Mobile Based Data Retrieval using RDF and NLP in an Efficient Approach

Dr. N. Srinivasan, Anandaraj Selvaraj

Professor,
Computer Science Department,
Sathyabama University,
School of Computing,
Chennai, India.
anandarajmsccst@gmail.com,

PG Student,
Computer Science Department,
Sathyabama University,
School of Computing,
Chennai, India.
professorsrini@gmail.com

Abstract

Nowadays Youngsters like IT people and various sectors are showing massive interest and moving towards agriculture farming for self sustainable energy for our generation. People are in metro environment with family owned lands they are in need of proper guidance from farmers having huge experience But there is a gap in medium in approaching the farmers by staying with them for months together to get Knowledge Transfer for facing issues and clarifications. The farmers and ancestors traditional methods and subject knowledge need to be transferred to upcoming generations. In Digital world, farmers are using smart phones so to document the knowledge from farmers the Subject Matter Expertise of agriculture (SME) under database and to provide solution by queries in efficient approach through RDF (Resource Description Framework) and Natural Language Processing Technique (NLP) in the form of mobile application internally using SPARQL queries. In this paper, WordNet text mining processing method is used

Keywords

NLP mobile application, RDF mobile app, NLP data retrieval, RDF data retrieval, agriculture on NLP, agriculture on RDF.

INTRODUCTION

We came far away from the roots of our traditional way of Agri cultivation and the generation of us are living a mechanical and medicinal life. Because every where the FMCG goods is imported and the White Revolution and Green revolution were made us forget the traditional methods in Fruits , Vegetables, Paddy varieties. It's a step moving towards building healthy generation. So there is way to re connect the distance between our farmers and our generation.

The amount of data available in internet is in millions and the experience of the real time farmers and their thought process are to be documented and the different ways is mentioned. The New comers have dilemma in deciding the way to follow more over the methods will vary from soil to soil, climate and water regions. Since retrieving data from large agricultural data set will be slow and not optimal at all times so we decided to apply semantic search which improves query efficiency and reduce operation cost by converting the data into ontology based repositories and Resource Description Framework so querying data will be fast. Here SPARQL query language is used to retrieve required data from ontology based repositories. Rest of the paper is organized as follows. New Farmer Registration and their land soil test details in mobile app. Module 1 is Recommended Plants based on Soil Test. Module 2 is Inter Crop Pest Information. Module 3 Weather Report and Marketing. In Section II interprets the related work. Section III explains the Proposed Approach. Section IV explains the Architectural Flow of the project. Section V explains the list of technologies used Section VI covers the result Section VII conclusion and Section VIII Web References of the paper.

RELATED WORK

[1]Yu Hua, Hong Jiang ,Dan Feng, "Real Time Semantic Search Using Approximate methodology for Large Scale Storage Systems" deals with the retrieving of large volume of data from large scale systems in cloud environment.

[2] Y. Hua, B. Xiao, B. Veeravalli, and D. Feng, "Locality-sensitive bloom filter for approximate membership query," IEEE Trans.

This paper deals with Locality sensitive hashing, minhashing technique to reduce the matrix size.. The data set is stored in the form of matrix Since LSH is space efficient and faster in query responsive

[3] J. Chou, K. Wu, O. Rubel, M. Howison, J. Qiang, Prabhat, B.Austin, E. W. Bethel, R. D. Ryne, and A. Shoshani, "Parallel index and query for large scale data analysis". It's a Java Based Index search applied for parallel execution while searching the given query.

This paper interprets the Fast Bit open source software for fast data retrieval by using MPI message passing library and indexing Array into sub arays

[4] C. Papadimitriou, P. Raghavan, H. Tamaki, and S. Vempala, "Latent semantic indexing: A probabilistic analysis".

