSN: 0975-8585

knowledge to stakeholders and the outcome was a report in terms current situation and future “what if” scenario outcomes.

RELATED WORKS

A. Marathe, R. Harris, D. K. Lowenthal, B. R. de Supinski, B. Rountree, M. Schulz, and X. YuanIn compare the top-of-the-line EC2 cluster to HPC clusters at Lawrence Livermore National Laboratory (LLNL) based on turnaround time and total cost of execution. When measuring turnaround time, they include expected queue wait time on HPC clusters. This results show that although as expected, standard HPC clusters are superior in raw performance, EC2 clusters may produce better turnaround times. Algorithm used in this paper is First Come First Serve and Back Filling Algorithms are used [1].

P. Nath, B. Urgaonkar, and A. Sivasubramaniam conduct an empirical evaluation of the benefits offered by CAS to a variety of real-world data-intensive applications. The savings offered by CAS depend crucially on (i) the nature of the data-set itself and (ii) the chunk-size that CAS employs. They investigate the impact of both these factors on disk space savings, savings in network bandwidth, and error resilience of data. Algorithm used in this paper is implicit algorithm to solve the 3D compressible Navier Stokes equations [2].

Y. Ke and R. Sukthankar examines (and improves upon) the local image descriptor used by SIFT. Like SIFT, our descriptors encode the salient aspects of the image gradient in the feature point’s neighbourhood; however, insteadof using SIFT’s smoothed weighted histograms, we apply Principal Components Analysis (PCA) to the normalizedgradient patch. This paper examines (and improves upon) the local image descriptor used by SIFT. Like SIFT, our descriptors encode the salient aspects of the image gradient in the feature point’s neighbourhood; however, instead of using SIFT’s smoothed weighted histograms, they apply Principal Components Analysis (PCA) to the normalized gradient patch. Algorithm used in this paper is Scale Invariant Feature Transform (SIFT).SIFT algorithm have following four stages (1) scale-space peak selection; (2) key point localization; (3) orientation assignment; (4) key point descriptor [3].

Y. Ke, R. Sukthankar, and L. Huston introduces a system for near-duplicate detection and sub-image retrieval. Such a system is useful for finding copyright violations and detecting forged images. They define near-duplicates as images altered with common transformations such as changing contrast, saturation, scaling, cropping, framing, etc. In this paper SIFT and LHS (Locality Sensitive Hashing) algorithms are used [4].

J. Liu, Z. Huang, H. T. Shen, H. Cheng, and Y. ChenIn devises a location visualization framework to efficiently retrieve and present diverse views captured within a local proximity. Novel photos, in terms of capture locations and visual content, are identified and returned in response to a query location for diverse visualization. Algorithm used in this paper is MPNG (Maximal Near-duplicate Photo Groups) discovery algorithm [5].

P. Indyk and R. Motwani present two algorithmic results for the approximate version that significantly improve the known bounds: (a) pre-processing cost polynomial in n and d, and a truly sub linear query time; and, (b) query time polynomial in log n and d, and only a mildly exponential pre-processing cost O ~ (n) O (1=)d . Algorithm used in this paper is PLEB (Point Location in Equal Balls) algorithm presents two techniques the first is based on a meth d similar to the Elias bucketing algorithm and works for an ylp norm, establishing Proposition. The second uses locality-sensitive hashing and applies directly only to Hamming spaces [6].

ARCHITECTURE OF THE PROPOSED SYSTEM

This paper proposes semantic queries based methodology called FAST. In order to improve query efficiency and reduce operation cost in smartphones, we need to reduce the redundant data such as identifying and filtering redundant data at the client side. Data is transformed into ontology-based repositories like Web Ontology Language (OWL) and Resource Description Framework (RDF).

OWL adds more vocabulary for describing properties and classes: among others, relations between classes, cardinality (e.g. "exactly one"), equality, richer typing of properties, and characteristics of properties (e.g. symmetry), and enumerated classes. Retrieving data from ontology-based repositories done using

SPARQL. The idea behind FAST is to explore and exploit the semantic correlation property within and among datasets. The FAST methodology can be extended to and well suited for multiple data types.

