An Approach for Knowledge Extraction Using Ontology Construction and Machine Learning Techniques

Dhanasekaran K *dhana_mec5684@yahoo.co.in*
Research scholar/Computer Science and Engineering
Info Institute of Engineering
Anna University of Technology
Coimbatore-641 107, India

Rajeswari R *rreee@gct.ac.in*
Assistant professor/Department of Electronics and Instrumentation Engineering
Government College of Technology
Anna University of Technology
Coimbatore, India

Abstract

In recent research, Ontology construction plays a major role for transforming raw texts into useful knowledge. The proposed method supports efficient retrieval with the help of ontology and applies combined techniques to train the data before taking into testing process. The proposed approach used the phrase-pairs to extract useful knowledge and utilized data mining techniques and neural network approach to express the knowledge well and also it improves the search speed and accuracy of information retrieval. This method avoids noise generation by analyzing the relevancy of tags to the retrieval process and shows somewhat better recall value compared to other methods. In this approach an optimized reasoner applied to reduce complexity in the key inference problem. The formulated ontology can help clearly expressing its meaning for various concepts and relations. Due to the increasing size of ontology repository, the matching process may take more time. To avoid this, this method forms a hierarchical structure with semantic interpretation of data. The system designed to eliminate domain-dependency with the help of dynamic labeling scheme using ontology as a base. In this paper, our proposed models were presented with ontology description using Ontology Web Language (OWL).

Keywords: Back Propagation, Domain Ontology, Knowledge Extraction, Agricultural Environment.

1. INTRODUCTION

Agricultural sector plays an important role in increasing economy of any nation. With this in mind, researchers nowadays, focusing on solving problems that causes either damages or disease with respect to plant production, rice production, etc., and they are trying to solve problem in respect of the issue for improving the productivity and increasing an economy and reducing the cost that is to be invested in production [1].

This method allows us to identify various qualifiers for retrieving information which are relevant to the user query. In survey there is a evidence that there are no efficient construction of agriculture ontology method that is available to address various issues which degrades growth in multiple dimension [1][2].In the proposed method, optimal tagging will be performed based on the weights assigned to the items in the list during analysis stage. Initially the data is trained with the help of back propagation method and learning is performed on those data to produce valuable data for the testing process. Finally, the data which are stored at various nodes at various levels in the hierarchy in some order will be processed and retrieved with improved recall value.
The mapping is carried out by checking for correspondence between input vectors and the expected output vectors. In addition, the similarity and differences in the concepts and relations

are identified to eliminate redundancies and ambiguity that are present in the knowledge repository.

In this method, various guidelines and rules are formulated by associating relevant features which are extracted from the hierarchical structure. This will support presenting some advisable ideas to the farmer and suggests some methods for preventing damages or controlling disease in cultivation of varieties, planting, etc., the new method improve efficiency in information extraction process with the help of a generated ontology which adopts some rules and criteria recommended by the researchers [4].

Ontology is mainly constructed as a formal specification of various concepts and relations between properties of those concepts. Ontology is defined as partial specification of conceptual vocabulary used for formulating knowledge-level theories about a domain of discourse. Ontology is applied in domains like natural disaster management system, medicine, military intelligence, cooking, enterprise, jobs, agriculture, Wikipedia, automobiles and so on. Ontology is also expressed as a formal representation of knowledge by a set of concepts within a domain and the relationship between these concepts. Ontology consists of four main components to represent a domain. They are:
i. Concept represents a set of entities within a domain.
ii. Relation specifies the interaction among concepts
iii. Instance indicates the concrete example of concepts within the domain
iv. Axioms denote a statement that is always true.

Let us take an example of a wine ontology and look at its components. The concepts of the wine ontology are, “Winery, Wine, Wine descriptor, Wine color, etc”. The relationships are given as Winery *produces wine*, wine *has* wine descriptor. The instances of wine color can be red, rose and white. The axiom in this example is ‘a winery must produce at least one type of wine’.
Ontologies can be constructed using three different approaches; single ontology approach, multiple ontology approach and hybrid ontology approach [22].

