RESULTS AND DISCUSSION

Evaluation Methods

The measurement of the system's performance was based on test cases in the domain of agriculture and divided according to the resources: text, thesaurus, dictionary and merging ontologies from all resources. To evaluate each methodology, we assumed that its inputs from the results of preprocessing are correct. To evaluate the results, the outputs of the system are compared with the results produced by the agreement of two experts of agriculture from Thai National AGRIS Center (2003). We use standard well-known measurements to assess the different approaches i.e. precision, recall and F1 variant of the popular F-measure. Precision (P) is the ratio of the number of extracted correct results to the total number of extracted results. Recall (R) is the ratio of the number of extracted correct results to the number of total correct results prepared by manually. Precision and Recall are given by standard formulas:

$$precision = \frac{\text{the number of extracted correct results}}{\text{the number of total extracted results}}$$

$$recall = \frac{\text{the number of extracted correct results}}{\text{the number of total correct results}}$$

The F-measure (F) combines the two parameters precision and recall. It was first introduced by Rijsbergen (1979). The standard formula is defined as:

$$F1 = \frac{(b^2 + 1) \cdot precision \cdot recall}{b^2 \cdot (precision + recall)}$$

b is a factor to quantify the value of precision and recall against each other. For the consequent test runs we use b = 1.

Results and Discussion

The experimental results of the system are summarized according to the resources: text, thesaurus dictionary. Moreover, the performances of the ontology integration and reorganization system are discussed in the end of this section.

1. Text-based ontology construction

The system's performance of this resource is evaluated according to the methodologies: cues-based ontology extraction and machine learning-based semantic relationship extraction from NPs.

1.1 Ontology Learning by using Cue

Concerning the calculation of the feature weights, the training corpus consists of 1,500 examples with positive and negative class. From the experiment by using several techniques: Information Gain, Information Gain Ratio and SVM (with linear kernel), for calculating the feature weights, we found that Information Gain and Information Gain Ratio give the similar results but Information Gain Ratio being better. SVM has a lower precision and it is difficult to set the proper kernel parameter. The weights of features from Information Gain Ratio are shown in Table 18.

Table 18 The Information Gain Ratio (or weight) of each feature.

	f_1	f_2	f_3	f_4	f_5
Information gain of lexico-syntactic pattern	0.001	**0.069**	**0.139**	0.014	0.029
Information gain of item list	0.050	0.190	0.045	0.043	**0.214**

From Table 15, we can conclude that the properties list feature (f_3) is the most important feature since it can prune the ambiguous cue words in patterns. Moreover, we found that the NE feature (f_2) is the next important feature for selecting the candidate term of the pattern-based ontology learning because NE usually occurs in the agricultural document. We can conclude that this feature is appropriate for extracting specific domain term. If the candidate terms had the same NE class as the related term they should be selected. In addition, the co-occurrence feature (f_5) has the crucial role for selecting hypernym term of the list item terms because the hypernym and hyponym terms have more co-occurrence value than the other terms in the candidate term set. Conversely, the head word compatible feature (f_1) rarely occurs with the lexico-syntactic pattern and this feature usually occurs in the general domain documents. Hence, this feature is not significant for using to select the candidate term in the specific domain, as well as, the topic term feature (f_4) on the item lists.

In the evaluation process, we test the system with 3 aspects. First, we measure the performance of the system classified by the type of cues. Next, the system is evaluated based on two different test corpora i.e. technical documents and Thai encyclopedia. Finally, we compare our methodology with the Hearst's technique.

1.1.1 Evaluation based on the type of cues.

Based on our assumption that cues, lexico-syntactic patterns and item lists, could be used as the heuristic information for hinting the ontological relationship, we test the system based on each cue. As shown in Table 19, the results from using item lists are 0.83 of the precision, 0. 81 of the recall and 0.82 of the F-measure while the precision, recall and recall when using lexico-syntactic patterns as cue are 0.64, 0.69 and 0.66, respectively. The results of item lists are higher than the results from lexico-syntactic patterns since the item lists do not have the problem of anaphora and cue word ambiguity. However, the errors of item-list cue technique occur because some bullet lists are composed of two classes, for example, disease and pest. Then the system has the error in the detecting the paragraph that contains the hypernym term of the items in the second list.

Table 19 The evaluation results of the system classifying by the cue types.

