Ontology-based Advisory System for Cotton Crop Farmers

S. Narayana
Professor, Gudlavalleru Engineering College
Gudlavalleru, Andhra Pradesh, India

ABSTRACT
Different parts of the country have recently witnessed a large number of suicide deaths by the farmers who found themselves in extreme distress. A variety of factors are held as responsible for the farmers' suicides which includes a) adverse seasonal conditions, b) lack of irrigation facilities, c) massive pest attack, d) wrong agricultural practices and failure of extension services: excessive use of pesticides, mono-cropping, wrong selection of seeds, dependency on private dealers for technical advice, e) spurious pesticides and seeds, f) increase in the cost of cultivation, and d) low yields and adverse prices. To address some of these problems, this work proposes the development of an expert system (Agro-Advisory System) which provides valuable answers to farmer queries. Agro-Advisory System is an ontology based knowledge system that provides query-answering support for farmers. Farmers encounter many questions on cropping depending on climate, diseases on crops, pests, preventions, timelines of various activities etc. To handle these queries, the old procedure was to collect the questions from farmers and record these questions, then search for the answers to these questions manually from documents provided by agricultural experts. Sometimes lack of experts and their expertise creates problem, so we choose expert system approach to solve this problem.

Keywords
Ontology, Cotton crop, Cotton ontology, RDF, Knowledge-based system.

1. INTRODUCTION
India has the largest cotton area in the world with about 96 lakh hectares under cultivation accounting for one-fourth of the global cotton area. It contributes to 16% of the global cotton produce and has emerged as the world's second largest cotton producer in 2006-07, edging past the USA, which held the second rank until recent past. China is the world's leading cotton producer [1]. It has been estimated that cotton contributes to approximately 30% of the Indian agricultural gross domestic product and considerable export earnings [2]. The organized sector of the Indian textile industry constitutes the largest single industrial segment in the country in terms of annual value of output and labour employed, both direct and indirect. There are more than 1,500 spinning mills, 250 composite mills in India having an installed capacity of approximately 35 million spindles and more than 1 lakh handlooms. The decentralized sector comprising power looms and handlooms, provide employment to over 2.5 million people. The fabric quality of handlooms has been widely acclaimed for fitness and comfort. Thus, cotton is an immensely important crop for the sustainable economy of the country and livelihood of the Indian farming community.

But unfortunately, different parts of the country have recently witnessed a large number of suicide deaths by the farmers who found themselves in extreme distress. A variety of factors are held as responsible for the farmers' suicides which includes a) lack of guidance b) adverse seasonal conditions, c) lack of irrigation facilities, d) massive pest attack, e) wrong agricultural practices and failure of extension services: excessive use of pesticides, mono-cropping, wrong selection of seeds, dependency on private dealers for technical advice, f) spurious pesticides and seeds, g) increase in the cost of cultivation, and h) low yields and adverse prices [3].

To address some of these problems, this work proposes the development of an expert system (Agro-Advisory System) which provides valuable answers to queries regarding type of soil/climate for a particular crop, type of pests within crop, different diseases and their cure, and timelines associated with each activity etc.

Agro-Advisory system is a flexible querying system which allows users to give the flexibility of querying the system as they like. Knowledge acquisition is done with the aid of Agro experts. Farmers encounter many questions on these cropping systems depending on climate, diseases on crops, pests, preventions, timelines of various activities etc. To handle these queries, the old procedure was to collect the questions from farmers and record these questions, then search for the answers to these questions manually from documents provided by agricultural experts. Sometimes lack of experts and their expertise creates problem, so we choose expert system approach to solve this problem.

In order to help farmers on the agricultural issues related to Cotton Crop, we are proposing Agro-Advisory System provides valuable answers to queries regarding type of soil/climate for a particular crop, type of pests within crop, different diseases, and timelines associated with each activity etc. In particular the objectives of this research include;

- To design and construct crop ontology (also called as knowledge repository) that provides information about crop, pests, diseases, climates, timeliness and associated timelines with various activities involved for cotton crop.
- To build Agro-Advisory system over this ontology.
- To use Agro-Advisory system to provide useful answers to queries posed by cotton crop farmers.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 describes the methodology of the proposed work. The experimental results are presented in section 4. Section 5 draws some conclusions and possible future work.

