CONCLUSION AND RECOMMENDATION

Conclusion

Ontology with a precise semantic is very important for improving information systems by and large, for automating reasoning, as well as for sharing and managing knowledge. In this work, we are interested in building an ontology that is both taxonomic and non-taxonomic; hence we included relations like meronyms, and functional relations. One of our aims is to solve the ontology development bottleneck problem by exploiting the enormous body of knowledge gathered over the years in various types of classification systems and thesauri. Then AGROVOC and a Plant dictionary have been used as a seed to transform these terms to ontological terms. Terms grow very fast in many domains. In the case of agriculture, this is illustrated by terms like, *plant species, disease name, chemical names, and pathogen names.* This is where text corpora become very useful, since they contain a lot of, frequently updated information. However, texts present also the hardest challenge, as they require much more work concerning the acquisition of ontological terms and mining their relationships.

We presented and evaluated here the learning methodologies for the automatic building of ontology that is composed of term and relation extraction. The ontology is constructed from several resources: text corpus, thesaurus and dictionary. In order to build ontology of Thai text, a shallow parser is used for candidate terms extraction, and cues-words: lexico-syntactic patterns and item list (numbering and bullet list) are used for relation extraction. Concerning the lexico-syntactic patterns, there are some problems of many candidate terms and cue word sense ambiguity, then the lexicon and the co-occurrence features of each candidate term are used to solve this problem. We also applied information gain ratio for weighting each feature to measure its relevance. This technique can be used to extract the hypernym term of the item lists from the set of candidate terms. One of the most important advantages of using cues is that it reduces the problems of concept and relation labeling which are the crucial problems of the research of ontological engineering.

Concerning with the thesaurus-based ontology construction, we propose three methodologies for data cleaning and semantic relationship refinement to solve the problem of producing well-defined semantics from poorly defined or underspecified semantics in a thesaurus. The system refines the semantic relationships though noun phrase analysis, WordNet alignment, and semantic relationship rules, some generated by experts and others generated from annotated examples by an inductive statistical machine learning system. Finally, the relationships were verified by the experts. Initial results are promising. Moreover, in order to extract ontological terms and relationships from a specific dictionary, a task oriented parser is used to build the ontological tree. Finally, all integrated sub-ontologies are integrated to the core-tree by using term matching technique and the ontology is reorganized for pruning the inconsistency relationships.

We consider our results quite good, given that the experiment is preliminary, but the vital limitation of our approach is that it works well only for documents that contain a lot of cues. Based on our error analysis the performance of the system can be improved and the methodologies can be extended to other sets of semantic relations and applied to other domain. However, concerning the process of pruning the redundancy relations, we analyze that these redundancy relations have the benefit for analyzing the using of words in text. For example, the relation *HYPONYM(animal, mammal)* and *HYPONYM(mammal ,cat)*. In our system, if we found the relation of *HYPONYM(animal, cat)*, the system will delete this redundancy relation. However, we found that texts sentence ‘cat is an animal.’ more than the sentence ‘cat is a mammal.’ then the redundancy relations have the useful for analyzing text and they can show the using of words in text.

Further works to complete the research are performing more tests on large corpora and evaluating the ontology by using in the real applications. Another research direction is to extract the semantic relation embedded in the sentence without the cues and to represent the confidence of relation by adding the frequency of the occurrence of each relation. Moreover, the techniques for translating Thai to English and selecting words' sense can be applied for helping the task of Thai WordNet construction.

Recommendation

Although there are many studies in the field of ontology engineering, it still has many open problems. Special attention to improve the field must be given to the following tasks.

Axiom learning: The only report we found on learning axioms is by HASTI (Shamsfard and Barforoush, 2003), which learns some axioms in restricted circumstances. This system learned the explicit axioms from conditional sentences in texts. The implicit axioms from text need more attention.

Task ontology learning: Most of the methods for building ontology are focused on domain specific ontology. However, task ontology is another important knowledge for supporting the process of real world application. Hence, automatic construction of task ontology needs more research.

Evaluating ontology learning systems: Before using the ontology in applications, it should be evaluate the content in the aspects of consistency, completeness and conciseness (Gomez-Perez, 1996). Currently ontology learning systems (Gomez-Perez, 1999; Gruninger and Fox, 1995) are evaluated their results in specific domains. Special attention must be given to find formal, standard methods to evaluate the ontology learning systems.

Managing the evolution of the ontology: Ontology is a set of dynamic entities that evolves over time. While maintaining it throughout its life-cycle is an endless task, since new terms are created, others becoming obsolete. In the domain of agriculture, certain terms grow very fast, for example, *plant species, disease name, etc.,* while others tend to disappear. They have lived their time, or they have simply become obsolete with the appearance of new technologies. For example: *pineapple* was an instance of an *economic plant* in the year *2002* while in *2004*, *oil palm* has become an economic plant instead as illustrated in Figure 39.

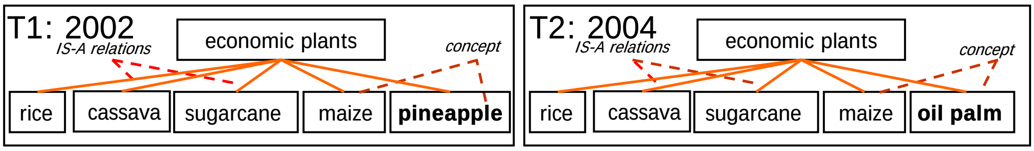

Figure 39 Examples of domain specific ontologies for an economic plant at different points in time

The management of ontology evolution and the relationships between different versions of the same ontology are crucial problems to be solved.