TH_WSD: Thai Word Sense Disambiguation Using Cross-Language Knowledge Sources Approach

J. L. Mitrpanont and P. Chongcharoen

***Abstract*—The ambiguity in Thai Word is still a significant issue in translating Thai language to English. This paper presents the TH_WSD, a framework for Thai word ambiguous resolution using cross-language knowledge sources of AsianWordNet (AWN) and PrincetonWordNet (PWN) for lexical and word sense explorers. A semi-automated Thai WSD approach for non-specific domain using four disambiguation techniques, word forms, and even window sizes is proposed. The disambiguation techniques include path, vector, vector_pair and lesk. The 250 context words from four target words group which are วัด (wat), หัว (hua), เก็บ (kep) and เกาะ (koh) from bi-text corpora of SEAlang and Concordance are studied. The experimental results show that using AWN with vector technique and PWN with path technique provides better accuracy. However, for Thai WSD included time consideration, the vector technique with AWN at five window size is suitable.**

***Index Terms*—Natural language processing, word sense disambiguation, Wordnet, cross-language, AseanWordNet.**

I. Introduction

The ambiguity in word sense (WSD) is a common problem in natural language processing (NLP) of many languages. Thai language also deals with ambiguous meanings or senses. For this reason, NLP applications for Thai language such as Thai-English word translation are confused and select incorrect word with the wrong meaning particularly in non-specific domain.

There are several approaches to determine sense for ambiguous words. Two widely used approaches are corpus-based and knowledge-based approaches or referred as supervised and unsupervised techniques. The corpus-based utilizes raw text from corpus that has sense-tagged for NLP applications [1]. However, the difficulty of manual sense-tagged in a training corpus decreases the NLP applicability. Many attempts have been made to solve the knowledge acquisition hindrance such as too many languages, too many words, too many senses, and too many examples per sense. Therefore, it is still an open problem of the supervised learning approach for Thai WSD. The knowledge-based approach disambiguates word sense by matching context with information from knowledge source [1] which consists of the dictionary, semantic network structure and definitions for the different senses of each word. Furthermore, it defines group of synonymous words by a synset that represents distinct lexical concept.

Manuscript received August 20, 2014; revised October 20, 2014. This work was supported in part by the Faculty of Information and Communication Technology (ICT), Mahidol University, Bangkok, Thailand.

The authors are with the Faculty of Information and Communication Technology, Mahidol University, Bangkok, Thailand (e-mail: jarernsri.mit@mahidol.ac.th, premchai@bangkok.go.th).

In general, manual sense-tagged in the corpus-based may limit its scalability and domain. Thus, the knowledge-based approach is selected which has fewer drawbacks and can encode fine-grained information [2] that is more appropriate for determining sense with higher precision than previous approach.

One of the most successful Thai WSD studied by Kanokrattanukul [3] which applied Machine Learning algorithms to create statistical models in order to perform WSD. The basic idea is using the decision list collocation algorithm to resolve disambiguation of two ambiguous Thai word sense such as หัว(hua) and เก็บ(kep) representing noun and verb forms, respectively. They analyzed the sense based on Thai dictionary. They suggested that the two positions of word surround the target word were sufficient for the disambiguation of both words. The sense indicators of both words are mostly on the right side. Besides, Pongpinigpinyo [4] worked on the multi-strategies with knowledge-based, corpus-based and hybrid-based approaches to resolve word ambiguity. They emphasized on corpus-based that employ an unsupervised method for disambiguation, perform the efficient and effective information retrieval technique called Latent Semantic Indexing (LSI) to disambiguate Thai noun and verb word sense. Their purpose was to use two Thai multiple-meaning word (polysemous), i.e., หัว(hua) and เก็บ(kep) for Thai noun and verb. They applied the vector-based distribution information measured for semantic disambiguation. The experiments showed the comparison to the baseline system for the disambiguation of หัว(hua) and เก็บ(kep). Although knowledge-based approaches have been applied in several fields of English NLP, they have not been consider much on Thai language, particularly in non-specific domain.