- To facilitate its use for developing agricultural domain terminologies, including ontologies, without requiring the terminologist to start from scratch;
- To enable the development of applications using semantic technologies;
- To enable interoperability between applications using these ontologies.

This paper proposes an ontology for the farmers to find the appropriated crop information to be cultivated during the specified monsoon. The architecture is explained in Figure 1. This architecture clearly stated the enrollment process, the registration process, the NPK values entry process by the farmer which was previously valuated from soil test department. Simulateanously the farmer know the weather details from their corresponding Weather station API. Based on the location and soil test value, the ontology suggest which type for plant would be suitable during the particular monsoon. In certain case, plans requires fetrizler mix to strength the process. Ontology and semantic analysis would recommended during the pest and fertriziler mixing process depends upon the soil test values.

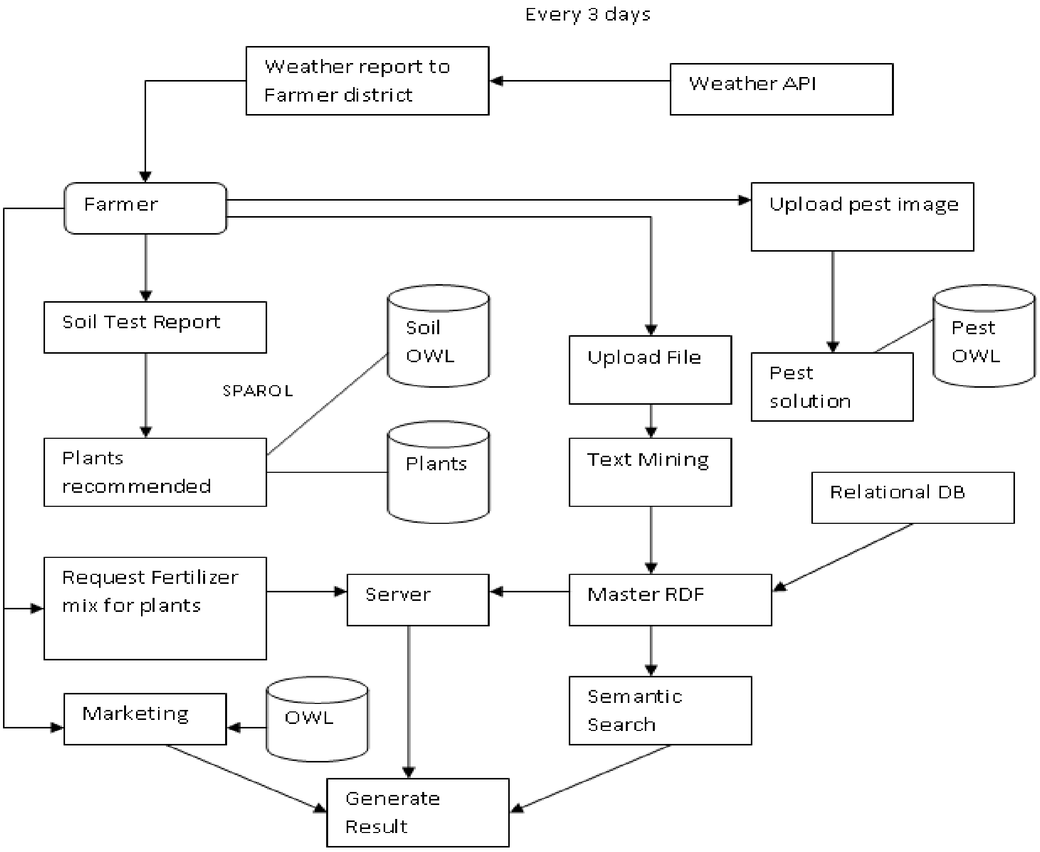

Fig. 1 Proposed System Architecture

AGRICULTURE AND CONTEXT-AWARE SYSTEMS

Context awareness plays a key role in agriculture domain [15]. A context aware system senses its environment and adapts itself accordingly. In order to cater with variations in contextual information, a context aware system is required to perform distributed and parallel processing, especially in the domain of agriculture. An agriculture domain suffers from a range of issues ranging from irrigation and pest management to fertilization. In order for a context aware system to address these issues, following requirements needs to be addressed [16]:

- Acquisition of information related to weather, crop and soil. This includes information such as moisture level of soil, temperature, humidity, type of crop and its properties.