Cues	P	R	F
Item list (a)	0.83	0.81	0.82
Lexico-syntactic pattern	0.64	0.69	0.66
Lexico-syntactic pattern with anaphora solving (b)	0.71	0.78	0.74
Both cues (a) + (b)	**0.74**	**0.78**	**0.76**

The important errors of pattern approach are caused by many sentences contained anaphora terms, then the system can not extract the correct ontological terms. The anaphoras causing these problems are the direct reference. They are definite NPs and zero anaphora. From observation, some of these anaphora terms can be solved by using heuristic rule by getting the subject of previous sentence. This method can increase the precision value by 7%, i.e. raising the level from 0.64 to 0.71. Even if including the anaphora resolution, the precision is not so high since it still has the problems of the ambiguity of the cue word /pen/ (be) that can be meant to other non-taxonomic relations i.e. being-state-of, being-status-of and made-of as shown in the example (18), (19) and (20), respectively.

*(18) */nai/(in) /raya/(period) /pen/(be) /tua-on/(young)…*
(Being in the young period)
*(19) */khao/(he) /pen/(be) /sot/(single)*
(He is single.)
*(20) */rongruean/(house) /pen/(be) /khrong-Lek/(steel structure)*
(House structure is made of steel.)

These problems can not be solved by keeping all non-taxonomic terms as a list like the property terms for pruning non-taxonomic relation of cue word /pen/ (be).

1.1.2 Evaluation based on different data set.

Dataset 1: Technical documents in the domain of agriculture. By testing with a dataset 1 (about 277,164 words), the system is able to extract about 2,043 concepts and 2,154 taxonomic relations when using both the lexico-syntactic patterns and the item list. The precision, recall and F-measure of the system when testing with this dataset are 0.75, 0.78 and 0.77, respectively.

Dataset 2: Thai encyclopedia in the topic of plant. The size of dataset 2 is about 25,476 words. By using both cues the system can extract about 224 concepts and 198 taxonomic relations. The precision, recall and F-measure of the system when testing with this dataset are 0.63, 0.72 and 0.67, respectively. The accuracy of the experiment with this dataset is less than the previous one because the documents in this genre contain a few ontological lists.

The evaluation results of the ontology extraction system classified by the data test set are shown in Table 20. From both data sets, the system can extract totally 2,228 concepts and 2,325 relations. The performances of the system, in total average, are 0.74 of the precision, 0.78 of the recall and 0.76 of the F-measure. As mentioned above, there are some ambiguities of the cue words /pen/ (be) that still remain in both datasets.

Table 20 The evaluation results of the system classifying by the data test set.

Data Set	**Test Corpus size(words)**	**No. of concepts**	**No. of relations**	**P**	**R**	**F**
DataSet1	277,164	2,043	2,154	0.75	0.78	0.77
DataSet2	25,476	224	198	0.63	0.72	0.67
Total	**302,640**	**2,228**	**2,325**			
average				**0.74**	**0.78**	**0.76**

1.1.3 Comparing with the Hearst's technique.

In this sub section, the system was evaluated by comparing with the Hearst's technique (Hearst, 1992), the most well known of pattern-based technique. Hearst's technique selected only terms that occur nearest the cue word of lexico-syntactic patterns to be the ontological terms. However, the ontology terms might occur far away from the cue word. This research then proposes the technique for selecting the correct ontology term even it was far away from the cue word. Figure 38 shows the comparison between applying Hearst's technique and our technique by varying the data test size. The experimental results show that our technique has the higher accuracy than Hearst's technique.