In the Information Retrieval techniques first time a approach a mathematical based approach of Semantic search indexing on index based files

[5] S. Deerwester, S. Dumas, G. Furnas, T. Landauer, and R.Harsman, "Indexing by latent semantic analysis"

The large matrix of semantic or related data can be grouped together in one single space for fast retrieval based on weightage of terms

[6] Ross, and N. F. Samatova, "ISABELA-QA: Query-driven analytics with ISABELA-compressed extreme-scalescientific data"

ISABELA compresses data from meta data of the original data. Only less than 10% of data can be used creating index from the original data. This is place where this approach differs from other approach.

[7] A. Viswanathan, A. Hussain, J. Mirkovic, S. Schwab, and J.Wroclawski, "A semantic framework for data analysis in networked systems".

A study of spllunk, wireshark like of tools to retrieve the related data in the name of Behavior based semantic analysis framework

[8] Y. Tao, K. Yi, C. Sheng, and P. Kalnis, "Quality and efficiency in high-dimensional nearest neighbor search"

Instead of Locality Sensitive Hashing approach here they go with LSB-tree approach which gives better quality by two orders of magnitude

[9] A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing".

NNS (Nearest Neighbor Search) and LSH (Locality Sensitive Hashing) are the algorithm used to increase the performance measures in data retrieval from large data set.

[10] Q. Lv, W. Josephson, Z. Wang, M. Charikar, and K. Li, "Multi-probe LSH: Efficient indexing for high-dimensional similarity search".

The above paper deals with the Multi-probe LSH in spite of normal LSH since it has less number of hash tables and query execution time is faster

[11] Tanzim Mahmud, K. M. Azharul Hasan, Mahtab Ahmed, Thwoi Hla Ching Chak[2]proposes a rule based approach to query the data from the database by a common man instead of a structured Query Language.

Himani Shukla, Misha Kakkar, Keyword Extraction[8], proposed an approach to extract the keywords from a video instead of providing the entire text data. This could help the user to understand how relevant the video is to their search. They use a regular expression grammar rule to detect the keywords from the video transcript.

PROPOSED APPROACH

Our proposed approach is to develop a mobile based app about agriculture process to get the optimal solution from owl files based on text processing using SPARQL query from NLP language. Data is transformed into ontology-based repositories like Web Ontology Language (OWL) and Resource Description Framework (RDF).Large dataset will be stored in Ontology-based repositories; querying semantic data will be fast. We also develop a mobile based application which guides a farmer from planting to harvesting in a timely manner.

Recommendations for planting based on soil conditions and guidelines for plant growth and current market status regarding the particular crop is also investigated and intimated to the farmers.

Module 1 - Recommended Plants Based on Soil Test

Once the Farmer Registration details done, the new comer enters the mineral details of the soil and the Plants are recommended according to soil test uploaded on server database and the details will be displayed are Duration, soil type, climate, season and quality of seeds, distance and no of plants.

Module 2 – Inter Crop Pest Information

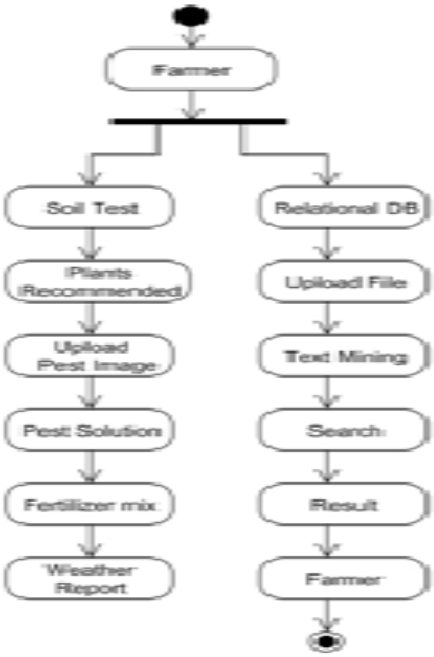
The screenshots of insect which attacks the crop is uploaded and the intercrop to be planted between the crops and pest Information to be applied is retrieved as solution. Sample solution is "Grow Agathi as Intercrop Trreat seeds with imidacluoprid 70 perc WS@12g / kg of seed. Apply carbofuran 3 perc G @ 33 kg. Hes nce this module recommends the maintenance methods to yield a good return.

