IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

A Semantic Based Accurate Answer Retrieval Technique for Plant Diseases

Johannah Jebaraj[1], Dr. J.G.R Sathiaseelan[2]

Research Scholar, Department of Computer of Science, Bishop Heber College, Trichy[1]

Associate Professor and Head of Department of Computer Science, Bishop Heber College, Trichy[2]

Abstract: This research paper presents an answer retrieval system for retrieving information about plant diseases and their control measures. The Question-Answering system proposed involves a semantic based technique using Ontology technology. The recommended method focuses on answering queries about different types of diseases that affect vegetables and their remedies. The outcome of this technique gives results with greater accuracy and which are more appropriate to context. This method is evaluated with a web forum that contains questions and answers about plant diseases. The evaluation results suggest that the proposed technique is more accurate and efficient.

Keywords: Question-Answering System, Semantic-Based, Answer- Retrieval, Agriculture Ontology.

I. INTRODUCTION

Ontology has been used in several domains since its introduction in the 1990's. Recently it has found great use in the field of Agriculture [1]. Question- Answering Systems (QAS) are used to answer queries in a particular domain. These QAS are different from search engines. QA system retrieves data from a dataset unlike search engines which retrieves data from massive databases. Examples of QA system include Apple's SIRI, Microsoft's Cortana, IBM's Watson, Wikipedia etc. These are the popular and frequently used Question –Answering services [2]. When QAS are created using the ontology technology then they have tremendous potential to provide information that is semantically relevant. Semantically relevant information is more accurate and precise.

In this paper the proposed technique is a Question Answering (QA) system that can answer queries related to plant diseases, their causes and their control measures. Such a system can have a wide application in the domain of agriculture. This system can be utilized by farmers, researchers and those dealing with agriculture data. It provides more specific answers when queries about details regarding the plant's disease. The main technology behind this technique is a robust technology called ontology [11].

Ontology is a representation vocabulary that is specialized to some specific domain. Ontology forms the heart of any system of knowledge representation for that domain. Ontologies can be realized as a 'Web of Data' i.e collection of data from different WebPages [3].The data from different web pages can be combined and reused. The semantic web standards constituted by the W3C consortium help to form a web of data and ways to access this data.

The following sections highlights the recent related works in this research domain as well as the ontologies developed in a similar fashion. It also focuses on the proposed technique and its methodology. Finally its performance is measured and it is evaluated with web forums to determine its accuracy and its relevance.

II. RELATED WORKS

There are many projects and proposals for the use of ontology and QAS in Agriculture domain. Some of them are highlighted below:-

Anusha et al developed a **user-centered ontology for Sri Lankan farmers.** This ontology application was launched as a website, where farmer's queries can be easily answered using SPARQL queries. Agricultural information has strong local characteristics with respect to climate, culture, history, languages, and local plant varieties. These local characteristics as well as the need to provide information in a context-specific manner were the main motivation behind building this ontology application for the Sri Lankan farmers. [4]

Online Agriculture Prescription Recommendation System (OAPRS) is a system that prescribes solutions for farmer's queries and suggests expert's name with whom the farmers can clarify their doubts by telephonic call or by

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

individual chat [5]. Agriculture Ontology Service Research Group of Agricultural Information Institute of Chinese Academy of Agriculture Sciences has carried out many ontology projects in China, pertaining to the domain of Agriculture. Their notable research application includes, FORS- **Floriculture Ontology Retrieve System [6].** This system answers queries related to horticulture especially the flowers found in China.

To assist the farmers in Nigeria and to mitigate Food Security, an ontology-driven information retrieval system is proposed called the **Nigerian Agricultural Ontology** (NAO). This proposal aims at providing important agriculture-related information to help farmers and thereby reduce food imports and achieve food security in Nigeria [7].

Agropedia platform is an agricultural Wikipedia for a wide range of application in agriculture in India and it is developed by Indian Institute of India-Kanpur (IITK). It consists of an online knowledge repository, a social networking platform and content distribution services. This knowledge repository consists of universal meta-model and localized content for a variety of users with interfaces that supports information access in multiple languages [8]. Apart from this application, **Cynthia et al** proposed a semantic based Question Answering System (QAS) for the tomato plant, which served as a major inspiration for the proposal presented in this paper [9].