II. TH_WSD Conceptual Design

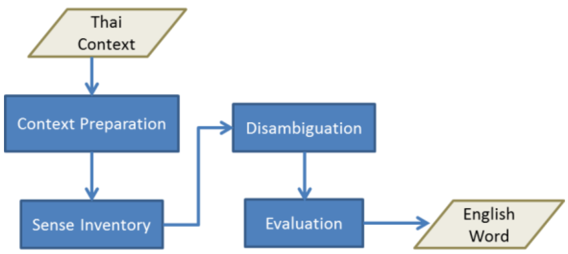


Fig. 1. TH_WSD system.

Currently, the multilingual machine translation is one of the top on-demand services. As a consequence, the idea for

DOI: 10.7763/IJCTE.2015.V7.997

Thai Word Sense Disambiguation (TH_WSD) using Cross-Language approach in order to utilize this method to reduce the problem of disambiguation in Thai-English Translation is proposed. Fig. 1 shows the conceptual design of the TH_WSD system functions.

TH_WSD consists of four modules common to the functions of general NLP tasks. Context Preparation module obtains Thai Context input and processes the word segmentation that parses sentence then eliminate non-informative words. Sense Inventory module is a control module to make sure that the translation of those Thai words to English using the lexical from both knowledge sources will be collocated to give variety of meaning and sense of the target word. Then construct the pair of sense between senses of ambiguous word and senses of word in the surrounding context in Disambiguation module to find the relatedness from the pair list by each algorithm automatically depending on word form type. Then mark the completion of the framework with Evaluation module.

TH_WSD is initially designed as a general framework then the concept and implementation methods are proved by a lot of work from our preliminary test [5] to clarify several processes, procedures and environment which will be designed, created and implemented. In addition, the test includes the physical structures used for test creation and implementation, as well as the logical interactions among those components intent to find the implementation possibility of TH_WSD framework. It involves NLP activity aimed at evaluating an attribute or capability of a framework and determining that it meets its required results. Some preliminary processes are listed below.

1) Thai Context Segmentation Testing using non-specific domain context such as news and novel
2) Knowledge-sources Testing and Measurement
3) Word Sense Disambiguation Techniques Testing
4) Implementation Language and Environment Setting
5) Thai Context Data Sets Selection and Testing

A. *Thai Context Preparation*

This module performs a process of resolving word-level boundary ambiguity because Thai context does not have space between words. There are many word segmentation techniques. In this module, KUCut [6], a word segmentation tool, is used however the linguist is still required to consider those segmented contexts correctness as a fine-grain verification to remove potential misleading word which will reduce amount of computing unnecessary words and minimize time for disambiguation. However, the un-informative elements or the stop list may be constructed accumulatively for use with other context domain automatically later.

B. *Sense Inventory*

In our approach, the cross-language is aimed. Thus, this module is designed as a control module to explore the existence of the sense of each Thai word on both knowledge resources. There are two main tasks in this module. First, use AWN [7] as Lexical to translate Thai to English words and verify existence of those words in PWN [8] to ensure the acquisition of the sense information of each word in the Disambiguation module. Second, the module computes the English word with PWN to explore all possible senses both ambiguous word and surrounding words. Obviously, not all Thai words are translated with AWN. However, the problem of senses from AWN and PWN unequally match will diminish the accuracy of the system. Practically, this module is designed to ensure that each English word exists in both sources; the amount of sense from PWN is not less than AWN. Consequently, the amount of Cartesian product of *Pair* process is exceedingly difference. Fig. 2 shows the component of Sense Inventory Module.

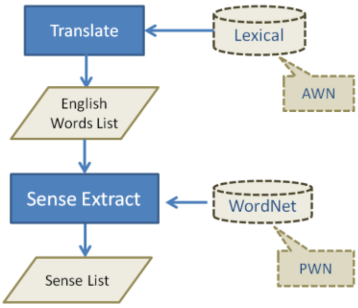


Fig. 2. Sense inventory module.

C. *Disambiguation*

This core module is designed to offer a semi-automated process to provide an appropriate choice of disambiguation techniques and word form selection. From [9], path similarity technique showed the highest accuracy to disambiguate noun word form with PWN. However, PWN does not provide relation between cross part of speech even in version 3.0. In our work, thus, the attempt to perform the word form mixture disambiguation in a semi-automatic way is done.