The single ontology approach is the simplest of all and it uses single global ontology for all information sources, which shares the vocabulary and the terminology to specify the semantics. The limitation of this approach is that it does not provide a perfect solution for information integration. This limitation has paved way for multiple ontology approach, where each information source is described by its own ontology thus each source will have its own local ontology. The main drawback of this approach is the construction of individual ontology. The hybrid ontology approach is the combination of single and multiple approaches [22].

In general, ontology construction could be done in three ways [23]:

- Manual: Ontology is constructed manually.
- Semi automatic: Human intervention is needed during ontology process.
- Fully automatic: The system takes care of the complete construction.

Ontology construction involves six basic steps.

1. Ontology Scope
2. Ontology capture
3. Ontology encoding
4. Ontology integration
5. Ontology evaluation
6. Ontology documentation.

It gives semantic expression for each and every term in the conceptual framework by means of combining relevant items together with the help of a Natural Language Processing tool. In my

study part, I could see that some methods are developed to construct ontology that refers to business model and data model for petroleum exploration and production domain feature, etc., [2][5][6].In ontology research, no one focused on the related research of agricultural domain that analyzes how planting method, use of ingredients, irrigation method affects land quality, productivity, as a result economy.

The Extractor in this method will retrieve the factor that characterizes planting, cultivation of varieties, various methods in irrigation, etc., It uses various concepts and relations for the core vocabularies of the considered field. With these, it makes it possible to regulate the extension of features and an ontology is used to describe them using XML as a base and OWL ontology language as a description language.

2. LITERATURE REVIEW

In the year 2009,Amal Zouaq and Roger NKambou have published a paper on "Evaluating the generation of domain Ontologies in the Knowledge Puzzle Porject".The author described the procedure to extract concept maps from texts that are followed by TEXCOMON,Knowledge Puzzle Ontology Learning tool. In this paper, they are evaluated ontology in three dimensions: structural, semantic and comparative. In structural evaluations, ontology is considered as graph based on a set of metrics. Semantic evaluation is carried out using human expert judgement.Finally comparative evaluation is done by comparing the output of current tools and new tools. This task has used the same set of documents for all cases.

They compared the ontological output in terms of concepts, attributes, hierarchical and non-taxonomic relationships. The method produced more interesting concepts and relationships but failed to avoid a lot of noise generation by lexico-syntactic patterns and their methods. They suggested developing method for improving the patterns.

Moreover, the OWL Java API of their project improved in terms of processing time. This paper consists of a future direction towards automating ontology evaluations in order to solve a number of problems such as., Ontology-learning, population, mediation and matching.

In the year 2004, Marta Sabou has published a paper on "Extracting Ontologies from Software Documentation: A Semi-automatic method and its evaluation". In his approach, he used software APIs to build domain ontology by extracting types of method functionalities. In that method, a small corpus is used for applying statistical techniques. The author has described that there is a need to enhance the corpus and to develop a better extraction method that suits the small corpora. This method is encouraging towards building an Ontology extraction method from software APIs.

In the year 2010, B.Saleena, Dr.S.K. Srivatsa has published a paper on "A Novel Approach to develop a self-organized Domain Specific Search mechanism for Knowledge Acquisition using Ontology". The authors have created a search method for semantic web in order to design a self-organized system to retrieve information about a particular topic based on user interest in learning.

For this they created a knowledge library for DBMS domain using Ontology and Knowledge management technologies. And then they followed a strategy to group the relevant information for the user in a single search. The search is implemented based on keyword which is a time-consuming process. The system has been developed in JAVA 2 API with OWL API for semantic web. It retrieves the inter-related contents, prerequisites and further readings needed to understand the topic as per user's interest. With the help of this, an e-learning framework is developed using Ontology based knowledge retrieval.