2. RELATED WORK
Cotton is the most important commercial crop of our country contributing up to 75% of total raw material needs of textile industry and provides employment to 60 million people. India has the largest area under cotton cultivation with relatively low productivity primarily due to the large area under rain fed cultivation with inadequate supply of inputs. Area wise, India ranks first in world, but it ranks second in production next to China. Only in India, all the four spinnable fibre yielding

species of gossypium viz., gossypium hirsutum, gossypium barbadense, gossypium arboreum and gossypium herbaceum are cultivated commercially. Hybrid cotton cultivation in about 45% of total cotton area contributing 55% of production is a significant milestone achievement in Indian cotton scenario [4].

Cotton is often referred as the White Gold/Kapas, consumes more than 45% of the total pesticides used in our country. Cotton is attacked by several insect pests reducing the crop yield to a greater extent. The insect pests that attack cotton crop may be classified into sucking insects (Aphids, Jassids, Thrips and White fly), Bollworms (American, Pink, Spotted Spiny), Stainers, Stem Feeders, Folliage Feeders etc. Of the total pesticides used in Indian Agriculture, about 45 per cent is sprayed on cotton crop alone. To reduce pesticide usage in cotton, several strategies like use of Genetic Resistance to insect pests, Integrated Pest Management (IPM), Insecticide Resistance Management (IRM) etc. are advocated. In recent times, BT cotton technology is found to be one of the best strategies to manage bollworms, the most important pest of cotton [4] [5].

Cotton agronomy research is oriented to develop best agro techniques for attaining maximum yield from cotton varieties and is very useful recommendation for cotton production technology that have been evolved under prevailing agro-climate of the State.

Unfortunately, the research knowledge repositories for plant production are not well organized and utilized. This is a main obstacle for the researchers to make use of the previous studies. Conventional search engines cannot interpret the sense of the user‘s search and often the ambiguity of the query leads to the retrieval of irrelevant information. BANKS [6] enables users to extract information in a simple manner without any knowledge of the schema or any need for writing complex queries. Knowledge-based applications on the other hand are applications that retrieve information from a knowledgebase like ontology. They are developed on the domain knowledge of a system.

A number of systems are now emerging that use techniques to apply ontology-based domain-specific knowledge to the indexing including: similarity evaluation, results expansion and query enrichment processes [7-9]. Ontology has been moving from the domain of Artificial-Intelligence laboratories to the desktops of domain experts. An important role of ontologies is to serve as *schemata* or *intelligent view* over information resources. Thus, they can be used for indexing, querying, and reference purposes over non-ontological datasets and systems.

The Rice ontology (RO) [10] developed in Japan, is an ontology specialized for genome informatics of rice and has been developed as biological domain ontology, applies to all organisms useful to exchange genome informatics. Up until now, ontology on a specific cotton crop has not been constructed. This research is therefore a pioneer work to develop an ontology prototype for plant production using the cotton production as a test case study. This prototype will be a model for other agricultural ontology developments in the future.

3. METHODOLOGY

In order to provide useful answers to cotton crop farmers, Agro-Advisory System uses the cotton crop ontology as a knowledge base. Following section describes the construction of cotton crop ontology.

3.1 Construction of Cotton Crop Ontology

To construct the Cotton Crop Ontology (CCO), data pertaining to cotton crop was collected from agricultural experts from different parts of Andhra Pradesh. The CCO was constructed from scratch in consultation with domain-experts. This ontology covers the domain of cotton production from cultivation to harvesting. Relevant knowledge related to cotton production was analyzed.

The classes describe the concepts of cotton crop domain like Crop, Soil, Activity, Pest, and Disease etc. The relationships are used to connect the classes. For example, the relationship *affected_by* connect the classes Crop and Disease which can be interpreted as **Crop *affected_by* Disease.**

3.1.1 Concepts of Cotton Crop ontology

The concepts of Cotton Crop Ontology were categorized as classes to provide an initial comprehensive framework that will incorporate every other relevant concept. The categories are summarized in the following table 1.