III.SEMANTIC BASED ANSWER RETRIEVAL TECHNIQUE

There are many methods and techniques available to retrieve data from a given database or from a data store. The proposed technique is designed to retrieve answers from ontology. To get a better understanding of ontology it is important to understand the concept of **Semantic Web**. Semantic web is an extension of the web that helps to represent web of data through the common standards by World Wide Web Consortium (W3C). [10]

Semantics refers to meaning or understanding of the data. This difference engenders a completely different outlook on storing, querying and displaying information. Resource Description Framework (RDF), Simple Protocol and RDF Query Language (SPARQL) and Web Ontology Language (OWL) are the W3C standards that allow creating a **'Web of Data'** also called as **'Linked Data'**. By adding formal semantics and context to linked data, it becomes understandable by the application.[10]

Ontologies are formal definitions of vocabularies that allow structure and relationships between vocabulary terms and members of classes defined. Ontology is queried using SPARQL queries and answers are retrieved in a semantic way which is more appropriate without any duplication. OWL is used to build ontologies using RDFs. Thus OWL is a collection of triples. OWL is what puts the "Semantic" in the "Semantic Web". [11]

The semantic based QA technique has three steps. Fig. 1 presents the steps which are Dataset preparation, Text analysis and Answer search. Dataset preparation and text analysis are done in offline and the answer search from RDF annotation is done in online. In dataset preparation, plenty of details about different types of plants, different types of plant diseases, the controlling and preventive methods of the diseases are gathered from different authorized web forums and web pages. These details are processed manually and formed as natural language questions and answers.

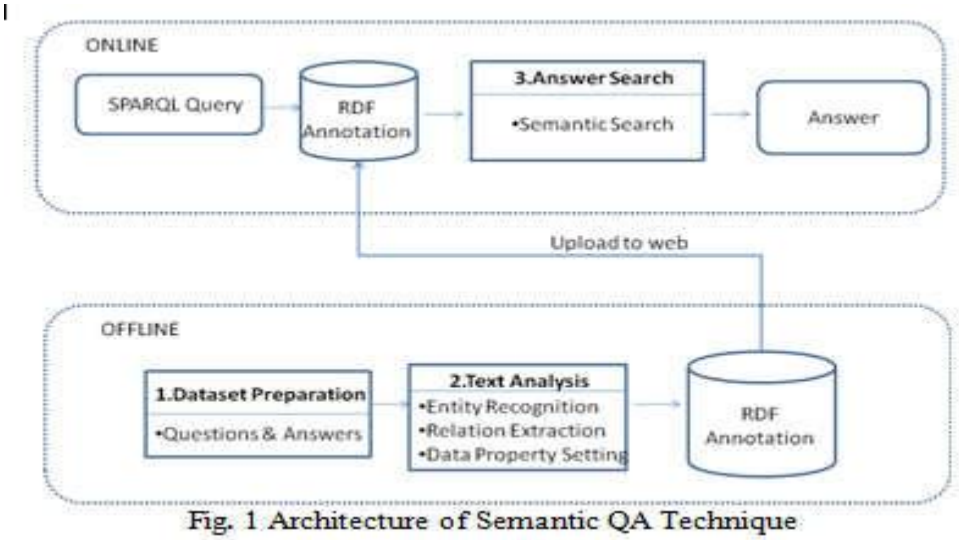

Fig. 1 Architecture of Semantic QA Technique

Fig 1 Architecture of the Semantic Question Answer Technique

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

In text analysis, question and answer dataset is used for entity recognition and relation extraction. In entity recognition, diseases and its classifications, plants and its classifications, symptoms of diseases, bacteria and fungus that cause diseases are identified and organized as classes and subclasses. Each class is instantiated and objects (instances) are created. Then the relationship between objects and the properties of objects are created. After entity recognition, these details are connected with each other as real world objects and the properties are set to the objects. This is known as ontology or web of data which are semantically related with each other. The reference ontology is given in Fig 2 below. This ontology is uploaded to the web. It is referred as RDF annotation that is going to be data source. This annotation is queried by SPARQL. Answer searching is done in semantic way and answer A for the SPARQL query is retrieved. The performance of QA techniques is often evaluated using precision and recall of the returned answers. When using semantic based retrieval techniques, it doesn't retrieve irrelevant answers. So to confirm the accuracy Recall alone is used. The recall percentage of this technique is compared with the recall percentage of the web forums from where the dataset is prepared. To calculate recall, Number of relevant answers from dataset that are not retrieved ARN, and Number of relevant answers retrieved ARR from dataset are taken into account.