This paper included future work to enhance the system for various domains and to add a large set of functionalities to the UI screen to improve the user-friendliness. Also it has been suggested to develop a method for automatic extraction of information.
In the year 2009, Song Jun-feng, Zhang Wei-ming, Xiao Wei-dong, Xu Zhen-ning has published their work "Study on Construction and Integration of Military Domain Ontology, Situation Ontology and Military Rule Ontology for Network Centric Warfare". The author have discussed that there is a need for knowledge infrastructure in network centric warfare in order to transform information into knowledge.

In this paper, they proposed approach to construct all three kinds of ontology mentioned above. They also addressed the integration approach using all these approaches. Then they have constructed scenario based knowledge infrastructure fragment using proposed approaches and techniques. They said that current research works are few in this respect, in future study; they are going to study what kinds of other component ontology are needed for the knowledge infrastructure. They also planned to implement experimentation and at present, they would like to use protégé basic tool.

The conceptual graph of ontology language (OML) lacks precise semantics. OWL is a new synthesis of research on ontology language. The expressiveness of all the languages are very limited and key inference problem has most complexity. So, there is a need for optimized reasoners.

In the same year, Zhang Rui-ling, XU Hong-Sheng have published a paper on "Using Bayesian Network and Neural Network Constructing Domain Ontology". In this paper, they have addressed that the current ontology construction methods have limitations. They are: 1) Requirement for human labor 2) Domain restrictions. To avoid these problems, they developed an approach to construct ontology based on a novel method which contains Projective Adaptive Resonance Theory (PART) neural network and Bayesian Network Probability theorem.

Their system could acquire key terms automatically. Finally it reasons out the complete terms in the classification framework in order to construct domain ontology. The ontology is stored using a Resource Description Framework (RDF).The Semantic Web can be deployed based on the rapid and efficient construction of the ontology. Some of the features of this work are: the PART architecture is included to overcome the lack of flexibility in clustering, and in the web page analysis, WordNet deals with the lack of knowledge acquisition.

Finally they said that there is a need to improve the precision of term location. Due to the accumulation of the number of documents in the ontology repository, the similarity calculation takes more time. This is unavoidable. So, if we build an approach to form a hierarchy of clusters, it will solve the problem. The current methods can build only a partially automated classification of terms". This involves a time-consuming process and costly procedure.

In the same year, Yi Zhang, Li Tan, Jie Liu, ChangChang Yu has published a paper on "A Domain Ontology Construction method towards Healthy Housing". Due to the presence of various domains, there is no efficient framework for ontology construction. The authors have introduced an improved ontology construction method. In this method, they used graphic language to represent domain knowledge for research domain. The system evaluated and verified the correction of relationships and hierarchy for constructed ontology. The proposed method for Healthy Housing solves the problems such as lack of semantic expression and understanding of expert knowledge.

In future work, it is stated that the research can be done to utilize this Housing ontology in the Healthy Housing Intelligent Synthesized Evaluation System by mapping and matching domain ontology. Then it will be possible to realize the interchange between natural language of professional field and conceptual-level ontology language which can be understood by machines.

In the year 2010, Zhang Dan, hang Li, Jiang Hao have published their paper "Research on Semi-automatic Domain Ontology Construction". They have applied Data Mining method and word partitioning technique to construct semi-automatic domain ontology. At certain level, they could recognize the effectiveness and quality of the ontology. In this paper, they said that the semi-automatic approach still poses a problem because of the difficulties in constructing a common tool. The reasons are: choosing data source is manual, extracting compound words without considering the characteristics of language, and analyzing the grammatical components of sentence to conclude the relations among concepts. The authors said that the methods can be tried to address these issue.

3. ONTOLOGY CONSTRUCTION METHOD

Existing ontology methods are focused on acquiring concepts, properties and relations, and also focus on emphasizing the description of knowledge ontology. Although it can eliminate ambiguity based on user's judgment, it fails to apply analysis and expression evaluation for predicting accuracy of constructing ontology. This paper gives solution for this problem.

