Overview of Agriculture Domain Ontologies

Prachi Dalvi, Varsha Mandave, Madhu Gothkhindi, Ankita Patil, S. Kadam, S. S. Pawar
Department of Computer Engineering D Y Patil College of Engineering, Akurdi, Pune Savitribai Phule Pune University

Abstract— Ontology is defined as shared specification of conceptual vocabulary used for formulating knowledge-level theories about a domain of discourse. Ontology has many applications. Ontology extraction is a process in which important concepts related to a domain are extracted and relationships between them is formed. India is an agriculture based economy. Majority of Indian population consists of farmers, but technologies are sparsely used for the aid of farmers. Ontology based modeling of agricultural knowledge can change this scenario. The farmers can understand it easily in their native language. This paper presents a review on various existing agriculture ontologies along with some of Natural Language Processing (NLP) techniques and models.

Index Terms— Ontology Modeling, Agriculture, NLP, Ontology Extraction, Domain Ontology.

I. INTRODUCTION

In India, over 70 percent of population has agriculture as a mean of livelihood. The agriculture domain is very vast. Large number of data has been written in books and till today lots of electronic data is available. But if we are searching for a specific data, the data becomes too large to read. India being a diverse country and language changes after every 20 kilometers it becomes difficult to communicate. And as majority of Indian farmers are not educated, it becomes difficult for them to handle English language. So it is necessary to have a system which will have farmers to gain knowledge in their native language.

Agriculture is the backbone of Indian economy. In most part of India traditional method[1] of farming is used. The traditional methods are having drawbacks like less yield of crops and diseases. Farmers need to learn new techniques like organic farming.

The most widely quoted definition of “ontology” was given by Tom Gruber in 1993, who defines ontology as (Gruber, 1993) [2]: “An explicit specification of a conceptualization.” Ontologies have proved their usefulness in different applications scenarios, such as natural language processing, semantic web, intelligent information integration, knowledge-based systems, digital libraries. Ontologies are developed to separate domain knowledge from operational knowledge. Reuse of domain knowledge and operational knowledge is possible using ontologies.

Ontologies in specific domains such as Health care has been developed on a large scale. In health care, the information regarding medical treatments are consistent worldwide. But in agricultural field, the information changes according to environmental conditions and geographic locations. Farmers in India belong from different states and different states have different languages. Language becomes a barrier as the farmers are unaware about other languages. Due to this, India is lacking in agricultural domain ontology and hence, it is a challenge to create an agriculture based ontology in native language. Successful ontologies like Gene Ontology and Disease Ontology are available in bio-medical domain.

Natural Language Processing (NLP) is a very active area of research and development in Computer Science. NLP applications are machine translation and automatic speech recognition. Natural language processing techniques are used to process input which is in the form of natural language i.e human understandable. The idea behind the natural language processing is to interpret input as whole by combining the structure and meaning of words that is interpretations are obtained by matching patterns of words against the input utterance.

The remainder of this paper is structured as follows:

In Section II existing agriculture ontologies are discussed in text and tabular format. In Section III, models and techniques in NLP(Natural language processing) are also discussed.

II. EXISTING AGRICULTURE ONTOLOGIES

Agriculture is considered to be a very important sector in creation of raw food items. For economic growth of country agro based industries play a vital role. It is important that all the data regarding agriculture domain should be well organized and properly arranged, so that the farmer can easily retrieve the inter-related data. Ontology extraction techniques can be used for extracting relevant information.
A. Integrated Agriculture Information Framework (IAIF)[3]

Integrated Agriculture Information Framework (IAIF) is one of the useful solution for ontology extraction. This IAIF technique makes knowledge extraction possible from various domain related repositories. Main functions of IAIF technique are combine, merge and aggregate the data in existing knowledge repositories. The three sub-ontologies included in IAIF agriculture ontology are Domain ontology, Resource Ontology, Linking Ontology. This is triple based approach used for ontology creation, in which two concepts are linked together based on specific property to generate a graph as directed.