Table 1: Concepts of Cotton Crop Ontology

Object Concept	Process Concept
Crop	Cultivation process
Climate	Harvesting process
Soil	Protection and control process
Fungi (pest and natural enemy)	Fertilizing process
Bacteria (pest and natural enemy)	Irrigation process
Virus (pest)	Soil preparation
Growth Stage	
Pesticide	
Disease	

Relationships connect classes (concepts) and provide meaning to the data. Table 2 shows the relationships present in the CCO.

Apart from the following relationships, a large number of hierarchical relationships such as *is-a* and *are* are used to represent the class hierarchies.

Using the collected cotton crop data, cotton crop ontology is created using Protégé tool [13]. Ontology is an explicit specification of a conceptualization, where a conceptualization is an abstract, simplified view of the world that we want to represent for some purpose [11]. It is a collection of concepts and relationships between them. Developing ontology is a complex task that requires a high degree of analytical and abstract thinking [14-17].

They are several semantic languages that are used to represent the ontology. However, the constructed ontology was represented in OWL (Web Ontology Language) [12] as it has more expressive power than RDF and RDFS. It is a W3C recommended language for semantic-based systems and is been supported by many free open source APIs (Jena, Protégé, Sesame, Swoop, etc.) and permits easy data transfer and reuse. The Cotton Crop ontology contains 162 concepts and 496 instances and 66 relationships and provides an organizational framework that allows reasoning about cotton crop.

Table 2: Relationships of Cotton Crop Ontology

Relationship	Inverse relationship	Type
hasSubClass	isSubClassOf	Hierarchical relationship
hasLexicallization	isLexicallizationOf	Equivalence relationship
hasSynonym	isSynonymOf	Equivalence relationship
hasCroppingSystem	isCroppingSystemOf	Associative relationship
hasVariety	isVarietyOf	Associative relationship
hasBotanicalDescription	isBotanicalDescriptionOf	Associative relationship
hasCropProfile	isCropProfileOf	Associative relationship
hasGeneralDescription	isGeneralDescriptionOf	Associative relationship
hasProductionPractice	isProductionPracticeOf	Associative relationship
influences	influencedBy	Associative relationship
hasComponent	isComponentOf	Associative relationship
expressedBy	isExpressionOf	Associative relationship
composedBy	isComposedOf	Associative relationship
hasFamily	isFamilyOf	Associative relationship
hasGenus	isGenusOf	Associative relationship
hasSpecies	isSpeceOf	Associative relationship
hasOrogin	isOriginOf	Associative relationship
Affects	isAffectedBy	Associative relationship
precede	precededBy	Associative relationship
useProcess	isProcessFor	Associative relationship

3.2 Proposed System

The main objective of this research work is to design a search interface that allows the farmers to post their queries related to Cotton crop which are answered by performing a search over the Cotton crop ontology. Examples of few queries which can be answered are;

- What are the symptoms of cotton curl disease?
- If cotton leaves dries and color changes to red then what is the disease?
- How to cure blight disease in cotton? Etc.

Information about cotton crop is represented as ontology that describes the practices and farming techniques for cotton crop. It gives complete information on good and bad practices of cotton. The concepts which best describe cotton crop are soil and climatic conditions, recommended varieties of cotton which can be location specific, diseases affecting the crop and the reasons for their occurrence, their symptoms, and their cure; similarly information on pests attacking the crop, their precautions and cure. The ontology also contains concepts on various activities of farming, hoeing, sowing, irrigation, fertilizing, spraying and harvesting, along with their timeliness.

A user application is designed and developed using this crop ontology to effectively answer farmer queries. The architecture of Agro-advisory System is shown in Figure 1.

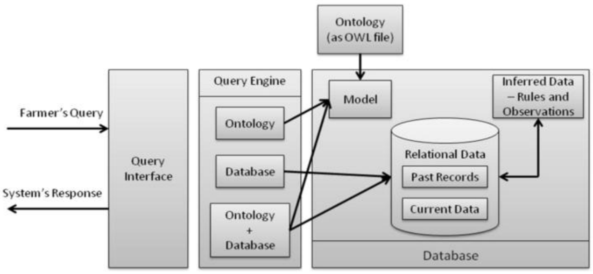

Fig 1: Agro-advisory System Architecture

As shown in Figure 1, the proposed system consist the following components.