Steps in the process of Ontology construction are as follows:

Step1: Domain Analysis Phase
In this phase, determination of extensibility and reusability of domain ontology are taken place.

Step2: Ontology Analysis and Design Phase
In this phase, Agricultural domain ontology will be established by defining the hierarchy of concepts and relations between different activities and then translates the professional knowledge and raw data of the domain to commonly used information. This phase acquires semantic information about concepts, relations, actions, etc.

Step3: Ontology Representation
In this phase, Individual classes are created based on properties of the class and domain ontology is represented using XML and OWL ontology description language.

Step4: Ontology Evaluation
Here, the rule set is formed and inference mechanism is applied and then the accuracy and correspondence between input and output vectors are evaluated. Hence, it checks for the consistency of the constructed model. These steps are shown in Fig.1.

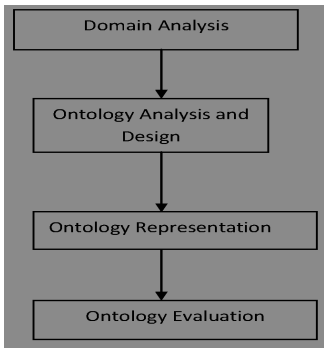

FIGURE 1 The generic process steps of Agricultural ontology construction

The proposed approach shown in Fig.2 include ten simple criteria: lawfulness (i.e. frequency of semantic errors), richness (how many of the semantic features available in the ontology are used by the processing element and data mining algorithm), interpretability (do the phrase pairs used in the ontology also appear in Phrase Net?), consistency (how many concepts in the ontology are involved in inconsistencies),clarity (do the phrases used in the ontology have many senses in Phrase Net?), comprehensiveness (number of concepts in the ontology, relative to the average

for the entire library of ontologies), accuracy (percentage of false statements in the ontology), relevance (number of statements that involve semantic features marked as useful or acceptable to the user/agent), authority (how many other ontologies use concepts from this ontology), history (how many accesses to this ontology have been made, relative to other ontologies in the library/repository)[24].

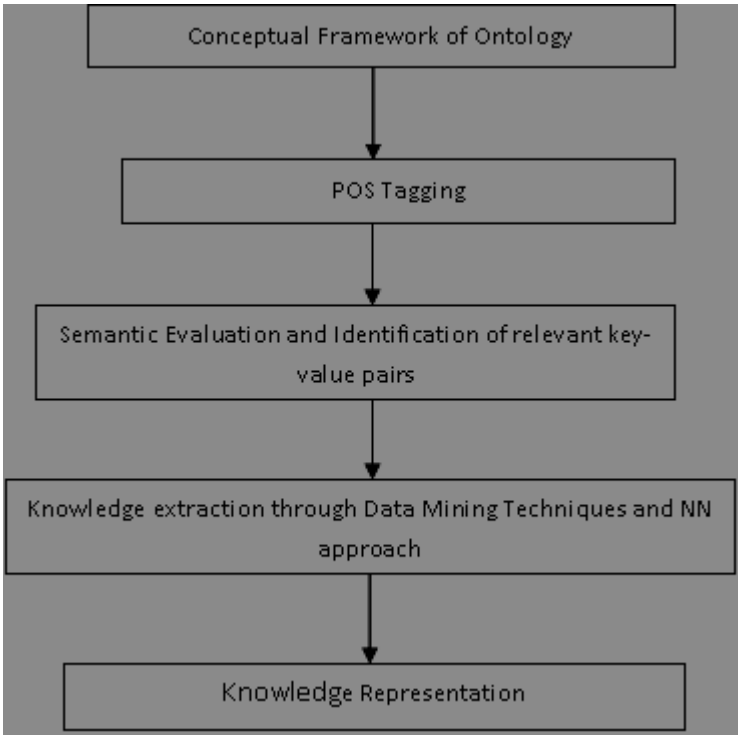

FIGURE 2 Proposed model of Knowledge Extraction

4. CONSTRUCTION OF AGRICULTURE ENVIRONMENT CONTROL AND MANAGEMENT ONTOLOGY

4.1 Domain Analysis

In this phase, grouping of the expert's opinion and the definition of the domain consists of four higher-level concepts such as., agricultural environment standard, climate analysis standard, watering service or irrigation standard, and land use environment standard.

