Rich Lexical Knowledge based Q&A System for Ubiquitous Knowledge Service

Asanee Kawtrakul
U-Know Center and,
Department of Computer Engineering
Kasetsart University Bangkok, Thailand
asanee_naist@yahoo.com

Mukda Suktarachan
U-Know Center and,
Department of Computer Engineering
Kasetsart University Bangkok, Thailand
naist_da_da@yahoo.com

Navapat Khantonthong
Department of Computer Engineering
Kasetsart University Bangkok, Thailand
navapatk@gmail.com

Aree Thunkijjanukij
ThaiAGRIS Center Kasetsart University Bangkok, Thailand
thunkijja@yahoo.com

Patrick Saint-Dizier
IRIT-CNRS 118 route de Narbonne
Toulouse, France
stdizier@irit.fr

ABSTRACT
We present the concept of a Question-Answering System for providing knowledge services. The system is based on a rice production and rice disease textual database which has been structured according to a number of ontological conceptual functions, and associated annotations. In this paper, the rich lexical knowledge is utilized for identifying semantic roles in a question, connecting with the domain knowledge base in ontology and text formats to response the questions.

Keywords
Question-Answering System, Knowledge Services, Lexical Knowledge, Ontology.

1. INTRODUCTION
In this short text, we summarize some aspects of the development of Question-Answering system for farmers through SMS by focusing on the language aspects. The communicating via SMS facilitates a ubiquitous and effective knowledge service in problem solving, decision making and early warning. Our application area is rice production and rice diseases.

To get a better grasp at the problem and to be able to characterize it in depth, we got a collection of 1000 questions raised in real life from farmers. We have annotated those 1000 questions and the text(s) identified as responses for each query. This allowed us to understand how questions can be answered. Since the agricultural knowledge base we are using (derived from Thai AGRIS: Agricultural Research Information System, specifications) has a rich conceptual structure [7], about 60% of the questions can be directly answered by transforming queries into a conjunction of conceptual functions of this schema via lexical descriptions and interpretation functions. However, for about 40% of the questions, this is not possible, in particular for evaluative questions (such as "what is the largest …") and How-to questions that are related to procedures. To deal with this latter set of questions, we developed a model based on response annotation in order to induce inference rules to match a question with its answer. This is particularly crucial when there is no straightforward response, for examples, when some forms of lexical inference are required, when the response is not a simple item, but a well-formed fragment of text, and chain of events leading to a consequence event, or a procedure [4] etc.

The project we present here emerged from a need of the real end-users, the Agricultural Land Reform Office, Ministry of Agriculture and Cooperative, Thailand, in the project of ALRO Cyber Brain [1,2], which is a social network framework that combines approaches based on knowledge engineering with language engineering. Conceptual knowledge is represented in ontology through ontology workbench [8] for responding the factoid questions. New knowledge in textual format is extracted for maintaining ontological knowledge and responding non-factoid questions. We present below a brief outline of the main problems we have encountered.

2. TOWARDS A 'REAL' QA SYSTEM: CHALLENGES
First, at the level of QA analysis, several problems arise to identify the facets of the question: the type of the question, its focus and the constraints that hold on the focus.

For complex questions, another challenge is to identify its contents. Our approach is, via a dependency parsing approach, to tag NPs (noun phrase) and PPs (preposition phrase) by means of semantic tags, which correspond to the categories of the AGRIS database. In natural language, a question can indeed be asked with different words and syntactic forms. This is particularly the case in Thai, which allows for a lot of optional terms with a large constituent order freedom.

Next, in most cases, questions and answers do not match directly because the clue words or focus words in the question never appear in the answers. This obviously causes difficulty in finding the expected answer. For this kind of Q&A matching problem, some lexical semantics devices or more elaborated reasoning schema, based on domain knowledge are needed to allow appropriate question-response matching [5,6]. This is realized in our project via text annotation and learning.

In some cases, some information is missing to elaborate a real diagnosis, in that case the user is asked to provide more details. We prefer to avoid settling a dialogue, since this may lead to unexpected data or directions. Users want a relatively fast response, therefore just asking for more free input is the best compromise. The second aspect of this problem is to be able to extract the complete text portion in a text that responds to the question. For that purpose we have developed an annotation methodology whose goal is to identify the different processes at stake and the needed resources. This method allows us to identify relevant text portions and then to delimit them appropriately.

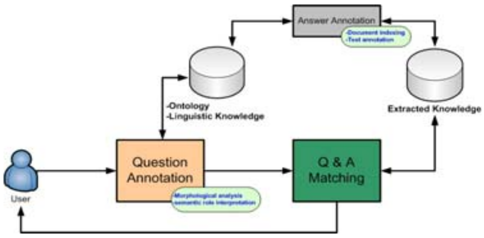

Figure 1. Question Representation and Corresponding text Indexing

Our question answering system is based on three sources of knowledge which interact:

- lexical data and in particular lexical semantics, and lexical inference,
- the domain data as represented by the rich conceptual functions ,i.e. Rice Ontology,
- some general purpose knowledge, useful for answering questions.