$$\text{Recall} = \frac{\text{ARN}}{\text{ARR + ARN}} \text{ X } 100$$

The result findings show that the proposed technique outperforms the performance of the existing work. The developed semantic based answer retrieval technique using ontology performs well in Factoid type questions and List type questions. Yes/No questions, definition type questions and hypothetical questions are yet to be developed. When the SPARQL query is transformed as Natural Language (NL) query, some variations in the recall is possible. The best part is, ontology uses a common set of vocabulary, so that, it could be combined with another ontology in the same domain and act as a huge dataset. Sharing conceptualism is achieved here.

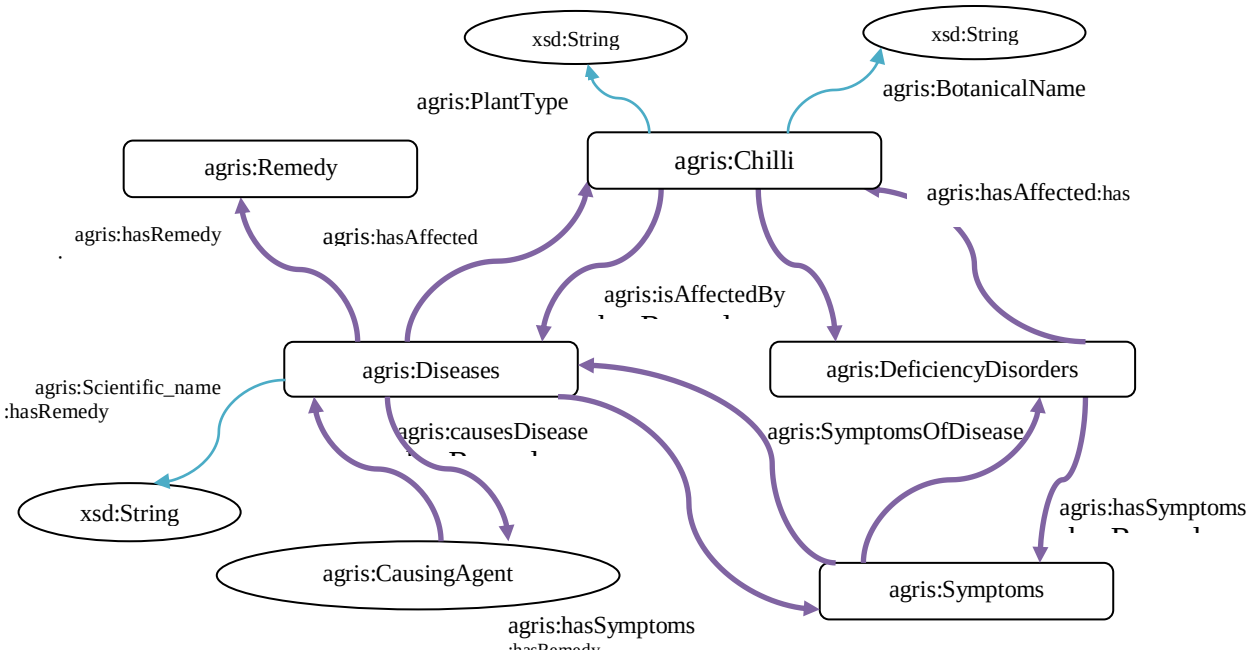

Fig.2. Reference Ontology for Semantic Question Answering System

A number of steps are involved in semantic answering method. The following steps are performed in the proposed Question Answering system:-

STEPS

1) When the SPARQL query, Q is given as input in the SPARQL query tab, it passes to the ontology that is hosted in the local host.
2) Based on the relationship between objects, the answer A is retrieved in semantic manner.
3) Check accuracy of the retrieved answer A, using Recall

$$\text{Recall} = \frac{\text{ARR (No.of Relevant Answers retrieved from Data set)}}{\text{ARR + ARN (No.of Relevant Answers NOT retrieved from Data set)}} \text{ X } 100$$

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

4) Recall is calculated for the retrieved answers from web forums from where the data is prepared and compared with recall percentage of proposed technique.
5) The more the recall percentage, the more the accuracy.