- Continuous (24×7) and real-time monitoring of land.
- Varying crop requirements at the same location.
- Varying fertilizer and water needs for different types of cropped land.
- Distinctive constraints of crops for varying weather, soil conditions and land. The requirements highlighted above are based on specific needs for a problem.

These requirements call for massively parallel and distributed processing in agriculture domain. This can enable monitoring of distributed terrains, dealing with varying crop requirements and employing specific solutions for particular portion/sector of land etc. Besides distributed processing, wireless sensor and actuator networks are needed for data acquisition from field as well as for actuation pertaining to different events. The raw level sensor data obtained can be used to compute high level contextual information that can be used by decision support systems for their operation.

Context Modeling in Agriculture Domain Context modeling enables acquisition of situational information based on currently sensed data. Also termed as situation identification [17], it can be defined as a technique to obtain an abstract view (higher level concepts) of complex, noisy, multi-dimensional and imprecise sensors data. A range of context modeling and reasoning approaches of different levels ranging from very simple to the complex models have been developed for the last 10years.These approaches can be divided into two main groups i.e. Specification based and learning based approaches [17]. Among the former approaches, ontology based modeling has been the most widely used technique [18]. Ontology is the formal and explicit representation of domain knowledge and a way to represent the relationships among the domain concepts. It provides integration of resources (such as over the web) described using ontology based metadata. Ontologies for several domains are built due to its power of persuasiveness, reasoning, inference and classification of knowledge. Ontology may also be formally defined as a 5-tuple as follows [19]: O ={C, S, I, R, A} where, O is the ontology being described, C is the collection of concepts, I is the set of instances of concepts with slots S, R is the restrictions imposed on concepts, instances, slots, and A is the set of axioms. In the domain of agriculture, several efforts have been taken to model contextual information [20], [21-24]. AGROVOC and Advanced Ontology Service (AOS) project was proposed by Food and Agriculture Organization of the United Nations (FAO) for the development of agricultural ontologies based on their multilingual thesaurus [24]. The basic reasons for development of agricultural ontologies are multifold i.e. Ensuring the semantic interoperability among interacting entities and thus enabling digital searching on web. Some of the examples are OntoCrop [22], Crop-Pest Ontology [23], AGROVOC, Irrigation Ontology [22] and AgriOnto [19] etc. To model contextual information of agriculture characterizing a situation e.g. PLANTS [21] and onto Crop etc. In the next paragraphs, the role of ontologies for context description is discussed in detail. The PLANTS ontology [21] framework was developed based on the notion of mixed societies conceptualization [21]. It considers plant as a digital entity such as ePlant and describes the basic ingredients of a plant. The PLANT ontology is composed of two sublayers: the PLANTS-CO (Plants core ontology) that defines a common vocabulary to ensure interoperability, and PLANT-HO (Plants Higher Ontology) that contains eEntity specific knowledge. Similarly, the OntoCrop ontology was developed for the domain of Horticulture. The proposed ontology was used to develop an expert system that identifies pests in crop, different types of diseases and the disorders of some common vegetable crops. The system also deals with control of these diseases in greenhouse. Considering the agricultural structure where the demand is based on the problems or requirement under consideration, ONTAgri ontology [23] is developed (Figure 1). It provides scalability and distributed operations, the basic demand of agriculture domain. It is based on service oriented approach to ontology development that allows extensions by inserting new agricultural services without changing the base ontology.

IMPLEMENTATION PHASES OF THE PROPOSES SYSTEM

Plants Ontology

The farmer has to register details. The server in turn stores the farmer information in its database. Farmer enters soil test value (N, P, K) to the server, now soil test value checks in Soil ontology. Each N, P, K value checks in each data type attribute in Ontology. Plants will be recommended to the farmer by the following strategy: Exact or above N, P, K values are checked with data type attribute in Soil Ontology, if both matches, relevant plants will be recommended first. Next one or two N, P, K values, if matches found then relevant plants will be recommended. Farmer Information saved as an individual in Farmer Ontology.