Module 3 – Weather Report and Marketing

Previous weather details are stored with proper comments and the current weather information is gathered through APIs and monitors the three subsequent day's weather details with the help of Google so the registered user can take necessary steps to protect the crops from rain or cyclone. With respect to the marketing the date in which the crop cultivated and the estimated price is stored and the other registered users for come to know the genuine sale to get mutual benefits for producer and consumer.

1. Architecture Flow

In the below architectural flow diagram initially the flow starts by entering the new farmer details. And then the farmer enters the ratio of the minerals received from the soil test report. Now the mobile app will recommend the plants, vegetables, fruits types based on respective Plant.owl vegetables.owl Pest.owl, Fruits.owl data files. Once the cultivation started based on the recommended methods and if any issues faced by the farmers are rectified through the images uploaded based on the insects name which is nothing but the fil name. Here the text mining is applied and the RDF SPARL query used to retrieve the data from the owl files. The fertilizer mixing and the natural way is mentioned in detailed manner at mobile app. Finally the marketing screen will allow the user to enter the cultivated crops ,fruits, vegetables details quantity and price list.



2. Natural Language Processing

NLP reduces the gap between computer and human by combining the database query into meaningful forms. Hence system manipulates semantic analysis. Here WordNet Processing – wordset is used for identifying the semantic similarities in the words of the dataset. It groups the semantically identical words to the word of interest.

Jena API

Jena API is an open source and it disk based storage and implemented based on metrics such as the load time of ontologies and the SPARQL query response times is very less. It handles small size ontologies efficiently. It supports SPARQL query language and RDQL. Oracle and MySQL are two databases are supported by Jena API.

Ontology	OWL	RDF	Open source
Jena API	Yes	Yes	yes

Data Set of OWL Files are categorized such as Vegetables.owl, Fruits.owl, Flowers.owl.Java Sample Code used to load .owl file is as below.

It's a boom for mobile application development. Here we used the Android studio SDK for developing mobile app are developed in the

Java support

While most Android applications are written in Java, there is no Java Virtual Machine in the platform and Java byte code is not executed. Java classes are

mobile devices with limited memory and CPU. J2ME support can be provided via third-party applications.

Handset layouts

The platform works for various screen sizes from smartphone sizes and to tablet size, and can potentially connect to an external screen, e.g. through HDMI, or wirelessly with Miracast. Portrait and landscape orientations are supported and usually switching between by turning. A 2D graphics library, 3D graphics library based on OpenGL ES 2.0 specifications is used.

Storage

SQLite, a lightweight relational database, is used for data storage purposes.

Native Apps

Android apps are also written in HTML.

Instant Apps

Android apps are hosted on a specific website path and load instead of the website itself. They are part-apps and load almost instantly without the need for an installation. One of the first apps being developed with such functionality.

My SQL

It is the world's most popular open source database. With its proven performance, reliability and easeo-of-use, MySQL has become the leading database choice for web-based applications.

Tomcat Application Server

Tomcat is an application server from the Apache Software Foundation that executes Java servlets and

AndroidStudio

IDE is for app development, based on IntelliJ IDEA. On top of IntelliJ's powerful code editor and developer tools, Android Studio offers even more features that enhance your productivity when building Androidapps, such as: A flexible Gradle-based build system

compiled into Dalvik executables and run on using Android Runtime or in Dalvik in older versions, a specialized virtual machine designed specifically for Android and optimized for battery-powered

renders Web pages that include Java Server Page coding. Described as a "reference implementation" of the Java Servlet and the Java Server Page specifications, Tomcat is the result of an open collaboration of developers and is available from the Apache Web site in both binary and source versions. Tomcat can be used as either a standalone product with its own internal Web server or together with other Web servers, including Apache, Netscape Enterprise Server, Microsoft Internet Information Server (IIS), and Microsoft Personal Web Server.