IV. SIMULATION RESULTS AND FINDINGS

The Semantic Based Answer Retrieval Technique was developed using Protégé 5.2.0 version to develop the RDF annotations i.e. the OWL file. The OWL file consists of classes and subclasses created from web forums (http://www.chillilicious.co.uk/chilli-pests-and-problems.html) and (http://www.worldofchillies.com/growing_chillies/chilli_pests_problems_diseases/chilli_diseases/chillidiseases.html) and WebPages such as (http://www.thechileman.org/guide_disease.php) and (https://plantvillage.org/topics/chilli-pepper/infos). All the agricultural information from Tamil Nadu Agriculture University is uploaded in the TNAU Agritech Portal. This portal also served as great source information in building the Ontology File.

The simulation results show that the present technique is very effective as it considers the context and not just the keywords. The results emerging from semantic-based search is more relevant to the users, as they give more accurate and has is not prone to keyword errors.

The screenshot for the various classes and subclasses in the OWL file are given below:-

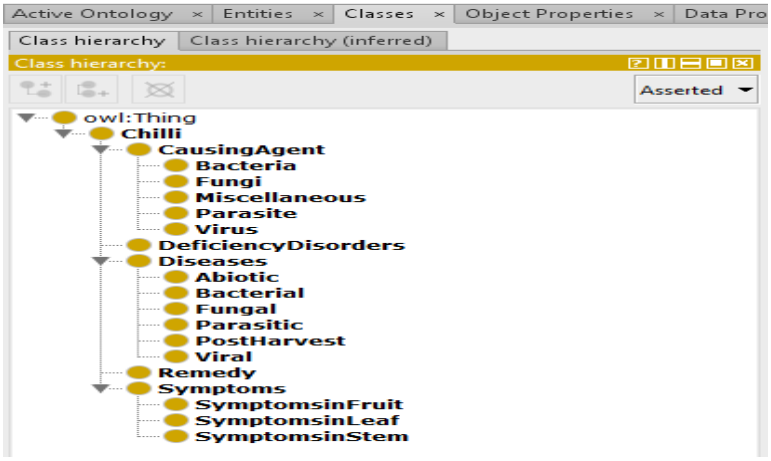

Fig. 3 Screenshot of Classes in Protégé

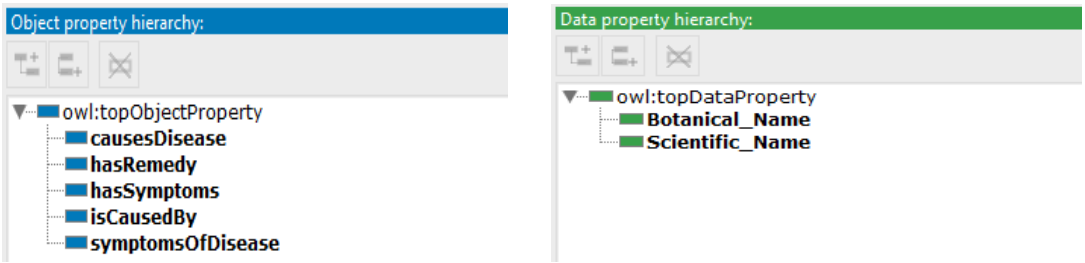

Fig. 4 Object and Data Properties in Protégé

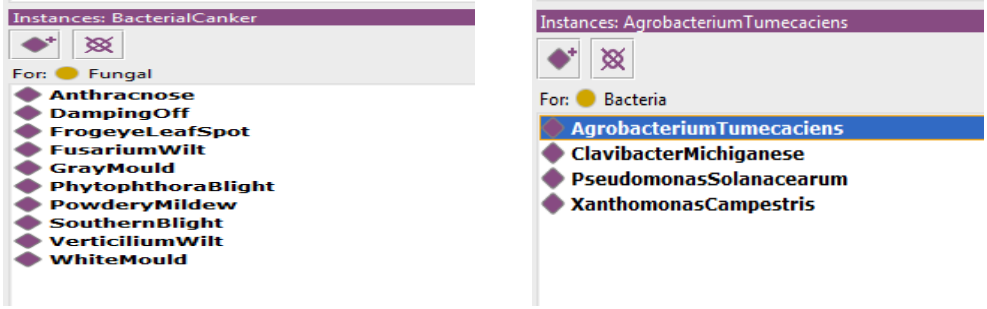

Fig. 5 Individuals of Fungal Diseases and Bacterial Causing Agent in Protégé

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

The ontology is saved as agriont.owl in local disk. Once the OWL file is saved we can now execute SPARQL queries. Before querying the ontology, five questions from factoid types and five questions from list type questions are prepared as Natural Language questions. Here we are considering questions related to the Red Chilli plant known as "Capsicum Anuum", which is the scientific name for Red Chilli.