Pest Ontology and Weather Report

Farmer uploads a pest image to the server. The server checks in Pest Ontology to retrieve the solution of an uploaded pest image. Everyday system monitors weather report of the farmer district from phone db and stores weather information in database. When the farmer login, past 3 days weather status informed to the farmer. If farmer request weather report and server provides solution from Weather Ontology.

Fertilizer Mixing and Marketing

Farmer request fertilizer mix for plants to Server by giving the input as acre, plants, reason. Server checks in Soil Ontology and retrieve N, P, K values of the plant. Then check in Farmer Ontology and retrieve farmer plants N, P, K values. By comparing both values and also checks the reason (Growth, pests), server provides fertilizer mixing for plants. Admin maintains price details of vegetables and fruits. Farmer request price and marketing information of plants, server provides solution from Marketing Ontology.

Semantic Search

The Resource Description Framework (RDF) is constructed for semantic data on a Relational Database containing Structured as well as unstructured data. A Schema is identified for the relational database and a RDF representing the schema of the database is constructed through model provided by the Jena API. The RDF is generated by mining the text contents uploaded by the users in blogs and the contents of the file are analyzed and the Meta contents are manipulated. Similar data's are grouped together that relate to the same resource. The data level processes are subjected to structure level processing by indexing the semantic data elements. Multiple RDF s are grouped and structured together to form a master RDF data that holds all the semantic information's of a Server that support reasoning in any formats of query processing. The Different resources are interlinked with high degree of relational factors by the predicates in the triples. The Query processing is handled directly in the RDF file by iterating the triples forming a discrete relation with the Service query and the URI representing the location of the resource is returned.