To identify the WSD techniques [9], the preliminary test on several techniques are performed and found that path similarity, lesk, vector and vector pair techniques provide the promising precision in specific context domain. As a result, they are adopted. This process will finally provide the automatic part of speech filter and automatically forward each sense to an appropriate technique. Each technique is briefly described below.

1) *Path similarity [9]*

It computes the semantic relatedness of word senses by counting the number of nodes along the shortest path between the senses in the IS-A hierarchy of the WordNet. The path length includes the end node. Since a longer path length indicates less relatedness, the relatedness value returned is the multiplicative inverse of the path length distance (D) between the two concepts:

$$R = \frac{1}{D} \quad (1)$$

If the two concepts are identical, then the distance between them is one; therefore, their relatedness (R) is also 1.

2) *Lesk [10]*

Lesk algorithm disambiguates by instance and compares glosses between each pair (P) of words in the window of context. If there are N words in the window of context then there are

$$P = N(N-1)/2 \quad (2)$$

There are a series of relation pairs that identify which synset provides the gloss for each word in a pair during comparison. For example, a relation pair might specify that the gloss of a synset of one word is to be compared with the gloss of a hypernym of the other word. The glosses to be compared are those associated with the senses given in the candidate combination that is currently being scored.

3) Context Vector [11]

The algorithm uses co–occurrence information along with the WordNet definitions to build gloss vectors corresponding to each concept in the WordNet. Numeric scores of relatedness are assigned to a pair of concepts by measuring the cosine of the angle between their respective gloss vectors. This measure is flexible in that it can make comparison between any two concepts without regarding to their part of speech.

4) Vector pair [9]

The word senses by second-order co-occurrence vectors of WordNet definitions. The relatedness of two senses is then computed as the cosine of their representative gloss vectors. Each gloss is converted into a second order vector by replacing the words in the gloss with co-occurrence vectors for those words. The overall measure of relatedness between two concepts is determined by taking the pair cosines between these expanded glosses. Then three pair cosine measurements are made to determine the relatedness of two concepts. The examples found in the glosses of two concepts are expanded and measured, so as the glosses themselves and the hyponyms of the two concepts. Then, the values of these three pair measures are summed to create the overall relatedness score.

These four techniques are candidates to our TH_WSD system. In Fig. 3, Disambiguation module selects each word from sense list and classifies part of speech of those pair with similar or different word forms. In case of the similar word form, path similarity technique is used; otherwise, three techniques of lesk, vector and vector pair are performed. PWN is used as the semantic knowledge source to compute the overall relatedness between each sense of the word. The results are determined by the similarity value.

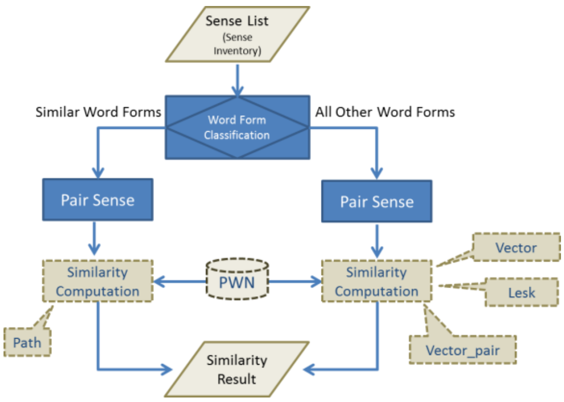


Fig. 3. TH_WSD disambiguation module.

D. Evaluation

This module makes sure that our framework contributes with valid results for the translation. It delivers the translation of ambiguous word from Thai open-domain context to English Word with correct sense. Practically, the translation has many level of justification. For instant, translation requires more fine-grained information about sense or meaning of that ambiguous word than second language understanding for general communication.