At the top-level more general concepts which describe the common properties are defined. On the other hand, at the lower level, more specific properties of the standards are defined which can be used as subclass. Its instance in the domain ontology plays a major role for sending input data and receiving output signal in order to perform snalysis, processing or other kinds of tasks.

4.2 Ontology Analysis and Design

The subclasses of core concepts are again divided into many lower level concept which supports extraction of correct knowledge for the users by comparing useful patterns of the original information. The entire structure of the environment is shown in Fig.2.

This method gives importance for generalizing the concepts into more specific concepts. Thus the constructed ontology will not have redundancies and ambiguity by finding the right classes, subclasses, instances, and properties. It supports getting better recall value for the user query.

In Fig.3, the subclass environment with appropriate relationships is derived among various conceptual elements. The agricultural environment standard has relationship with concepts such as., climate analysis standard, this in turn, allows us to express the concepts such as., temperature analysis, precipitation control, monitoring service, information management.

4.3 Ontology Representations

After the analysis and design of ontology, it will be represented using 5-tuple.They are defined as follows:
Definition: DOR :=(SC, RC, DFE, LLC, PLLC)
Where,
SC: Super Class which acts as root of the hierarchy
RC: Relationships among Concepts
DFE: Data and Functional Elements
LLC: Lower Level Concepts which represents some knowledge at the detailed level.
PLLC: Properties of Lower Level Concepts which represents some derived and specific properties of its own.

1) Super Class (SC)

This represents concepts of ontology by showing some top-level concepts like watering service setting, land use analysis and so on. Also it shows some attributes like planting date, location, yield and so forth.

2) Relationships among Concepts (RC)

The basic relationships in Agricultural domain ontology are shown in table1.

Relation	Definition of relation
Kind-of	The inheritance of concepts
Part-of	The relation between mass and parts
Instance-of	The relation between instance and concept

TABLE 1: The basic relations of Agricultural Domain

For example, in Fig.3, Temperature and Precipitation have Part-Of relationship with Weather class.

3) LLC (Lower Level Concepts)

The lower level mostly has kind-of character with its higher-level. In Fig.3, the Weather control and Management environment has been defined with properties and relationships as follows:

```
Def Category Weather control and Management
{
Property: Weather planning
      : Type string
      : Comment “plan of weather control”
Property: Management planning
      : Type string
      : Comment “plan of information management”
}
```

4) PLLC (Properties of Lower Level Concepts)

This represents implementation of class by writing class with derived values and some additional values of its own.