Sample Implementation Output Screens

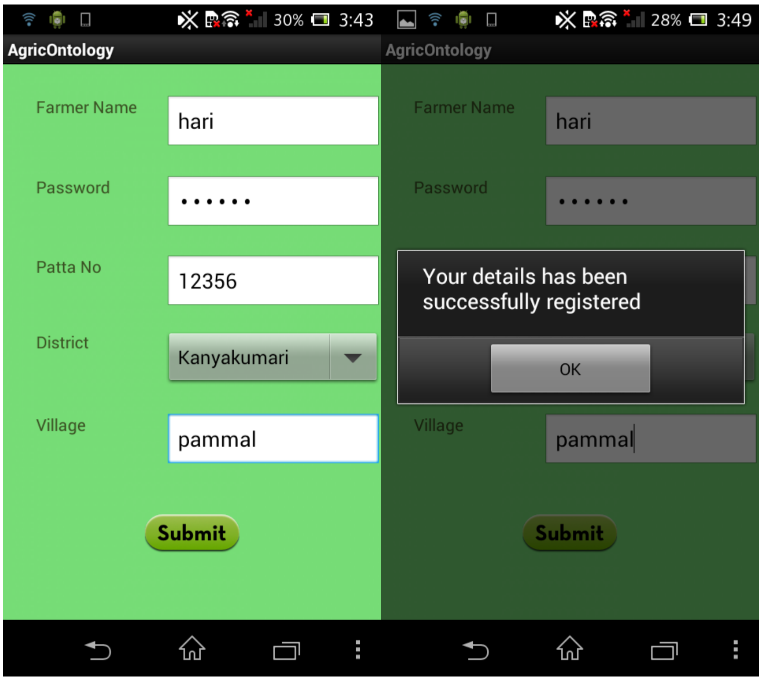

Fig. 2 Registration Page

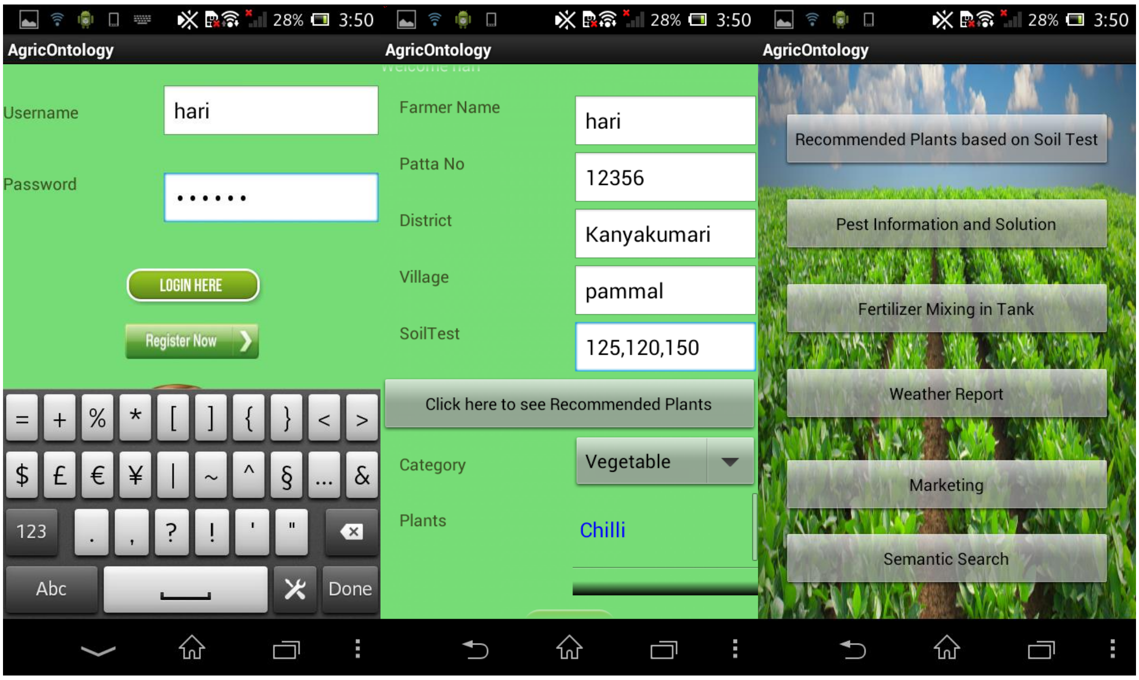

Fig. 3 Functions of Agriculture Ontology

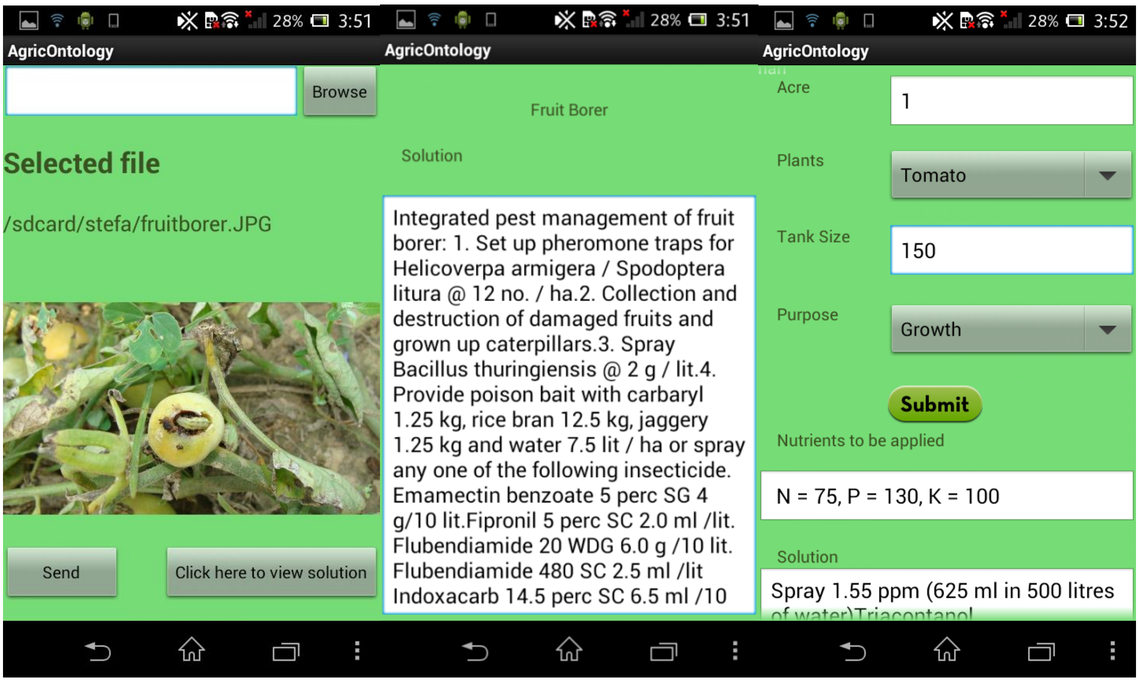

Fig. 4 Pest and Fertilizer Information

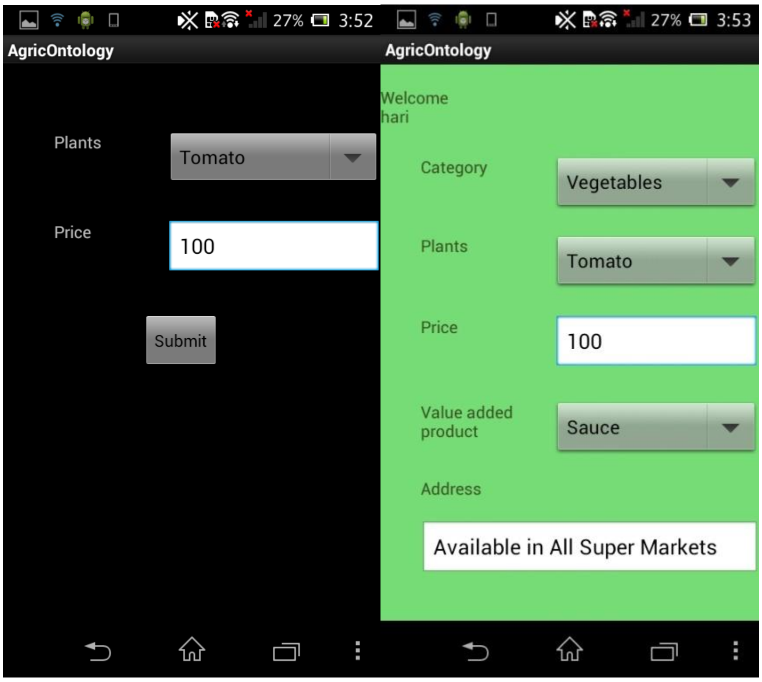

Fig. 5 Marketing Agricultural Products

PERFORMANCE ANALYSIS

This application aims to improve query efficiency using FAST methodology. The operation cost is low in smartphones. The web ontology language is vocabulary for describing properties and classes. SPARQL is used for retrieving data from the ontology. This enables the development of applications using semantic technologies and also enables interoperability between applications using these ontologies.

CONCLUSION

The PLANTS ontology framework was developed based on the notion of mixed societies conceptualization. It considers plant as a digital entity such as ePlant and describes the basic ingredients of a plant. The PLANT ontology is composed of two sublayers: the PLANTS-CO (Plants core ontology) that defines a common vocabulary to ensure interoperability, and PLANT-HO (Plants Higher Ontology) that contains eEntity specific knowledge. Similarly, the agri vocabulary ontology was developed for the domain of Horticulture. The proposed ontology was used to develop an expert