3. *REFERENCES*

[1] Yu Hua and Dan Feng, "SmartEye: Real-time and Efficient Cloud Image Sharing for Disaster Environments" in 2015 IEEE International Conference on Computer communications

[2] Yan Ke, "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors" in Proceedings of InternationalConference on Electrical Information and Communication Technology (EICT 2015)

[3] "Yufei Tao1 Cheng Sheng1" Quality and Efficiency in High Dimensional Nearest Neighbor Search " in IEEE Transactions on Knowledge and Data Engineering

[4] M. Bawa, T. Condie, and P. Ganesan. Lsh forest: self-tuning indexes for similarity search. In WWW, pages 651–660, 2005.

[5] K. P. Bennett, U. Fayyad, and D. Geiger. Density-based indexing for approximate nearest-neighbor queries. In SIGKDD, pages 233–243, 1999.

[6] S. Berchtold, C. Bohm, H. V. Jagadish, H.-P. Kriegel,andJ. Sander. Independent quantization: An index compression technique for high-dimensional data spaces. In ICDE, pages 577-588,2000

[7] S. Berchtold, D. A. Keim, H.-P. Kriegel, and T. Seidl.Indexing the solution space: A new technique for nearest neighbor search in high-dimensional space. TKDE,12(1):45–57, 200

[8] M. E. Houle and J. Sakuma. Fast approximate similaritysearch in extremely high-dimensional data sets. In ICDE,pages 619–630, 2005.

[9] Q. Lv, W. Josephson, Z. Wang, M. Charikar, and K. Li.ulti-probe lsh: Efficient indexing for high-dimensional similarity search. In VLDB, pages 950–961, 2007.

[10] N. Roussopoulos, S. Kelley, and F. Vincent. Nearest neighbor queries. In SIGMOD, pages 71–79

[11] R. Weber, H.-J. Schek, and S. Blott. A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. In VLDB.

[12] A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing," inProc. 25th Int. Conf.

[13] Y. Tao, K. Yi, C. Sheng, and P. Kalnis, "Quality and efficiency in high-dimensional nearest neighbor search," inProc. SIGMOD Int.Conf.

4. RESULTS

The mobile application for the new comers in the field of agriculture is the enhancement we implemented in this approach and the screenshots are given below

Figure 1 - Home screen

Above Figure 1 is the Mobile App Home page.

Below Figure 2 and 3 Agri app Menu screen and Recommended Vegetable screen .based on soil tests.