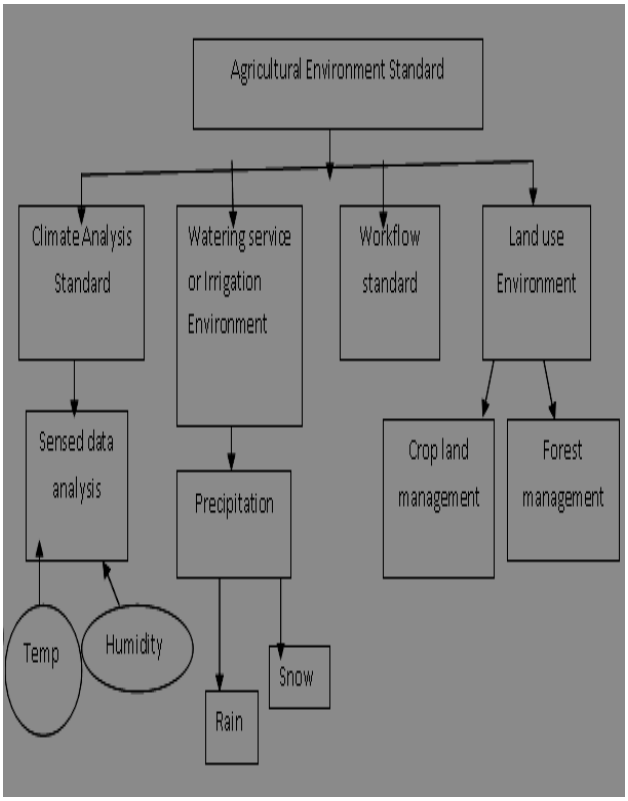

FIGURE 2: The structure of Agricultural environment standard

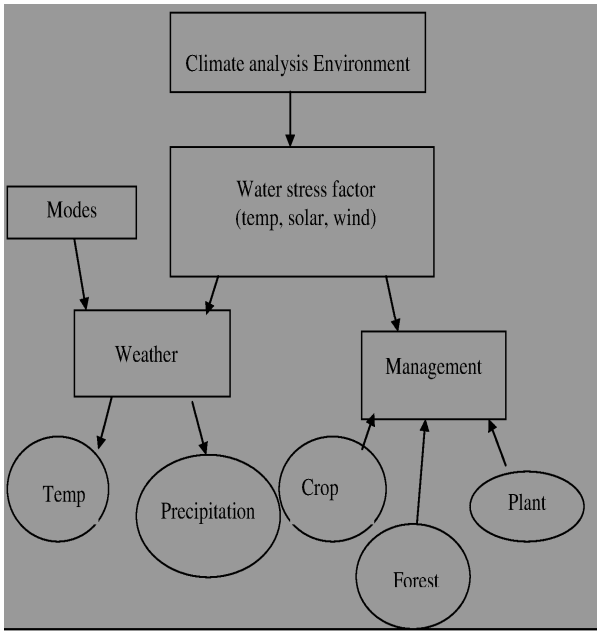

FIGURE 3: Relationship between weather and management

In Fig.4, considering Environmental factor, they can be classified as Nonbiotic factor determination and outcome planning.

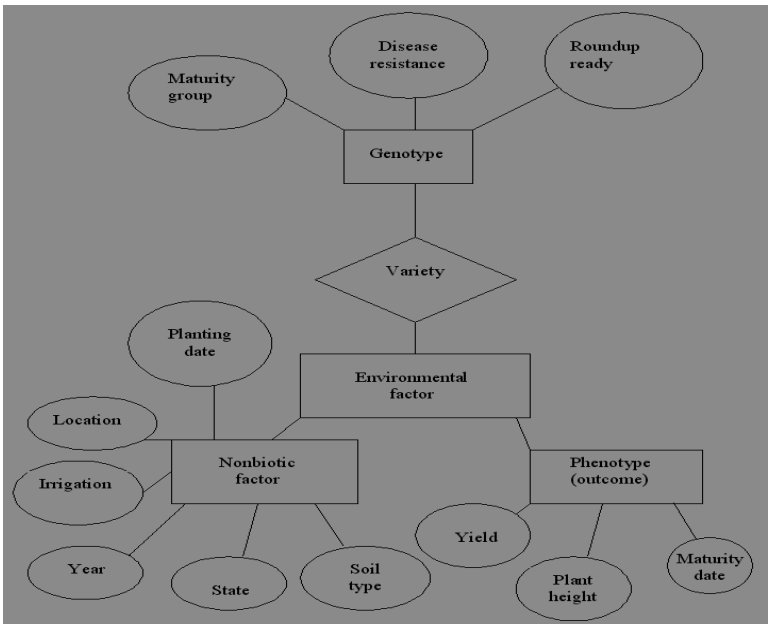

FIGURE 4: A model of agriculture agronomy

These factors may affect the environment and they are defined as follows:

Def Environment Analysis

{

Nonbiotic factor determination: Make decision for the parameter such as., planting date, location, irrigation, year, state, soil type, etc.,

Outcome generation plan: Make decision for generation of yield, maturity date, plant height.