system that identifies pests in crop, different types of diseases and the disorders of some common vegetable crops. The system also deals with control of these diseases in greenhouse. Considering the agricultural structure where the demand is based on the problems or requirement under consideration, ONTAgriontology is developed. It provides scalability and distributed operations, the basic demand of agriculture domain. It is based on service oriented approach to ontology development that allows extensions by inserting new agricultural services without changing the base ontology.

REFERENCES

[1] Islam N, Shaikh ZA, Rehman AU, Siddiqui MS (2013) HANDY: A Hybrid Association Rules Mining Approach for Network Layer Discovery of Services for Mobile Ad Hoc Network. Wireless Networks pp. 1961 – 1977.

[2] Aqeel-ur-Rehman, Shaikh ZA (2011) ONTAgri: Scalable Service Oriented Agriculture Ontology For Precisionfarming, in International Conference on Agricultural and Biosystems Engineering pp. 1 – 3.

[3] Chen H, Perich F, Finin T, Joshi A (2004) SOUPA: Standard Ontology for Ubiquitous and PervasiveApplications, in the First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services pp. 233 – 258.

[4] Gutierrez C, Robbes R (2013) WEON: Towards a Software Ecosystem Ontology, in 2013 International Workshopon Ecosystem Architectures pp. 16 – 20.