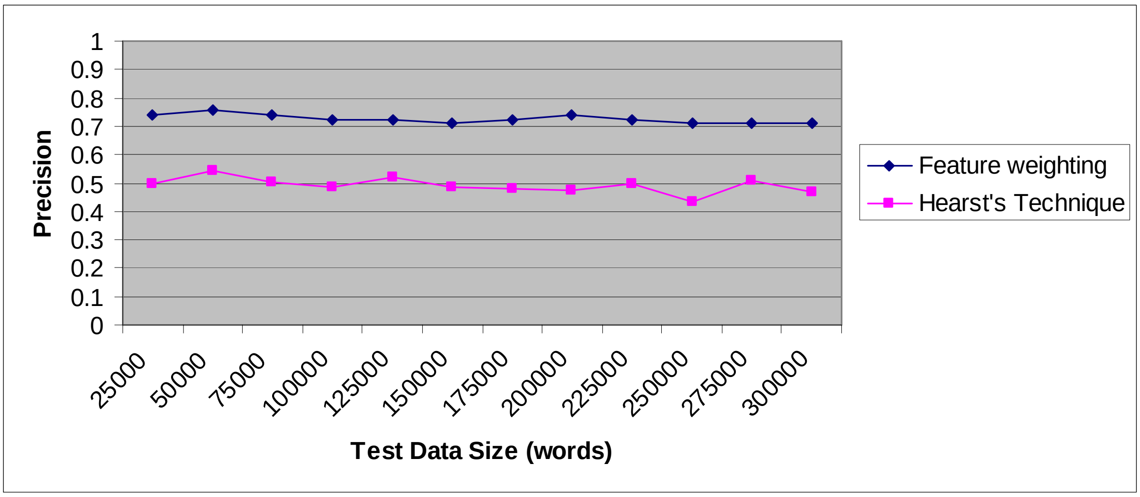

Figure 38 The accuracy of the systems classifying by the technique of selecting term.

1.2 Discovery of the semantic relations of nouns in NPs

In our experiment, we select 1055 pairs of head and modifier of NPs expressing 10 relationships. The 1055 pairs are compound noun that composed of two or three nouns. There are only 107 pairs that composed of three nouns then they need to be segmented or bracketed for identifying the head and the modifier. From the experimental result, the precision of this bracketing process is 0.71 and the errors are caused by the incorrect pairs of words are occurred more than the right one. For instance, "[*/kak/(lees) /malet/(seed)*] */fai/(cotton) (cotton meal)*". Altogether, from 1055 pairs, the system can translate and disambiguate them, 347 pairs, by using a thesaurus to achieve a precision of 0.96 and a recall of 0.34. The remaining term pairs are processed by using a dictionary to obtain 420 pairs from 708 NPs. with a precision of 0.97 and a recall of 0.59. Table 21 shows the experimental results of the system for translating and selecting words' sense. When combining thesaurus- and dictionary-based translation, the precision and recall of the system are 0.97 and 0.73, respectively. We found that many terms that could not be processed by our system were domain specific terms absent from our knowledge resources.

Table 21 The experimental results for translating and selecting words' sense

Technique	precision	recall
Thesaurus-based Translation	0.96	0.34
Dictionary-based Translation	0.97	0.59
Thesaurus-based + Dictionary-based Translation	**0.97**	**0.73**

Concerning the learning of semantic relations, we did an experiment with 767-tagged examples for 10 semantic relations. The decision tree learning generated about 90 classification rules and the SVM learning system generated 10 learning models accordingly to the relations. The experimental results are shown in Table 22. Both learning systems show that the 'Topic' relation has the highest precision, as all NPs expressing this relation have the head word in the class 'data' or 'information'. The following examples show the NPs having the class 'data#1' as head word.

(21) */khomun/(data#1) /phum-akat/(weather#1)*
(weather data)
(22) */khomun/(data#1) /sathiti/(statistic#1)*
(statistical data)

Besides, the 'IS-A' relationship achieves the lowest precision. This is due to the fact that there is a problem of data sparseness then it is difficult to discovery the common semantic class of the head word or modifier word of NPs which embedded this relationship. The data sparseness is caused by a variety of heads and modifiers such as *plant* or *animal species* as follows:

(23) */khao/(rice#1) /bale/(barley#1)*
(Barley)
(24) */nok/(bird#2) /krachokthet/(ostrich#2)*
(Ostrich)
(25) */phlia/(aphid#1) /kai-chae /(bantam#1)*
(Durian psyllid)

Table 22 The evaluation results concerning the discovery of the semantic relation from NPs.

Relation	**SVM**		**Decision tree**	
	P	R	P	R
IS-A	0.77	0.70	0.61	0.75
container	0.70	0.54	0.70	0.54
location	0.82	0.80	0.75	0.76
made-of	0.82	0.82	0.76	0.85
purpose	0.83	0.67	0.65	0.57
possessor	0.85	0.86	0.86	0.74
property	0.85	0.76	0.88	0.74
source	0.88	0.82	0.80	0.67
part-whole	0.89	0.83	0.86	0.76
topic	1.00	0.90	1.00	0.90
Average	**0.84**	**0.77**	**0.79**	**0.73**

2. Thesaurus-based ontology construction

In this process, the AGROVOC thesaurus relationships are cleaned and refined before converting to ontological relationships. The BT/NT relationships are cleaned and converted to hypernym/hyponym relationship by using expert-defined rules, NP analysis and WordNet alignment techniques while the USE/UF relationships are cleaned and converted to synonym relationship by using only WordNet alignment technique. The RT relationships are refined to more specific relations by using expert-defined rules and training rules. We ran an experiment testing the training rules technique for refining the RT relationship by using 100 examples with 5 semantic relationships i.e. 'scientificNameOf' *(Malus <scientificNameOf> apples)*, 'usedToMake' *(almond <usedToMake> almond oil)*, 'subclassOf' *(avocados <subclassOf> tropical fruits)*, 'producedFrom' *(goat milk <producedFrom> goats)*, derivedFrom *(rice straw <derivedFrom> rice)*. It produced around 10 classification rules. The experimental results using these rules as well as expert-defined rules, noun phrase analysis, and WordNet Alignment are shown in Table 23.

