

Aurawan Imsombut 2007: Automatic Thai Ontology Construction from Corpus, Thesaurus, and Dictionary. Doctor of Engineering (Computer Engineering), Major Field: Computer Engineering, Department of Computer Engineering. Thesis Advisor: Associate Professor Asanee Kawtrakul, D.Eng. 117 pages.

Ontology has a crucial role to play in information retrieval; however, its building by experts is an expensive task, and also a never-ending process which relies on evolution of knowledge especially in science. Hence, we suggest learning ontologies automatically in order to spare experts the bulk of the job.

We present here a hybrid approach for especially building and maintaining (semi-) automatically an ontology from corpus, and also thesaurus and dictionary. Concerning the corpus, we propose a methodology for extracting ontological concepts and taxonomic relations by using explicit cue expressions, i.e. lexico-syntactic patterns and an item list. However, this technique has several difficulties i.e. cue word ambiguity, item list identification, and numerous candidate terms. We solved these problems by using the lexicon and co-occurrence features and using information gain for weighting these features. Moreover, we fill the ontology with the semantic relations embedded in Thai NPs by translating Thai words into English, with a method of selecting the word sense from WordNet and by applying machine learning techniques to learn the semantic relations. In order to extract ontology from a specific dictionary, a task oriented parser is used to build the ontological tree. Moreover, we refine the thesaurus' relationships to ontological relations by using machine learning and some heuristic rules. Finally, we integrated all the ontological sub-trees collected by using the technique of term matching and then the ontology is reorganized for consistency. We tested the system by using Thai corpora in the domain of agriculture and the accuracy of the final result from 3 resources is 0.86.

---

Student's signature

---

Thesis Advisor's signature