[5] Islam N, Abbasi AZ, Shaikh ZA (2010) Semantic Web: Choosing the Right Methodologies, Tools and Standards, in International Conference on Internet and Emerging Technologies, Karachi, Pakistan pp. 1 – 5.

[6] Wark T, Corke P, Sikka P, Klingbeil L, Guo Y, et al. (2007) Transforming agriculture through pervasive wirelesssensor networks. IEEE Pervasive Computing pp. 50 – 57.

[7] Fu B, Noy NF, Storey MA (2013) Indented Tree or Graph? A Usability Study of Ontology VisualizationTechniques in the Context of Class Mapping Evaluation. In 12th International Semantic WebConference(ISWC 2013).

[8] Katifori A, Halatsis C, Lepouras G, Vassilakis C, Giannopoulou E (2007) Ontology Visualization Methods – ASurvey. ACM Computing Surveys v: 39.

[9] Prud'hommeaux E, Seaborne A (2008). SPARQL Query Language for RDF.

[10] Seaborne A (2004). RDQL - A Query Language for RDF.

[11] Liu K1, Hogan WR, Crowley RS (2011) Natural Language Processing methods and systems for biomedicalontology learning. J Biomed Inform pp. 163 – 179.

[12] A. Bernstein and E. Kaufmann (2006) GINO - A Guided Input Natural Language Ontology Editor. In TheSemantic Web - ISWC 2006, Cruz I, Decker S, Allemang D, Preist C, Schwabe D, et al.In The Semantic Web- ISWC 2006 pp. 1 – 14.

[13] Fortuna B, Grobelnik M, Mladenic D (2007) OntoGen: Semi-automatic Ontology Editor. In: (GS Michael, SmithJ,)Human Interface and the Management of Information. Interacting in Information Environments. pp. 309 – 318.

[14] Lewis WJ1, van Lenteren JC, Phatak SC, Tumlinson JH 3rd (1997) A total system approach to sustainablepest management. ProcNatlAcadSci U S A pp. 12243 – 12248.

[15] Fukatsu T, Hirafuji M (2005) Field monitoring using sensor-nodes with a web server. Journal of Robotics andMechatronics pp. 164 – 172.

[16] Aqeel-ur-Rehman, Abbasi AZ, Islam N, Shaikh ZA (2011) A Review of Wireless Sensors and Networks'Applications in Agriculture. Computer Standards & Interfaces pp. 263 – 270.

[17] Ye J, Dobson S, McKeever S (2011) Situation identification techniques in pervasive computing: A review.Pervasive and Mobile Computing pp. 36 – 66.

[18] Strang T, Linnhoff-Popien C (2004) A Context Modeling Survey. In the First International Workshop onAdvanced Context Modelling, Reasoning and Management, in conjunction with UbiComp 2004, University ofSouthampton, Nottingham, England.

[19] Khattak AM, Pervez Z, Lee S, Lee YK (2011) Intelligent Healthcare Service Provisioning using Ontology withLow Level Sensory Data.KSII Transactions on Internet and Information Systems pp. 2016 – 2034.

[20] Roussey C, Soulignac V, Champomier JC, Abt V, Chanet JP (2010) Ontologies in Agriculture. In InternationalConference on Agricultural Engineering (AgEng 2010), Clermont-Ferrand, France pp. 1 – 10.

[21] Goumopoulos C, Christopoulou E, Drossos N, Kameas A (2004) The Plants system: enabling mixed societiesof communicating plants and artefacts. Ambient Intelligence pp. 184 – 195.

[22] Maliappis MT (2009) Using Agricultural Ontologies. Metadata and Semantics pp. 493 – 498.

[23] Beck HW, Kim S, Hagan D (2005) A Crop-Pest Ontology for Extension Publications. In 2005 EFITA/WCCAJoint Congress on IT in Agriculture, Vila Real, Portugal pp. 1169 – 1176.

[24] Koenderink NJJP, Top JL, VlietLJv (2005) Expert-Based Ontology Construction: A Case-Study in Horticulture.In 16th International Workshop on Database and Expert Systems Applications (DEXA'05) pp. 383 – 387.