}

In this paper, Protege tool used for establishing ontology by defining various attributes, activities, objects, and relationships. The land use environment is described as a subclass of Agricultural Environment Standard using OWL language as follows:

<? xml version=1.0"?>
<rdf: RDF
Xmlns:sqwrl=http://sqwrl.standard.edu/ontologies/built-ins/3.4/sqwrl.owl#
xmlns="file:/D:/MyFolder%18files/Protege3.4.1/Agricultural environment standard.owl#"
xml:base:"file:/D:/MyFolder%18files/Protege3.4.1/ Agricultural environment standard.owl">
<owl:Ontology rdf:about="file:/D:/MyFolder%18files/Protege3.4.1/ Agricultural environment standard.owl">
<owl:importsrdf:resource=http://sqwrl:standard.edu/ontologies/built-ins/3.4/sqwrl.owl/>
<owl:importsrdf:resource=http://sqwrl:standard.edu/ontologies/built-ins/3.3swrla.owl/>
</owl: Ontology>
<owl: Class rdf: ID="Sensed data analysis">
<rdfs: subClassOf>
<owl: Class rdf: ID="Climate analysis standard"/>
</rdfs: subClassOf>
</owl: Class>

5) Ontology Testing

The accurateness of constructed ontology is verified by comparing the MADRE method. The ontology is evaluated in two ways: 1) verification of relations 2) verification of hierarchy. Both of these verification compare the result that are obtained by inference mechanism of domain ontology in this paper and then if it satisfies the requirements then it will be added into the domain ontology library,otherwise,going back to the analysis phase, it finds new ontology and definition.

5. CONCLUSION

Ontology helps us to make knowledge acquisition and retrieval process in very easier manner. The existing methods and patterns for information retrieval do not allow us to accurately retrieve information so that various methods have been developed so far, to address issues in different domains. In this paper, we have discussed a new approach to produce somewhat better result for the selected problem in agricultural domain with the help of an efficient ontology construction, data mining and neural network approach. The method is developed to address disaster and disease control issues in the selected domain irrespective of tasks such as., planting, cultivation of varieties, irrigation, etc.It shows somewhat better performance compared to other methods and act as a new framework for agricultural domain. Also this work enables the user to utilize an ontology query method by using pair-wise tagging. The optimal pair-phrase is generated using Natural Language Processing tool focusing on identified key issues. The formulated method can clearly express its meaning for various concepts and relations. The limitations in the keyword or term based extraction is eliminated by implementing query transformation technique which generates intra-query value pair. This gives suggestions and lists previous histories for the user query by reducing the computation time and cost and producing the improved recall value. Although this method produces encouraging results with an agricultural data set, future work will involve libraries for testing the performance of the proposed method on other data set. It should also be possible to generalize our approach to learning simulataneously several subtrees in the ontology tree.

6. REFERENCES

[1] A.Thunkijjanukij, A.Kawtrakul, S.Panichsakpatana , U.Veesommai.” Lesson learned for ontology construction with Thai rice case study”. “in press”.*World Conference on agricultural information and IT*,2008,pp.495-502.

[2] W.Bi-long , H.Li.”Method of building petroleum exploration and production domain ontology”.“in press” *Computer Engineering and Applications*, 2009.pp.1-3.

[3] Z.Rui-ling, X.Hong-sheng.”Using Bayesian network and neural network constructing domain ontology”.“in press” *World Congress on Computer Science and Information Engineering*, 2009, IEEE 2008, pp.116-231.

[4] Guarino.N. “Formal ontology in information systems”.“in press”, in *Proc.IOS*,1998.

[5] Elena P,Sapozhnikova.” Multi-label classification with art neural networks”. “in press” *Second International Workshop on Knowledge Discovery and Data Mining*, 2009 IEEE, pp.144-147.