A. FACTOID QUESTIONS
1. What is the Botanical Name of Red Chilli?
2. What is the scientific name of Bacterial Canker Disease?
3. Which disease has the scientific name "Xanthomonas Campestris"?
4. Which disease has Symptoms of "Water Soaked Spots"?
5. Which Bacteria causes the Disease Bacterial Leaf Spots?

B. LIST TYPE QUESTIONS
1. Lists all the Individuals from Fungal Diseases in Red Chilli.
2. List out the Scientific Names of all the Bacterial Diseases in Red Chilli.
3. Lists all the diseases of Chilli plant
4. List out all the symptoms of Bacterial Canker Disease in Tomato
5. List out all the Bacteria causing Diseases.

C. SPARQL QUERY
The SPARQL query to find out the Botanical Name of Red Chilli is given below:-

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX agris: <http://www.johannah/agriont.owl#>

SELECT ?name
WHERE {
  agris:RedChilli agris:Botanical_Name ?name
}
```

This query is executed in the Local Server in the Protégé Software. The results are retrieved in a semantic manner and the output is given with great accuracy.

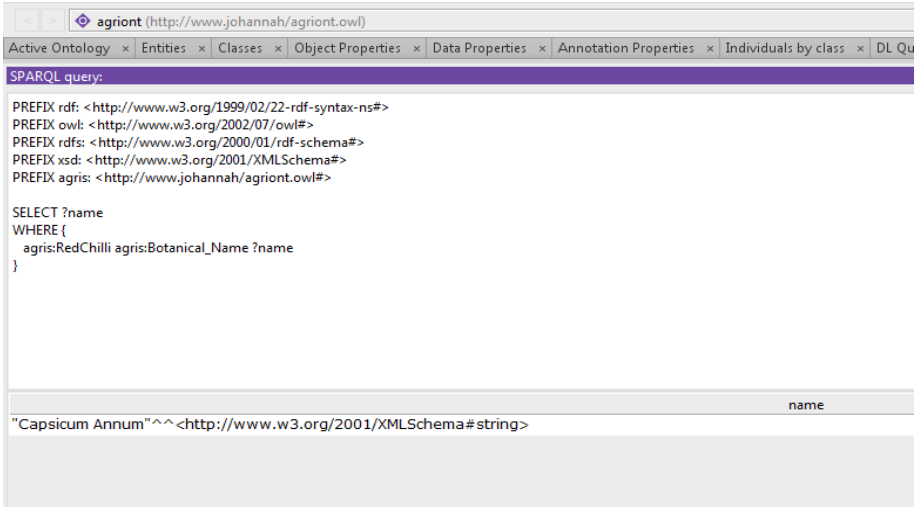

Fig 6.SPARQL Query and Results

All the 10 questions are executed in the SPARQL window and the results are noted down. These answers are then compared to the answers acquired from web forums relating to diseases in Red Chilli plant. By comparing the results from the web forums with the results in the semantic based answering system, we can easily conclude that the proposed technique is much more accurate and gives perfect context relevant answers than the web forums.
The Fig 7 given below depicts the comparison of the results of List Type questions in the web forums as well as in the proposed semantic system.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

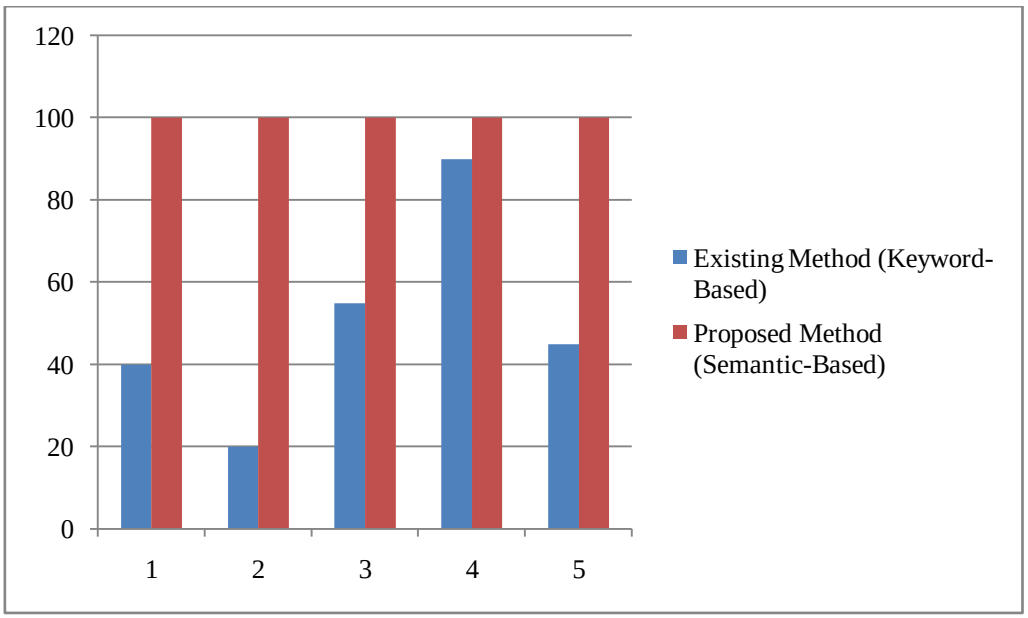

Fig. 7 Comparison of Existing and Proposed Method based on Recall Percentage

V. CONCLUSION

The proposed semantic based answer retrieval technique helps the users such as farmers or Agriculture researchers to get details on plant diseases which they need more accurately. Many answering systems presently existing communicate with the farmers through Short Message Service (SMS) or by making direct phone calls. But by using this QA system the users can get more semantically related data with less error. It is available always and so users can get their queries answered without any delay. The proposed system can be implemented either as a website or as a mobile application, thus it can cater to a huge audience.