3.2.1 Query Interface

This is an interface where farmer can post his query. Query is mainly a keyword-based query. This query is parsed, tokenized and mapped to resources in the ontology. Algorithm 1 is used to tokenize the input query to get useful tokens that are used in the search process.

Algorithm 1: Tokenize the input query

Algorithm **tokenize**(query)
// string[]: query – input taken from user
{
1. Remove special characters from the query if present
2. Split query into array of words, using delimiter 'space'.
3. For each word in array, check the length
4. If length is greater than 'zero', consider the word.
5. Return array of words.

}

Tokenize is an algorithm, which splits the input query into an array of words. It ignores whitespaces and special characters used in the query and return the array of words as an output. For example, the users query w*hat are the symptoms of leaf curl disease?* is tokenized into array of eight words as 1) what 2) are 3) the 4) symptoms 5) of 6) leaf 7) curl 8) disease.

Algorithm 2 is used to get useful tokens from the array of tokens.

Algorithm 2: Finds Useful Tokens

Algorithm **usefulTokens**(tokens[], requ2redTokens[])
// string[]: tokens – tokens extracted from **tokenize()**
// string[]: requiredTokens – List of useful tokens for identification.
{
1. For each word in requiredTokens do {
   i. For each item in tokens do {
      a. If(checkSimilarity(item, word) >= 0.65) { consider item }}
   }
2. Return considered items;

}

From the useful tokens which are obtained from the Algorithm 2, key words are created which are used for searching the answers in the ontology. Algorithm 3 is used for this purpose.

Algorithm 3: Create Keywords from the useful tokens
Algorithm **createKeywords**(usefulTokens[])
//string[]: usefulTokens – tokens extracted from **usefulTokenize()**
{
1. Create list of tempKeywords by createAllCombinations(usefulTokens)
2. For each item in tempKeywords do {
   i. keywords.Add(string.join('_',item)) }
3. Return keywords

}

Since, algorithm 3 creates a large number of key words for the given query so it is required to get the key words that best describe the users' context. Algorithm 4 serves this purpose.

Algorithm 4: Identify Required Keywords
Algorithm **createAllCombinations**(usefulTokens[])
//string[]: usefulTokens – tokens extracted from **usefulTokenize()**
{
1. Return generateCombinations(usefulTokens,0, emptyList)

}

Algorithm 5: genearateCombinations(usefulTokens[], position, prevList)
// string[]: usefulTokens – tokens extracted from **usefulTokenize()**
// integer: index – starting position for generating combinations.
// List: prevList – List of combinations extracted from previous call.
// List of allCombinations – global variable
{
1. For(i=position to usefulTokens.Length, i++) {
   i. Create new List for currentCombinations
   ii. currentCombinations.Add(prevList)
   iii. currentCombinations.add(usefulTokens.elementAt (i))
   iv. allCombinations.Add(currentCombinations)
   v. generateCombinations(usefulTokens,i+1, currentCombinations)
   }

}

After obtaining the required key words, the Levenshtein distance (LD) measure is used to compute the similarity between the key words and the concepts present in the ontology. The Levenshtein distance (LD) is a measure of the similarity between two strings, which we will refer to as the source string (s) and the target string (t). The distance is the number of deletions, insertions, or substitutions required to transform s into t. For example,

- If s is "test" and t is "test", then LD(s,t) = 0, because no transformations are needed. The strings are already identical.
- If s is "test" and t is "tent", then LD(s,t) = 1, because one substitution (change "s" to "n") is sufficient to transform s into t.

The greater the Levenshtein distance, the more different the strings are. The smaller the Levenshtein distance, the more similar the strings are. Based on the similarity, the matched results are then returned to the user.

The greater the Levenshtein distance, the more different the strings are. The smaller the Levenshtein distance, the more similar the strings are. Based on the similarity, the matched results are then returned to the user.

3.2.2 Query Engine

This is the core of the system which handles farmer query. Farmer query can be any of the 3 types: ontology-based query, database query or ontology & database query. Currently we are handling ontology-based query. This is taken care by designing an algorithm which returns a path that best matches cluster of resources selected by user query.

3.2.3 Database

This is the repository which holds complete information about the crop in the form of ontology. This ontology is stored as a graph in database. Also, past records of the farmers are present as relational data which give us the farming practices done by them. Based on these generic rules observations are made which help in validating the ontology. Advice is generated by the system based on the current farming which helps him to take precautionary measures in his practice.