Lexical representations of verbs are based on conceptual functions from Framenet. The general form is: Verb + argument selectional restrictions: conjunction of conceptual functions (with variables corresponding to argument positions). For example: **resist:** *verb, [X:NP, Y:NP],[X:plant, Y:insect ∨ disease], X 'isResistantTo' Y*. Nouns are associated with their types as defined in the domain ontology.

While the semantics of verbs can be represented on the basis of conceptual functions, more complex situations, e.g. the adjunction of constraints, often expressed by syntactic adjuncts, need further developments. The first difficulty is to develop a compositional framework that can integrate various modifiers. For that purpose, we reuse the semantic representations we developed based on the Lexical Conceptual Structure principles that we have integrated into the PrepNet lexical base. PrepNet proposes semantic representations for a large number of forms of adjuncts based on a notion a prepositional modification, which is what is encountered in complex questions. For ontology based Question and Answering system, we apply Thai Rice Ontology [7] as a source of rich knowledge. The corresponding answer could be extracted with simple algorithm by matching query with ontological relation and then grasping it's associations with inference rules as an answer as in Figure 2.

> *Q: What are the disease of rice caused by fungi ?*
> With inference Rule*: X isDiseaseOf Y ∧ X isa Fungi* we can get the answer as the following.
> *A: Magnaporthe grisea*

Figure 2. Example of an ontological inference rule.

Then the answers will be collected more from full-text by matching the question to extracted full-text (See Figure 3).

3. CONCLUSION

This short paper presents some ideas on how to utilize the rich lexical knowledge for annotating question and answer in both indexing and text level. The application of presented methodology has been implemented on knowledge services for the Thai farmers in Rice domain and is under testing. Moreover, with the Rice Ontology with 2322 concepts, 5603 terms, 57 associative relations, 60% of questions can be directly answered by transforming queries into one or more conceptual functions via lexical inference rules and interpretation functions.

```
Function Matching (Question Q, Answer A){
    Match = false;
   // Relevant document
   If  (Q.focus  =  A.index)  then
         // Relevant answer
         If  (Q.type =  A.task type) then
               //Detect Answer for the Question
               If (Q.focus = A.title) then
                    Match = true;
               Else if (Q.action = A.action and
                         Q.theme = A.theme or
                         Q.agent = A.agent) then
                    Match = true;
               End If
         End If
    End If
    Return Match;}
```

Figure 3. Full-Text Q&A matching Algorithm

4. ACKNOWLEDGMENTS

The work described in this paper has been supported by the NECTEC No. NT-B-22-KE-12-50- 19, within the project, I-Know II: CAT, EAT, RATs, and Agricultural Question & Answering Service System, granted by the KURDI, Kasetsart University. We would also like to thank the French CNRS PICs programme.

5. REFERENCES

[1] Kawtrakul, A. et. al. 2009. Problems-Solving Map Extraction with Collective Intelligence Analysis and Language Engineering. *Book Chapter 18, Medical Information Science Reference in Information Retrieval in Biomedicine.* ISBN: 978-1-60566-274-9; pp 460

[2] Kawtrakul, A. et. al. 2009, From CyberBrain to Q&A Services: A Development of Question - Answering Services System for the Farmer through the SMS. In *Proceedings of WCCA2009.* Grand Sierra Resort, Reno, Nevada, USA.

[3] Moldovan, D. et.al. 2000. The Structure and Performance of an Open-Domain Question Answering System, *Proceedings of the 38th Meeting of the Association for Computational Linguistics (ACL)*, Hong Kong.

[4] Estelle Delpech, Patrick Saint-Dizier. 2008. Investigating the Structure of Procedural Texts for Answering How-to Questions, *LREC2008*, Marrakech.

[5] Leonard Talmy. 1985. *Lexicalization Patterns: Semantic Structure in Lexical Forms, in Language Typology and Syntactic Description 3: Grammatical Categories and the Lexicon, T. Shopen(ed.),* 57-149, Cambridge University Press.

[6] Takechi, M. et.al. 2003. Feature Selection in Categorizing Procedural Expressions, *The 6th International Workshop on Information Retrieval with Asian Languages (IRAL2003)*:49-56.

[7] Thunkijjanukij, Aree, Ontology development for Agricultural research Knowledge Management: a case study for Thai rice, PhD Dissertation, Kasetsart University, 2009

[8] Kawtrakul A. et.al.2008. "Ontology based Knowledge Map Construction for a Smart Knowledge Service" *IAALD AFITA WCCA 2008*, Tokyo, Japan, 24 - 27 August.