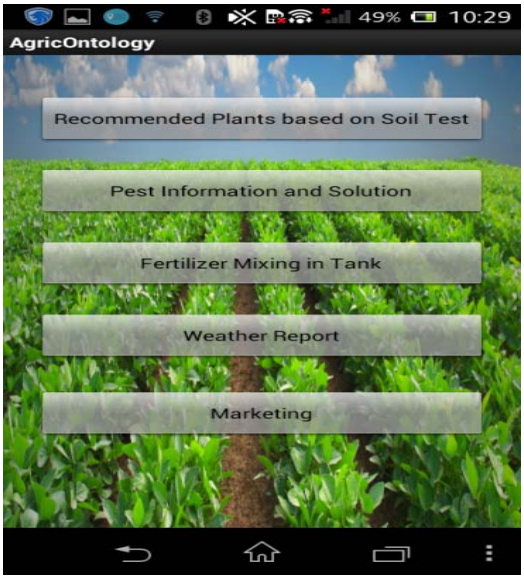


Figure 2: Mobile App Menu screen

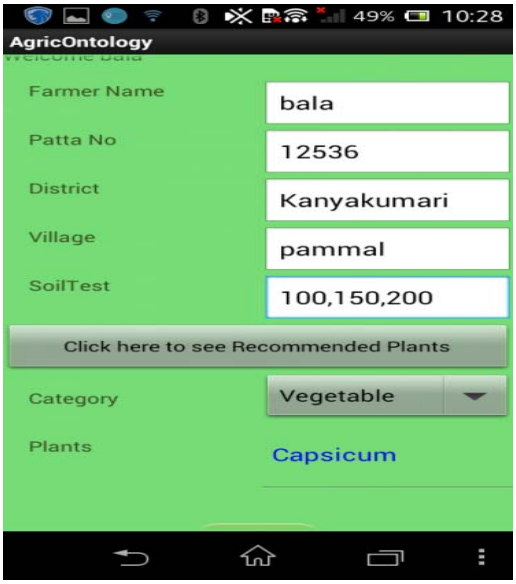


Figure 3: Recommended vegetable screen

The below Figure 4 interprets the Flower recommended based on soil test inputs.

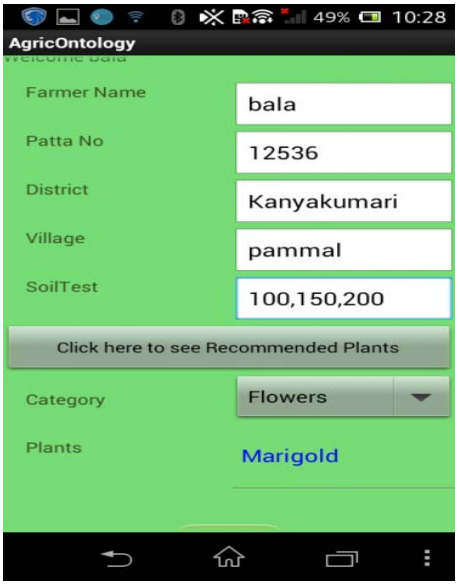


Figure 4: Recommended flower screen

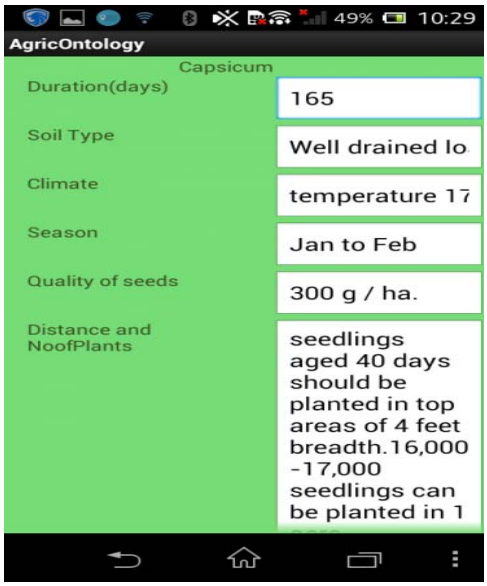


Figure 5: Intercrop solution screen 1

The Figure 5 explains the steps to be maintained and followed while planting the recomended plants as per Figure 3 and 4.

The Below diagram 6 interprets the insects uploaded, attacked the crops and user is expecting the remedy for the same

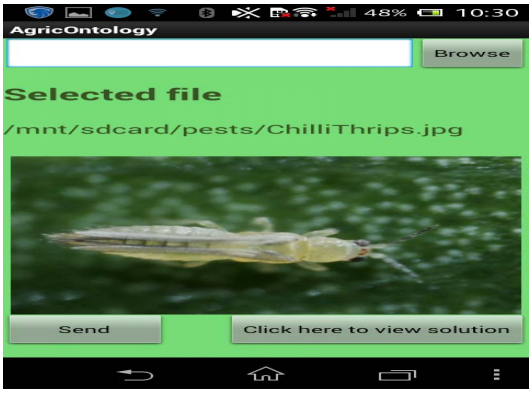


Figure 6: Insects uploaded screen

Figure 7: Intercrop solution screen2

The Figure 7 has detailed solution for the pest attacked in terms of planting inter crop.Below is the Tomcat server started screen for reference.

5. CONCLUSION

Hereby we have gone through an methods and developed mobile based app which bridge the gap between the farmers and new comers in the field of agriculture.