Table 23 The experimental results classified by relationship

Relationship	**No.**	**No. of refinement**	**Expert-defined rules**		**NP Analysis**		**WordNet Alignment**		**Training Rules**	
			No.	P	No.	P	No.	P	No.	P
BT/NT	32176	21072	16587	1.00	2062	0.95	2423	0.95	**	**
USE/UF	21605	3553	-	-	-	-	3553	0.70	**	**
RT	27589	1420	622*	1.00*	-	-	-	-	798*	0.72*
Total	**81370**	**26045**	**17209**	**1.00**	**2062**	**0.95**	**5976**	**0.80**	**798***	**0.72***

Remarks: - indicates this technique can not revise this relationship, * indicates the experiment is run with some data, ** indicates the experiment is in initial state.

Based on an expert's review of a small sample of data, some initial rough estimates were made regarding the precision of the methods. The precision of the Expert-defined Rules technique was estimated to be around 1.00 and 0.95 correctness for NP Analysis. The WordNet Alignment technique was estimated to be lower, about 0.94 precision, because some synonym relationships in WordNet should be replaced with the 'abbreviation_of' relationship. For example, in *AMP* <synonym> *Adenosine monophosphate*, <abbreviaton_of> should be used. The precision of the Training Rules technique was estimated to be about 0.72. Sources of error include ambiguity in concept classes used as arguments for a given rule, such as the following, 'If class X is *food#1* and class Y is *food#1*, and X RT Y, then X <*usedToMake*> Y' where, because X and Y belong to the same concept class, the system cannot distinguish between X and Y and may generate erroneous relationships, e.g., *pork <usedToMake> hams*, and *hams <usedToMake> pork*. These cases can be revised by using information from text such as the sentence '*pork is used to make ham*' and it needs the technique of verb analyzing for identifying the semantic relation.

3. Dictionary-based ontology construction

The experiment in this step, we can extract 37,110 terms with 21,620 relations from Thai Plant Name dictionary. By random checking, the accuracy of the system is 1.00 because this specific dictionary contains explicit taxonomic structures then it is not a complex task for extracting ontological terms and relationships.

4. Ontology Integration and Reorganization

The process of ontology integration is iteratively occurred when the system adds each extracted concepts and relationships to the core tree. The system integrated 1,452 relationships that extracted from corpus and dictionary to the core tree with term matching technique and 590 relationships with partially terms' head words matching technique. The accuracies of these techniques are 0.82 and 1.0 as shown in Table 24.

Table 24 The evaluation of ontology integration

Technique	Nb. of Rel.	Accuracy
Term matching technique	1,452	0.82
Partially terms' head words matching	590	1.0

Although these methods are looked simple, they give a promise results. The errors of term matching technique are caused of the terms have the same label with different meanings. For example, the term "/*kaew*/" can mean a flower or a variety of mango. Another example is /*chomphu*/ that can be '*rose apple*' or '*guava*' (called in the south of Thailand). This problem will be solved in the future work by defining the rules or constraints e.g. flower is a disjoint concept of fruit and not considering the local names in the merging process.

Moreover, we evaluate the accuracy of the ontology integration system according to the problem's categories as shown in Table 25.