4. EXPERIMENTAL RESULTS

The Cotton Crop Ontology (CCO) was constructed from scratch in consultation with domain-experts. This ontology covers the domain of cotton production from cultivation to harvesting. Relevant knowledge related with cotton production was analyzed. The Cotton Crop ontology contains 162 concepts and 496 instances and 66 relations and provides an organizational framework that allows reasoning about cotton crop.

The classes describe the concepts of cotton crop domain like Crop, Soil, Activity, Pest, and Disease etc. The relationships are used to connect the classes. For example, the relationship *affected_by* connect the classes Crop and Disease which can be interpreted as **Crop *affected_by* Disease.**

The Agro-advisory system has been implemented using .Net framework as the front end and Oracle database as the back end. The cotton crop ontology is stored as triples in the oracle database. A flexible user-interface is provided for entering the farmer queries (Figure 2). When the user enters the query through the user interface, the query is first converted into meaningful keywords. The keywords are then compared against the ontology present in the database and the matched results are returned as answers to the query. For example, Figure 2 shows that the user is querying for ‘What are the boolwarms and types of boolwarms?’ The result of this query is displayed in Figure 3.

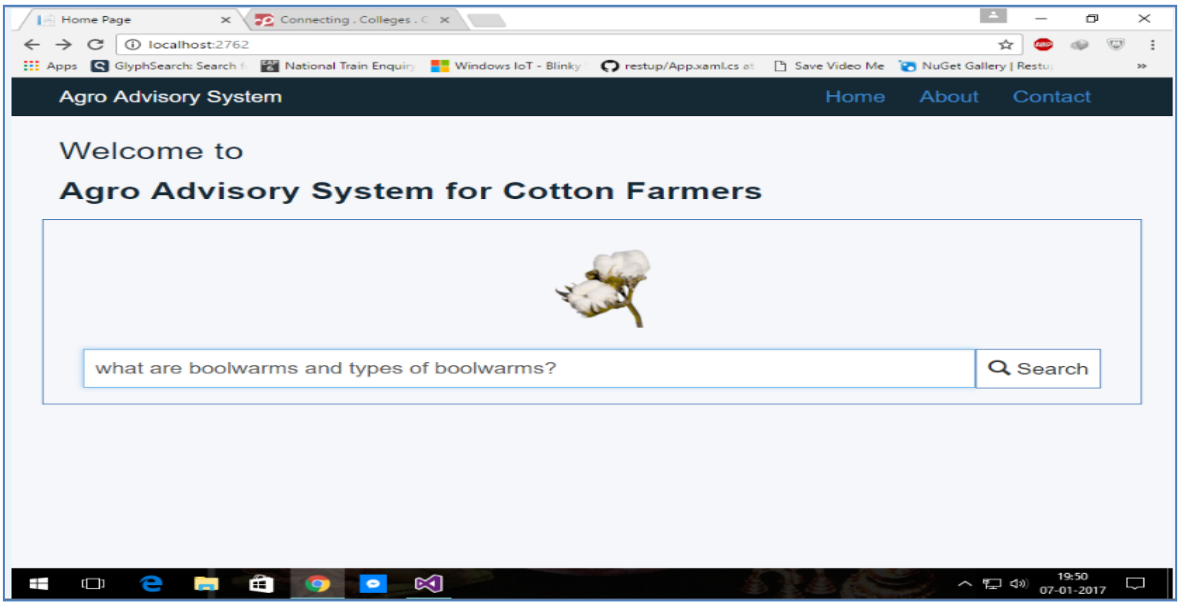

Fig 2: User interface for entering the query

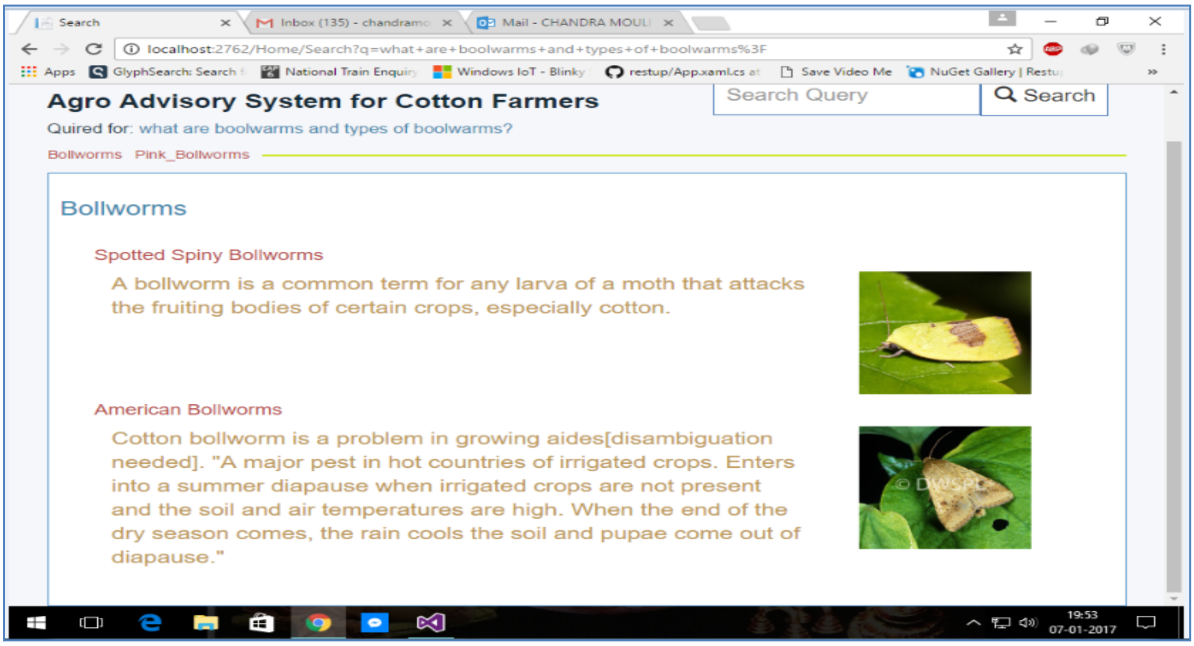

Fig 3: Result of the query ‘What are the bollworms and types of bollworms?

Figure 4 shows the result of the query “What are the symptoms of cotton leaf curl disease?” and Figure 5 shows the result of the query “How to control leaf curl disease?”