For this reason, the rules are constructed to classify the accuracy of the framework. There are 4 classes (A, B, C, D and F) as described in Table I. These classification rules are used to partition the accuracy in TH_WSD. It consists of rules number, the *First Condition;* illustrates the ambiguous word (w_1) and correct word (w_2) comparison; the *Second Condition* shows the correctness of the sense.

TABLE I: TH_WSD ACCURACY CLASSIFICATION RULES

Rules	First Condition	Second Condition	Class Name	Translation Theory
1	$w_1 = w_2$	$sense_{w_1} = sense_{w2}$	A	Sense translation
2		$sense_{w_1} \neq sense_{w2}$	B	Literal translation
3	$w_1 \neq w_2$	Same meaning	C	Sense translation
4		Mismatch	F	-

Class A: Completely correct word sense and POS
Class B: Correct word but wrong sense
Class C: Different word but still has similar meanings
Class F: Absolutely incorrect

To test on the Evaluation process, the parallel corpus from two sources are used which have pair of sentence consisting of word วัด(*wat*) in Thai context and word *temple* in English context also. Both of them are English-Thai Parallel Concordance [12] and SEAlang Library Thai [13]. The relevant five sentences of Thai context and translation to English from each corpus are chosen. The evaluation is divided into two parts to disambiguate the word by using Cartesian product of all senses of the word from two sources. From the evaluation testing, it is found that the differences on the number of senses would affect the accuracy. Given a structure to represent word and sense format as Word#POS#Sense_number. Table II shows correct sense of the translations with AWN while Table III shows the result from two ambiguous senses with equivalent similarity value (same path length) with PWN.

TABLE II: THE RESULT OF MAXIMUM PATH SIMILARITY VALUE WITH AWN

Ambiguous word	Ambiguous sense	Reference Sense	Max value
church_service	church_service#n#1	ceremony#n#3	0.2
Monastery	monastery#n#1	crematory#n#1	0.1
Temple	temple#n#1	pagoda#n#1	0.5
Measure	measure#n#4	activity#n#1	0.5

TABLE III: THE RESULT OF MAXIMUM PATH SIMILARITY VALUE WITH PWN

Ambiguous word	Ambiguous sense	Reference Sense	Max value
church_service	church_service#n#1	activity#n#1	0.2
Monastery	monastery#n#1	crematory#n#1	0.1
Temple	temple#n#1	pagoda#n#1	0.5
Temple	temple#n#3	crematory#n#1	0.2
church_service	measure#v#1	populate#v#1	0.2
Monastery	quantify#v#2	call#v#9	0.2

Table IV shows the result from the experiment based on single process to measure the performances of each disambiguation techniques used in TH_WSD. At the beginning, ten contexts are tested with 330,069 pairs.

TABLE IV: THE TIME CONSUMING FOR EACH TECHNIQUE

Techniques	Pair	Time(sec)	Sec/pair
Path	39,456	209.18	0.0053
Vector	96,871	683.07	0.0071
Vector_pairs	96,871	1,117.38	0.0115
Lesk	96,871	983.20	0.0101

From our implementation, 250 testing data contexts are used (each with 30 words in average) generating a number of 4,525,369 pairs which require a lot of processing time for on the fly disambiguation. To minimize this problem, a database is prepared to store the results or the relatedness values from the calculations to the database for use later. However, those pre-calculated solutions influence on scalability because the context from open domain have individual word area.

III. IMPLEMENTATION

To implement the TH_WSD, most of the work is based on empirical study so the sources of bi-text corpora and input Thai contexts are significant. The inputs consist of sentences and contain exactly one disambiguated target. TH_WSD system is implemented as a web-based system; uses on-line web service to connect to the AWN and PWN; and creates offline lexical database of English language using in translation and sense explorer respectively.

The input sentences are generated from the English-Thai Parallel Concordance and SEAlang Library to mix between news and translated novel domain. There are 4 target words for inputs of วัด(wat), หัว(hua), เก็บ (kep) or เกาะ (koh). The total of 250 contexts consisting of individual meaning, part-of-speech and senses are used to test the TH_WSD framework so it can compare with the previous research [3], [4] on Thai words. Table V shows the input for Thai ambiguous words. The number of each word form, the sense number of each word and context number from each source are shown. The variations of Thai word form are prepared.