Table 25 The evaluation of ontology reorganization system classified by the problem's categories

Problems Categories	Frequency	Accuracy
Term Mismatch	142	0.46
Redundancy in the class hierarchy	223	1.00
Conflict relationship	13	0.77

The accuracy of the system for merging the terms that have different names or have the problem of term mismatch is 0.46. The result shows that our proposed technique by using edit distance for solving the problem of term mismatch is work well only for transliterate terms e.g. */dai then em 45/ (Dithane M-45)* and */dai thaen em 45/ (Dithane M-45)*. However, this technique does not work with noun phrases that have the same semantic generated from different terms. These terms do not have the similarity when measured by using edit distance technique. For example,

(21) */ya/(medicine) /kha/(kill) /malaeng/(insect)*
(Insecticide)

(22) */sankhemi/(chemical) /kamchat/(eliminate) /malaeng/(insect)*
(Insecticide)

For solving this problem, the system needs the knowledge of lexicon meaning for mapping the synonym terms such as */kha/(kill)* has the same meaning as */kamchat/(eliminate)* and the techniques of paraphrase resolution are helpful. Concerning the problem of redundancy in the class hierarchy, the system can delete all redundancy relationships then the accuracy is 1.00. Moreover, the accuracy of conflict-relationship problem solving is 0.77. We used the statistical technique to solve this problem by deleting the relationship that has less frequency than another conflict relationship. However, some less frequent relationships could be used to imply the correct relationship e.g. the correct relation *HYPONYM(soil, loam soil*) has less frequently than the incorrect relation *HYPONYM(loam soil, soil).* Some relationship like this example can be solved by analyzing the head word of NP i.e. *soil* is the head word of phrase *loam soil* then *soil* should be the hypernym of *loam soil*.

Table 26 shows the number of concepts and relationships, extracted from each resource, as well as the total number of concepts and relationships in the ontology resulting from the process of organization.

After a random check with 1,000 integrated terms, the organizing system's accuracy is 0.86 and the coverage is 0.90. The errors are due to the particular characteristics of the corpus extraction terms

Table 26 Experimental results classified by the resources

Source	Methodology	Nb. of Terms	Nb. of Rel.	Accuracy	Coverage
Text (90 docs.)	Ontology learning by using Cues	2,228	2,325	0.74	0.78
	Relation Extraction in Phrase Level	585	767	0.97*0.84 = 0.82	0.73*0.77 = 0.56
Thesaurus		33,450	26,045	0.87	0.87
Dictionary		37,110	21,620	1.00	1.00
3 Sources		**59,971**	**41,677**	**0.86**	**0.90**

When evaluating the system with three criteria, accuracy, coverage and portability, we can conclude as follows:

Accuracy: We evaluate the accuracy of the system by using the precision value. The precision of the ontology extraction system based on the thesaurus and dictionary is very good since they are structure sources. When considering only text-based ontology extraction, the average accuracy of the system for extracting both taxonomic and non-taxonomic relations is 0.76. Concerning the relation extraction in phrase level, it composed of two sequence steps that are 1) word translating from Thai to English and word disambiguating process and 2) semantic relationship learning, and then we multiply their precisions (i.e. 0.97*0.84) as the precision of this task that is 0.82 of the accuracy. The errors of text-based ontology construction are analyzed as previous mentioned i.e. cue word ambiguity and various semantic classes of head noun and modifier noun of NPs. Based on this analysis and proposed solutions, the performance of the system can be improved.

Coverage: The recalls are considered as the coverage of the system. The recall values of thesaurus- and dictionary-based ontology extraction system are equal to the precision values because these sources contain exactly number of terms and relationships. Hence, the number of total extracted results is equal to the number of total correct results then the recall value is equal to precision value. These two sources give promise recalls. Concerning the text-based ontology extraction, the coverage of the system is 0.71. The recall of relation extraction in phrase level is very low because there are many terms that are domain specific terms and absent from our knowledge resources i.e. dictionary and thesaurus. Thus, these terms can not be translated. For solving this problem, we can apply NE classes for tagging semantic classes of these terms, e.g. /*namdokmai*/ is a kind of mango and its NE class is '*plant*', and the coverage of the system will be increased.

Portability: There are very difficult for evaluating this criterion as a quantity value. Accordingly, we evaluate this criterion by analyzing the competency of our techniques for applying to other domain and other language. Concerning text-based technique, our cues are very general, i.e. the cue word: /chen/(such as), /dai-kea/(i.e.), /pen/(be) and the item lists: bullet list and numbering list, and they can occur in other domains and other languages. By this reason, we can conclude that our proposed techniques can be applied to other domain and they can be adapted for other languages by changing some linguistic grammars e.g. NP patterns. In addition, the proposed technique for extracting ontology based on thesaurus can be directly applied to other thesaurus and the dictionary-based ontology extraction technique, proposed here, can be applied to other dictionaries that have explicit structure by modifying the rules of task-oriented parser in the process of structure analysis. For extracting ontology from other dictionaries that do not have the explicit structure of terms, the ontological elements can be extracted from words and their definitions by applying the approach of text-based ontology extraction.