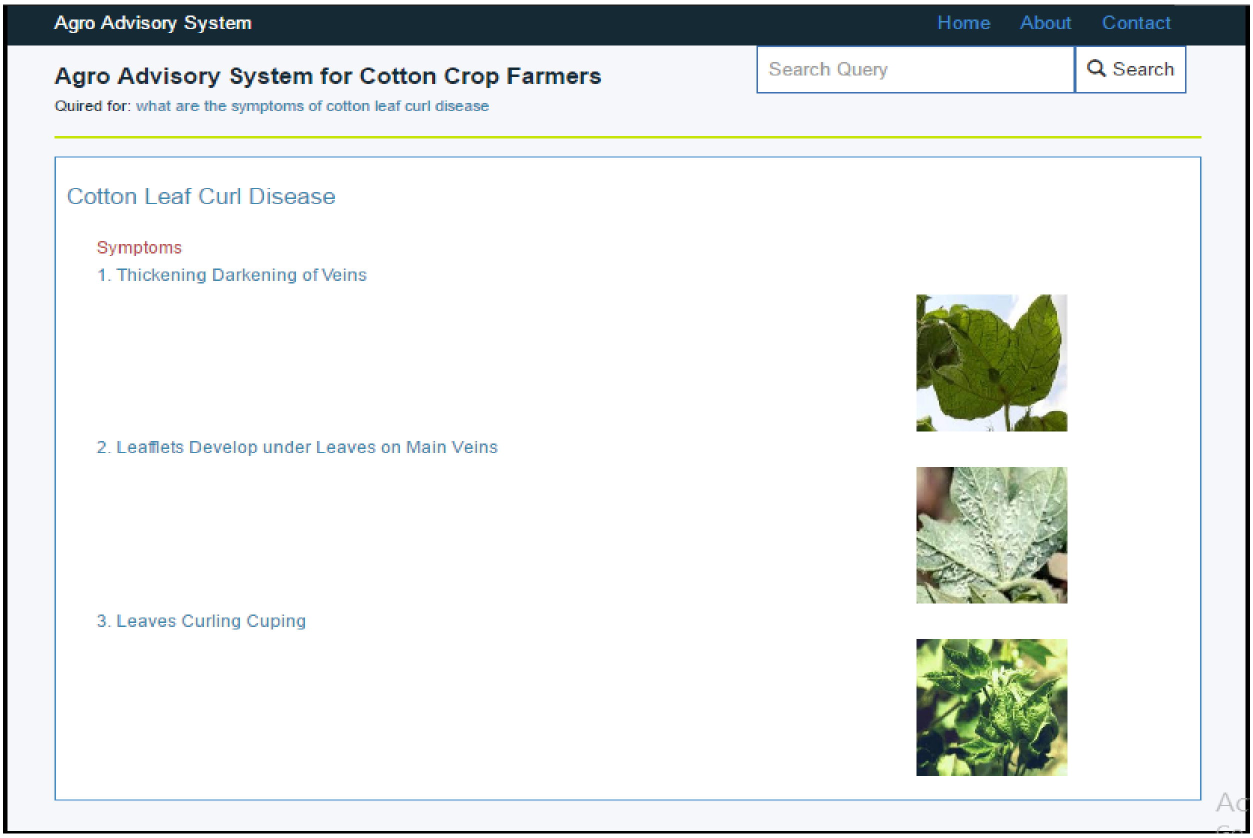

Fig 4: Result of the query “What are the symptoms of cotton leaf curl disease?”