B. Scalable Service Oriented Agriculture Ontology for Precision Farming (ONTAgri)[3][4]

Scalable Service Oriented Agriculture Ontology for Precision Farming (ONTAgri) is proposed to used in agriculture domain and this domain consist of several farming practices such as irrigation fertilization and pesticides spraying. ONTAgri provides sensor technology support for agriculture domain, which helps in acquisition of real time values. It composes of two major parts which are system ontology and domain ontology. Domain Ontology is divided into two parts i.e core and services. To support agriculture requirements number related domains ontology can be linked which introduces scalability features. Utilization of same components is also possible. These are separated from each other for reduction of dependency and this technique makes it easier when a new service that utilizes the same system component is added. New agricultural services are introducing easily without changing the structure in this ontology.

C. AGROVOC[5][6]

AGROVOC is a structured thesaurus created in 1980, by FAO and European Communities. It covers the fields of food, agriculture, forestry, fisheries, etc. It is a multilingual thesaurus. It covers upto 23 languages that include Arabic, Chinese, Czech, English, French, German, Hindi, Hungarian, Italian, Japanese, Korean, Lao, Malay, Persian, Polish, Portuguese, Russian, Slovak, Spanish, Telugu, Thai, Turkish and Ukrainian. The main purpose behind this AGROVOC is to improve the efficiency and effectiveness of the information searching process.

In thesaurus the relationships between concepts are not explicitly defined. When the concepts are defined explicitly they become more meaningful. So the AGROVAC thesaurus was converted to concept schema.

The AGROVAC concept schema is structured concept based system. It is represented in three levels and those are concepts, terms and term variants. It consists of concepts with their lexical representation, specific concepts and also relationships between them. AGROVAC has two models one is based on OWL(Web Ontology Language) and other is based on SKOS(Simple Knowledge Organization System).

D. Agricultural Ontology Service (AOS)[3]

AOS is designed for utilization of AGROVOC encyclopedia at its core. It also serves as a common set of core terms and relationships as well as the richer relationship which can be shared among knowledge organization system. The main purpose of AOS is to achieve interoperability among different agriculture systems. AGROVOC thesaurus only provides a conceptual hierarchy of terms which are not interlinked to provide relationships among the terms. To overcome this, AOS has been developed to provide the concepts between the terms and the specifications of relationships among them. Agricultural resources of multiple languages are combined together using AOS. It provides a framework for sharing common ideas and terms within agricultural community.

E. World Agriculture Information Center (WAICENT) [7]

WAICENT's is a multilingual knowledge management system. It is powered by FAO. With the help of WAICENT, FAO knowledge of agriculture is available to users around the world through internet. WAICENT's includes organizing and linking information in order to facilitate user access. WAICENT is a multilingual system it is available in five different languages and they are English, Arabic, Spanish, Chinese and French. Users can access WAICENT through the WAICENT Information Finder. Keyword search is available in WAICENT.

Table 1. Review of existing ontologies

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontology</td>
<td>Domain, Resource and Linking Ontologies</td>
<td>System and Domain Ontologies</td>
<td>AGROVOC Thesaurus</td>
<td>AGROVOC Thesaurus</td>
<td>System Ontology</td>
</tr>
<tr>
<td>Knowledge Representation</td>
<td>RDF-OWL</td>
<td>OWL/SWRL</td>
<td>RDF/SKOS-XL</td>
<td>RDF/SKOS-XL</td>
<td>RDF/XML</td>
</tr>
<tr>
<td>Ontology</td>
<td>Files</td>
<td>Files</td>
<td>Files and</td>
<td>Files and DBMS</td>
<td>Files and DBMS</td>
</tr>
</tbody>
</table>
III. NATURAL LANGUAGE PROCESSING

Natural language processing which is also known as “language technology ” or “language engineering”, (Mannig, Schut and press 1999) [8] is concerned with all those theories and hypotheses that deal with automatically processing textual information based on human knowledge of language, computational linguistics and speech language processing etc.