TABLE V: THE QUANTITY OF EACH TARGET WORD

Thai	Part of speech		Sense	Source	
Word	Noun	Verb	Number	SEAlang	Concordance
วัด *(wat)*	97	9	6	23	83
หัว *(hua)*	45	0	16	45	0
เก็บ *(kep)*	1	38	8	25	14
เกาะ *(koh)*	56	4	5	29	31

Fig. 4 displays the example of Sense Inventory in which the system explores the sense of each segmented word. The system verifies the existence of each word and translates. The system uses the AWN web service to check existence and retrieves their translation and variety of senses. The sense information of each word are stored in database.

Fig. 5 shows the output screen of Thai-English translation using AWN after the Sense Inventory.

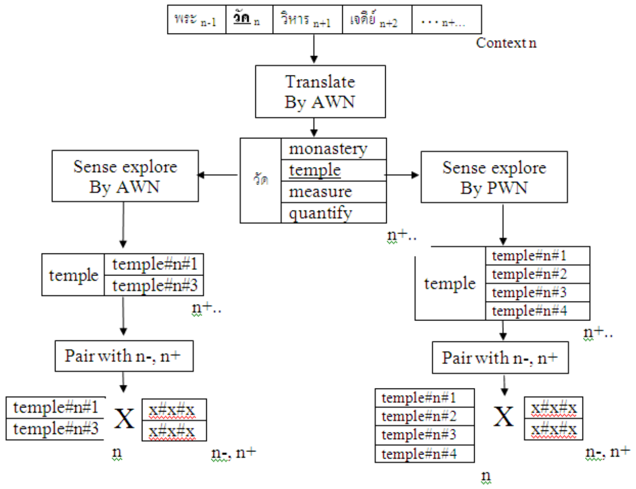
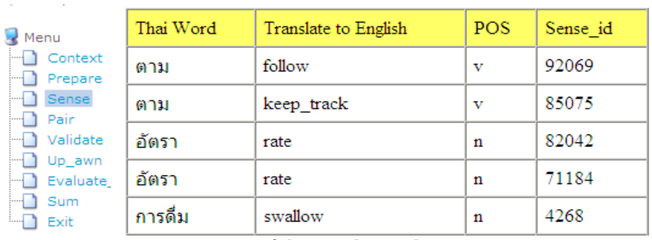


Fig. 4. An example of the sense inventory of Thai text.

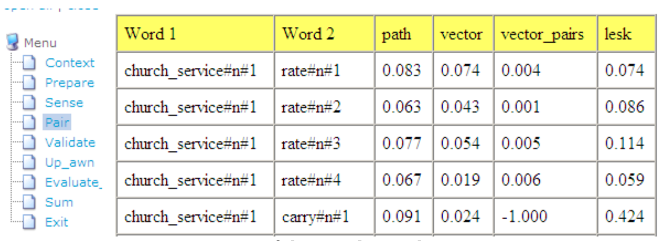


Thai Word	Translate to English	POS	Sense_id
ตาม	follow	v	92069
ตาม	keep_track	v	85075
อัตรา	rate	n	82042
อัตรา	rate	n	71184
การดื่ม	swallow	n	4268

Fig. 5. A screen of the translation that use AWN.

The ambiguous word selected from Thai word list and actual sense of English word will be ready to submit follow the sense list showed in synset format. After the Disambiguation module, the relatedness scores of each pair are then stored in the database.

Fig. 6 shows a screen of the output of the pair of word sense and PWN to compute the relatedness using by four disambiguation techniques.



Word 1	Word 2	path	vector	vector_pairs	lesk
church_service#n#1	rate#n#1	0.083	0.074	0.004	0.074
church_service#n#1	rate#n#2	0.063	0.043	0.001	0.086
church_service#n#1	rate#n#3	0.077	0.054	0.005	0.114
church_service#n#1	rate#n#4	0.067	0.019	0.006	0.059
church_service#n#1	carry#n#1	0.091	0.024	-1.000	0.424

Fig. 6. A screen of the translation that use PWN.