[6] X.LI, l.ZHAO, l.WU.”A feature extraction method using base phrase and keyword in Chinese text”. “in press” in *Proc. 3rd International Conference on Intelligent System and Knowledge Engineering*,pp.680-685, 2008 IEEE.

[7] S.Zhao R.Grishman,”Extracting Relations with Integrated Information Using Kernel Methods”, “in press” *in Proc.43rd Annual Meeting of the ACL*, pp. 419–426, Ann Arbor, Jun. 2005.

[8] A.Nag!, S. Biswas*, D. Sarkar*, P.P. Sarkar*, B. Gupta**.(2010,Jun.). “A Simple Feature Extraction Technique of a Pattern By Hopfield Network”, “in press” *International Journal of Advancements in Technology*, ISSN 0976-4860, Vol 1, No 1,pp.45-49.

[9] H.Liu1, 3, A.Abraham2, and B.Yue3," Nature Inspired Multi-Swarm Heuristics for Multi-Knowledge Extraction", "in press" *Advances in Machine Learning II*, pp. 445–466, 2010.

[10] Spyns,P."EvaLexon:Assessing triples mined from texts".*Technical Report 09*,STAR Lab,Brussels,Belgium,2005.

[11] Velardi,P.,et al."Evaluation of OntoLearn,a methodology for automatic learning of domain ontologies.In:Ont.Learning from Text:Methods Evaluation and Applications".*IOS Press*,2005.

[12] Weng,S.,Tsai,H,Liu,S.,and Hsu,C."Ontology Construction for information classification ,*Expert Systems with Applications*",pp.1-12.

[13] Christine W. Chan.(2004)"From Knowledge Modeling to Ontology Construction". *International Journal of Software Engineering and Knowledge Engineering (IJSEKE)*.

[14] G.Cui, Q.Lu, W.Li, Y.Chen.Corpus Exploitation from Wikipedia for Ontology Construction: pp.2125-2132.

[15] C.Chou, F.Zahedi, H.Zhoa."Ontology for developing websites for natural disaster management: methodology and implementation", 2008.

[16] Harith Alani."Position paper: Ontology construction from online ontologies".in *Proc. the 15th international conference on World Wide Web* . 2006,pp. 491 - 495.

[17] Antonio M. Rinaldi."An Ontology-Driven Approach for Semantic Information Retrieval on the Web". *In ACM Transactions on Internet Technologies*, 2009,Vol. 9, Article 10.

[18] X.Binfeng,Luo, Xiaogang P.Cenglin, H.Qian."Based on ontology: construction and application of medical knowledge base".*IEEE International conference on complex medical engineering*, 2007,pp.586- 589.

[19] A.Rafea, Hesham A. Hassan, M.Yehia Dahab."TextOntoEx: Automatic Ontology Construction from Natural English Text". *International conference of Artificial Intelligence and Machine Learning,* 2006.

[20] A.A. Barforoush, S.Soltani.(2009)."Web Pages Classification Using Domain Ontology and Clustering". *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*,.

[21] A.Moreno and D.Sánchez.(2008)."Learning non-taxonomic relationships from web documents for domain ontology construction".*Intl.jl.of Data and Knowledge Engineering*.

[22] S. Hubner, H.Neumann, H.Stuckenschmidt, G.Schuster, T.Vogele, U.Visser, H.Wache."Ontology-Based Integration of Information-A survey of Existing Approaches".2001.

[23] M.A. Ismail, M.Yaacob & S.A.Kareem."Ontology Construction: An Overview".Malaysia, 2006.

[24] Thai National AGRIS Centre,Dept.of Soil Science,Faculty of Agriculture,Kasetsart University,Dept.of Computer Engineering,Kasetsart University,(2009)."Rice Production Knowledge Management:Criteria for Ontology Development".*Thai Journal of Agricultural Science*, Thailand.