In future we can also use Image Processing Techniques to include pictures with queries and results. This will be more useful to the farmers. It will also be very beneficial if Natural Language Processing were used to translate the SPARQL queries into natural languages, especially in local languages so that it can be used in different parts of the country.

REFERENCES

[1] J. Johannah and J.G.R. Sathiaseelan, "An Exploratory Study on Agriculture Ontology: A Global Perspective", International Journals of Advanced Research in Computer Science and Software Engineering, vol. -7, no. -6, pp. 202-206, 2017.

[2] "Question answering", En.wikipedia.org, 2017. [Online]. Available: https://en.wikipedia.org/wiki/Question_answering. [Accessed: 26- Jul- 2017].

[3] B. Chandrasekaran, J. Josephson and V. Benjamins, "What are ontologies, and why do we need them?", IEEE Intelligent Systems, vol. 14, no. 1, pp. 20-26, 1999.

[4] A. Walisadeera, A. Ginige and G. Wikramanayake, "User centered ontology for Sri Lankan farmers", Ecological Informatics, vol. 26, pp. 140-150, 2015.

[5] Qingtian Zeng,, Zhichao Liang, Weijian Ni, Hua Duan, "OAPRS: An Online Agriculture Prescription Recommendation System"7th International Conference on Computer and Computing Technologies in Agriculture, 2014,pp.327-336.

[6] X. SU, J. LI, Y. CUI, X. MENG and Y. WANG, "Review on the Work of Agriculture Ontology Research Group", Journal of Integrative Agriculture, vol. 11, no. 5, pp. 720-730, 2012.

[7] E. Ukpe, "Agriculture Ontology for Sustainable Development in Nigeria", IOSR Journal of Computer Engineering, vol. 14, no. 5, pp. 57-59, 2013.

[8] D. P, M. G, P. PB, S. N and R. Akerkar, "A Comprehensive Agriculture Ontology: Modular Approach", International Journal of Innovative Science, Engineering & Technology, vol. 1, no. 10, 2014.

[9] T. Cynthia and P. Calduwel Newton, "A Semantic Based Answering Technique Using Ontology in MCC", International Journal in IT and Engineering, vol. 04, no. 11, pp. 9-13, 2016.

[10] V. A. and A. A., "Semantic Web Mining using RDF Data", International Journal of Computer Applications, vol. 133, no. 10, pp. 14-19, 2016.

[11] T. Cynthia and P. Calduwel Newton, "A Refined Latent Semantic Analysis (LSA) Technique to Improve Accuracy of Agriculture Data in Mobile Cloud Computing (MCC)", International Journal of Control Theory and Applications, vol 9, no. 27, 2016.