With our sample size, it is almost 5 million pairs to be evaluated. Thus, the classification rules play a significant role when it is classified earlier particularly in Class B that has correct word but wrong sense.

TABLE VI: THE FINAL REPORT OF EVALUATION RESULT

Techniques	A	B	C	F
Path	4	0	0	6
Vector	3	0	0	7
Vector_pairs	0	1	1	8
Lesk	3	0	0	7

Table VI shows an example of evaluation result from the ten Thai contexts during the testing phase of the implementation. It classifies the TH_WSD accuracy from the four WSD techniques. In contrast, the results of class A+B signify the high accuracy translation and A+B+C for baseline accuracy translation. More details can be found in [5].

IV. Experimental Results

There are several related modules in TH_WSD system. Each module is interconnected. For example, the context preparing process, its precision depends on segmentation techniques used to segment the context into words. Two sources are used. The SEAlang represents Library resources which contains complex scripts, while the Concordance is translations repository of department of Linguistics with translation from novel. The precisions are 80% and 76.43% for SEAlang and Concordance, respectively. In terms of sources, it is found that word segmentation for SEAlang provides better precision than Concordance and found that the main Concordance context contains more compound words and is full of pronoun.

In terms of sense extraction with AWN, the tool used to translate Thai word to English word and provide sense of each English word which was determined by the editor of Thai Computational Linguistics Laboratory manually using references from PWN. Subsequently, AWN has 73,660 of 117,659 senses [7] from PWN and that is the reason why Thai words using AWN for translation or extraction will result in less number of senses than using AWN for translation and extract sense with PWN as a result in Fig. 7.

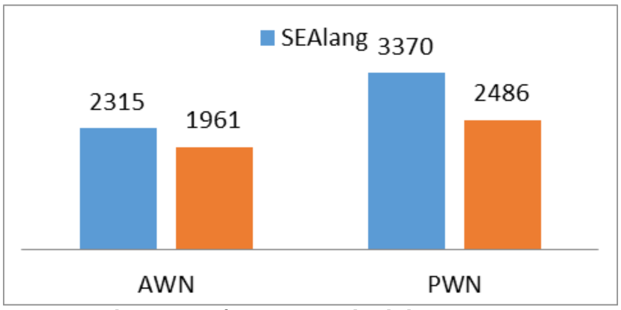


Fig. 7. The amount of sense extract divided into two resources.

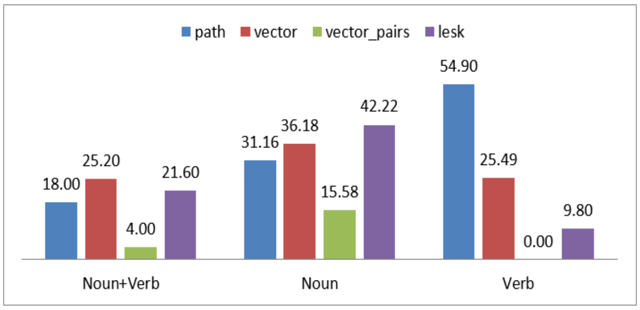


Fig. 8. The percentage of separate words form computation.

For word form aspect, an attempt to identify the factors that have influence on the precision of the framework is made. It is found that separating the input context with Part-of-speech to reduce the noise would increase the efficiency, particularly by feeding the right part-of-speech to the appropriate sense determination technique. Start with mix word forms, the vector technique provides the highest percentage for correct word and sense or Class A at only 25.2%. Then, using a specific word form, lesk technique provides the highest percentage for noun form at 42.22% while path technique reports 54.90% for verb form. Fig. 8 demonstrates the result from each technique group by Noun or Verb word form and both Noun and Verb word forms.

In PWN, the adjectives are not arranged in a hierarchical structure which prevents path based and information content measures from being applied. However, adjectives have glosses associated with their senses, so gloss based measures are useful. As a result in Fig. 9, observe that vector technique gives the highest percentage of 27.2 %. The result shows how few relations there are to and from adjectives.