Language processing can be summarized into the following six criteria (Jurafsky & Martin, 2008) : [9][10]

1. Phonetics and Phonology: the study of sounds (Sibawayhi, the Arabic grammarian of the 8th century was one of the first phonologist to study the vibration of sounds and words correlations (Edzard, 2000)).


3. Syntax: the study of the relationships between words.


6. Discourse: the study of linguistics as a complete unit.

A. Natural Language Processing Models

1) Rule based model [11]:

Rule based models help you to write the rules explicitly. A rule based system consists of a set of rules, a working memory for storing states, a schema for matching the rules and a conflict resolution schema if more than one rule is applicable. Rule based models are very expressive. Rule based models are easy to write and debug than decision trees. But they require more memory. Sometimes debugging becomes difficult.

2) Conditional Random Fields(CRFs) model [11]:

Conditional random fields (CRFs) are a probabilistic framework for labeling and segmenting structured data. The idea is to define a conditional probability distribution over label sequences given at a particular observation sequence. CRF is an discriminative model. It does not assume the features that are independent.

B. Natural Language Processing Techniques for Ontology Extraction

Some of the natural language processing techniques are as mentioned below:-

1) Part-of-Speech Tagging [12][8]:

Natural Language Processing is used to understand human language. Part-of-Speech tagging is one of the technique. The words of a language are grouped into classes. The words having similar syntactic behaviour are categorized into class. The classes are the parts of speech i.e. noun, verb, adjective. Noun consists of places, people. Verbs are used to express nouns. Properties of nouns are described by adjectives. A term named POS tagger is used which tags each an every word in a sentence about its part of speech. For example, it will indicate “Ram” as a noun because Ram is a noun grammatically.

2) Pattern Matching:

The idea behind the natural language processing is to interpret input as whole by combining the structure and meaning of words that is interpretations are obtained by matching patterns of words against the input utterance. There are large number of patterns required even for restricted domain in deep level of analysis in pattern matching.

3) Semantic grammars:

In semantic grammar the categories used are defined semantically. It also contains some mixture of syntactic
elements. Semantic grammars are used typically in language processing for limited domain contexts.

4) Word Stress [12]:

As the name indicates this method gives stress to each an every word in a particular given input text or a sentence. It weighs the words accordingly as it will be used in the sentence. It is a very important natural language processing technique used in speech recognition.

5) Text Normalization [12]:

Text Normalization is used for synthesizing the required input text properly. Sometimes few important aspects are generally taken for granted and to contemplate that text normalization is use. This method is used for dealing with confusions occurred regarding abbreviations, punctuations, differences in places where capital letters should be used. It is a technique which perfectly splits a sentence at punctuations and white spaces.

6) Syntactically driven parsing:

The way that words can fit together to form higher level units such as phrases, clauses and sentences is called syntax. Syntax analysis are obtained by application of grammar that determines what sentences are legal in the language that is being parsed.

IV. NLP FOR AGRICULTURE ONTOLOGY

To create an ontology raw text will be required and to process it nlp techniques are used. Techniques like POS Tagging, Pattern Matching, Text Normalization, Semantic Grammars etc can be used to process the raw text while developing ontology. Processing models like Rule based and CRF model can be used to set rules and to segment structured data respectively.

V. CONCLUSION

This paper reviews existing agriculture ontologies along with their features, knowledge representation and efforts for modelling. We based our study on the tasks expected. NLP models and techniques are also reviewed.

ACKNOWLEDGMENT

We express our deepest gratitude to the college authorities for technical guidance and infrastructure. We are very thankful to the researchers and reviewers for their contributions because of which we could complete this work. Further we extend our deepest the online community for their support without which the work would not have been possible.

REFERENCES