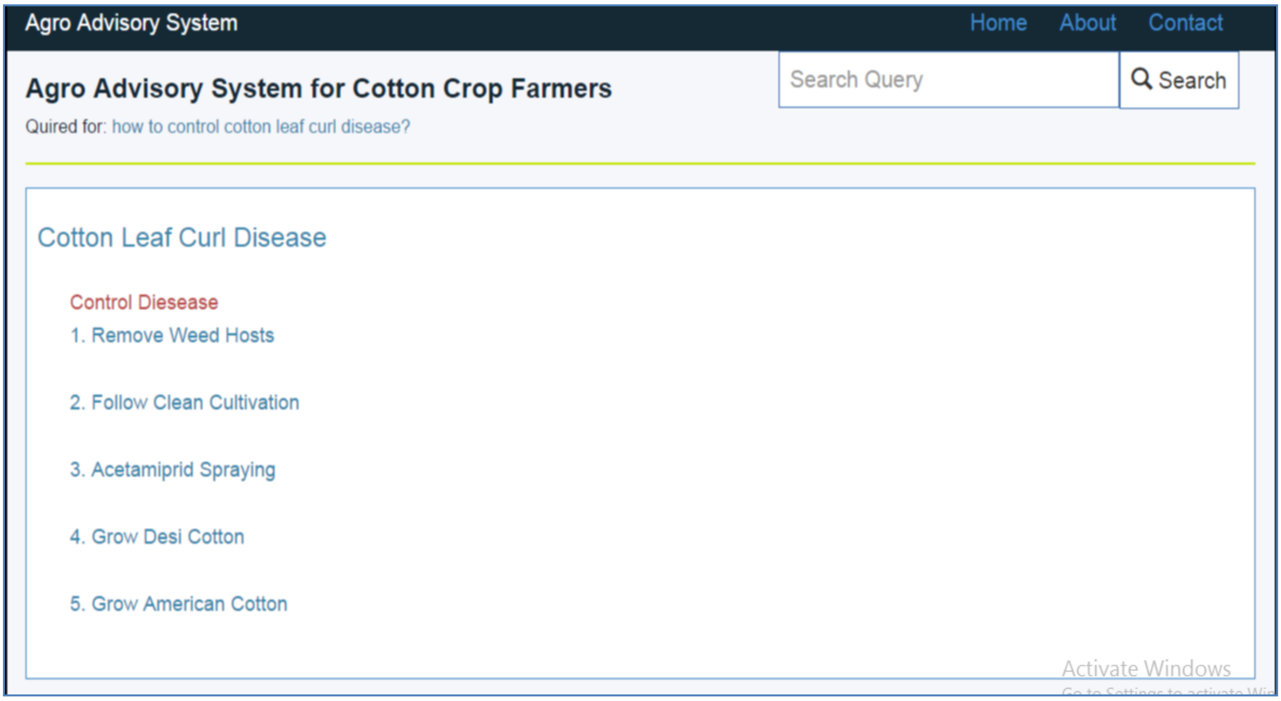

Fig 5: Result of the query “How to control leaf curl disease?”