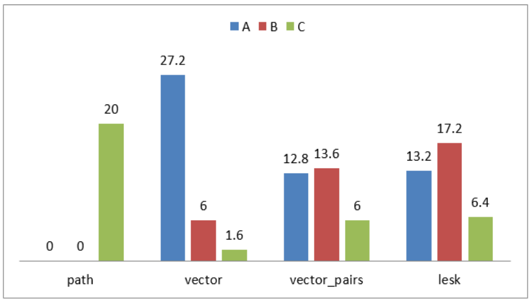


Fig. 9. The percentage of adjectives perform.

Considering on the window size factor, in our work the maximum word count in 250 contexts is 58. The line graphs below express comparison techniques on both sources for their combination (Class A+B). Fig. 10 shows that the highest percentage of AWN came from vector technique with almost 43%.

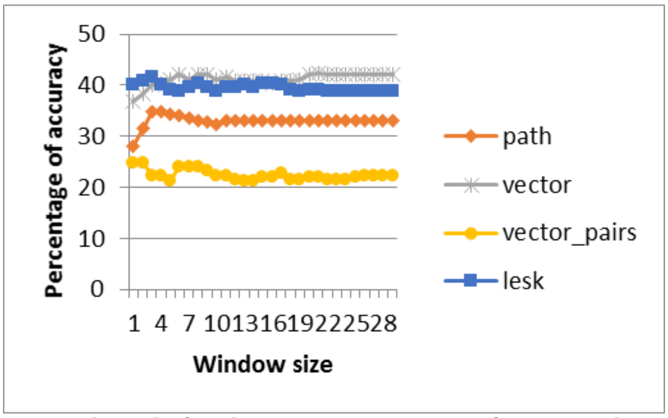


Fig. 10. The result of window size variation using sense from AWN (Class A+B).

Based on our WSD Accuracy Classification Rules, Table VII shows the accuracy results of one target word วัด(wat). The Thai word วัด(wat) appeared in 28 contexts from 106 contexts or 26.41%. The result shows that lesk technique generates the highest accuracy on the class A.

TABLE VII: The Accuracy for Word วัด(Wat)

วัด *(wat)*	A (%)	B (%)	A+B (%)	C (%)	A+B+C(%)
path	10(9.43)	7(6.60)	17(16.03)	1(0.94)	18(16.98)
vector	20(18.86)	8(7.547)	28(26.41)	14(13.20)	42(39.62)
Vector _pairs	4(3.77)	10(9.43)	14(13.20)	2(1.88)	16(15.09)
lesk	28(26.41)	4(3.774)	32(30.18)	18(16.98)	50(47.17)

Table VIII shows the accuracy in the word form or part-of-speech perspective. It illustrated that the word *measure* gave the highest accuracy for verb form while the word *island* which is in noun form provided the highest accuracy.

TABLE VIII: THE ACCURACY FOR EACH TARGET WORD

POS	Verb		Noun			
Word	keep	measure	head	island	monastery	temple
Total	39	9	45	56	44	53
Retrieved	15	7	25	37	5	13
Accuracy (%)	38.46	77.77	55.55	66.07	11.36	24.52

V. CONCLUSION

There are some interesting issues about TH_WSD framework worth to mention here. We have attempted to design a semi-automated TH_WSD for a non-specific domain WSD using cross-language knowledge sources and the initial implementation has shown an acceptable result. The Testing data with sufficient information to perform WSD across languages is used. However, in the empirical evaluation in NLP, the amount of testing data is an evident factor to identify the contribution of this work. Thus, 250 contexts which translated by bi-text corpora are used with some additional manual sense tagged by the linguist. Our framework is designed to resolve word sense disambiguation by mixing two digital knowledge sources and NLP techniques. Although the preparation process used manual correction as a semi-automate system, the core of system performs full automated disambiguation process. If the AWN has been developed further then TH_WSD will be beneficial in determining ambiguous sense on Thai to English sentence translation application automatically and accurately. In addition, TH_WSD has shown the potential of multilingual machine translation with WSD from the AWN and PWN.

REFERENCES

[1] R. Navigli, "Word sense disambiguation: A survey," *ACM Computing Surveys*, vol. 41, no. 2, pp. 14, February 2009.

[2] R. Mihalcea, C. Corley, and C. Strapparava, "Corpus-based and knowledge-based measures of text semantic similarity," in *Proc. American Association for Artificial Intelligence (AAAI 2006)*, Boston, July 2006, pp. 775-780.

[3] W. Kanokrattanukul, "Word sense disambiguation in Thai using decision list collocation," Master Thesis, Dept. Linguistics, Chulalongkorn Univ., 2001.

[4] S. Pongpinigpinyo and W. Rivepiboon, "Distributional semantics approach to Thai word sense disambiguation," *World Academy of Science, Engineering and Technology*, vol. 2, no. 6, pp. 1034-1035, 2008.

[5] P. Chongcharoen and J. L. Mitrpanont, "TH_WSD: A framework for Thai word sense disambiguation using knowledge-based approach," Master Thesis, Mahidol Univ., Bangkok, Thailand, 2013.

[6] S. Sudprasert and A. Kawtrakul, "Thai word segmentation based on global and local unsupervised learning," in *Proc. the National Computer Science and Engineering Conference (NCSEC 2003)*, Chonburi, Thailand, October 2003.

[7] S. Thoongsup, K. Robkop, C. Mokarat, T. Sinthurahat, T. Charoenporn, V. Sornlertlamvanich, and H. Isahara, "Thai wordnet construction," in *Proc. the 7th Workshop on Asian Language Resources*, ACL-IJCNLP 2009, Suntec, Singapore, August 2009, pp. 139-144.

[8] C. Fellbum, *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge, MA, 1998.

[9] T. Pedersen, S. Patwardhan, and J. Michelizzi, "WordNet: Similarity-measuring the relatedness of concepts," in *Proc. the Nineteenth National Conference on Artificial Intelligence (AAAI 2004)*, San Jose, CA, July 2004, pp. 1024-1025.

[10] M. Lesk, "Automatic sense disambiguation: how to tell a pine cone from an ice cream cone," in *Proc. 1986 SIGDOC Conference*, Association for Computing Machinery, New York, 1986, pp. 24-26.

[11] K. Abdalgader and A. Skabar, "Unsupervised similarity-based word sense disambiguation using context vectors and sentential word importance," *ACM Trans. Speech and Language Processing*, pp. 2-7, May 2012.

[12] W. Aroonmanakun, "English-Thai parallel concordance," Chulalongkorn Univ., Bangkok, Thailand, 2009.

[13] M. R. Haas, "SEAlang library," *The Mary Haas Thai Dictionary Project*, Univ. of California, Berkeley.

Jarernsri L. Mitrpanont was born in Bangkok, Thailand in 1958. She got her B.Sc. degree in physics in 1980, M.Sc. degree in applied mathematics in 1983 from Mahidol University, the Prestige University in Bangkok, Thailand and her Ph.D. degree in computer science from Oklahoma State University, Stillwater, Oklahoma, U.S.A. in 1993.

She is currently an associate professor and the dean of the Faculty of Information and Communication Technology (ICT), Mahidol University, Bangkok, Thailand. She has been with Mahidol University since 1980. She is the first dean who established and developed the ICT Faculty since 2009. She was previously the director of the Mahidol University Computing Center (MUCC) from 2003 to 2009. During that time, she also acted as the head of the Computer Science Department, Faculty of Science, Mahidol University. Her research interests are in the area of data and knowledge management, NLP, AI, decision support systems, and expert systems.

Assoc. Prof. Mitrpanont is a member of IEEE. She also serves as a committee of the Electronics Government Agency (EGA) and the Software Industry Promotion Agency (SIPA) of Thailand.

Premchai Chongcharoen was born on May 8, 1980 in Lopburi, Thailand. He received B.Sc degree in computer science from Suan Sunandha Rajabhat University, Thailand in 2001. Currently he is doing his M.Sc. degree in computer science in the Faculty of Information and Communication Technology at Mahidol University, Thailand. He has been working as a computer technical officer, professional level in Bangkok Metropolitan Administration, Thailand. His research interests are in natural language processing, word sense disambiguation and semantic technology.