5. CONCLUSION

Farmers encounter many questions depending on climate, diseases on crops, pests, preventions, timelines of various activities etc. To handle these queries, the old procedure was to collect the questions from farmers and record these questions, then search for the answers to these questions manually from documents provided by the agricultural experts. Due to lack of experts and their expertise, this process not only inefficient but also more time consuming. Hence, to provide meaningful answers to farmer queries more effectively, a knowledge based system called 'Agro-Advisory System' was developed. As part of this research work, cotton crop ontology was developed that acts as the knowledge base.

In order to provide useful answers to farmer queries, Agro-Advisory System uses Levenshtein distance to compute the similarity between the farmer query and the concepts present in the ontology. Based on this similarity, the answers are retrieved from the ontology and then presented to the user. It was observed that the similarity above 65% has produced the better results.

6. ACKNOWLDGEMENTS

I would like to express deep sense of gratitude and sincere thanks to UGC (University Grants Commission) of India for providing financial assistance to carry out this research work.

7. REFERENCES

[1] James, C (2008) Global status of commercialized biotech/GM crops: 2008, ISAAA Brief No. 39, ISAAA, Ithaca, NY

[2] ICAR (2006) Handbook of Agriculture. Published by ICAR, New Delhi. pp 1018-1036.

[3] V. Kumar, V. Dave, R. Bhadauriya and S. Chaudhary, "Krishimantra: Agricultural Recommendation System", in proceedings of the 3rd ACM Symposium on Computing for Development, Article No. 5, ISBN: 9781450318563, DOI:O.1145/2442882.2442933, 2013.

[4] Knowledge model for cotton, by : Neha Arora, Sonali Sahni ,Technical report, IIT Bombay, 2011.

[5] Approved Package of Practices for Cotton: Punjab State, Cotton Database, Available: http://www.cicr.org.in http://www.kapasindia.com, http://www.ikisan.com.

[6] G. Bhalotia, C. Nakhey, A. Hulgeri, S. Chakrabarti, and S. Sudarshanz. Keyword Searching and Browsing in Databases using BANKS. Proceedings of International Conference on Data Engineering, *2002.*

[7] Margherita Sini (FAO), Vimlesh Yadav (IITK), Contributors: Claudio Baldassarre (FAO),Revisers: Jeetendra Singh (IITK), TV Prabhakar (IITK),―Building Knowledge Models for Agropedia Indica v 1.0, Requirements, Guidelines, Suggestions.

[8] Amit Tripathi,Vimlesh Kumar Yadav and T.V. Prabhakar, IIT-Kanpur―Agropedia-An ICT TOOL FOR EXTENSION SERVICES IN INDIAN AGRICULTURE, Available: http://agropedia.iitk.ac.in/

[9] Vo Thi Ngoc Chau.Agroadvisory system, 2010.

[10] Aree Thunkijjanukij, ―Ontology Development for Agricultural Research Knowledge Management:A Case Study for Thai Rice.

[11] Gruber, T. 1993. A translation approach to portable ontology specifications. Knowledge Acquisition l5(2): 199-120.

[12] Owl web ontology language reference. http://www.w3.org/TR/owl-ref/.

[13] Protégé Tutorial, available: http://www.coode.org/resources/tutorials/ProtegeOWLTutorial-p4.0.pdf

[14] Natalya F. Noy and Deborah L. McGuinness, ―Ontology Development 101: A Guide for Creating Your First Ontology.

[15] Toby Segaran, Colin Evans, and Jamie Taylor, ―Programming the Semantic Web

[16] Oreilly Publications, ―Practical RDF

[17] Sergiy Nikitin, Vagan Terziyan, Yaroslav Tsaruk, Andriy Zharko, ―Querying Dynamic and Context-Sensitive Metadata in Semantic Web.