Suggestion Mining and Knowledge Construction from Thai Television Program Reviews

Kanda Phawattanakul and Pramote Luenam

***Abstract*— In today's fast-changing business world, businesses must be able to understand customers' thoughts more effectively and efficiently in responding to customer's needs and expectations. The Internet allows people to share their thoughts, opinions and suggestions regarding purchased products or services in the form of online reviews. Especially, suggestions are valuable to businesses and are not supposed to be ignored. However, the enormous amount of information is mixed up with suggestion, facts, and opinions and unstructured text expression in the reviews make them difficult to be used by businesses. This study examines the characteristics of suggestions and proposes a suggestion mining framework for classifying suggestion sentences in customer reviews. Two main processes were employed in this study: (1) constructing knowledge based and (2) classifying each suggestion in the reviews as either a 'suggestion' or 'non-suggestion' sentence.**

The data used in the experiments are television watcher reviews obtained from several sources including the Thai Public Broadcasting Service (TPBS) website, facebook.com, and other related websites. These reviews are subjective opinions of television watchers regarding programs and broadcasting services. The data set consists of 2,561 sentences. In conjunction with linear SVM classification, we use a knowledge based approach that is a combination of keyword selection with specific Part-of-Speech, association wordlists, and domain specific wordlists. The SVM classifier has obtained precision, recall and F-Measure of 0.83, 0.94 and 0.88, respectively. Results show that overall accuracy of the SVM classifier is satisfactory.

***Index Terms*—Text mining, Suggestion mining, Customer feedback extraction**

I. INTRODUCTION

To increase competitiveness and success, businesses need to emphasize customer satisfaction. Internet nowadays is a more popular channel for exchanging information between customers and businesses. Customers are willing to share their thoughts and experiences on web forums, blogs, and social media websites. The increasing number of customer reviews, as well as unstructured text expression in the reviews, makes them very difficult to be used by businesses.

From the aforementioned problem, we propose using the computer-aided statistical text mining, in conjunction with natural language processing techniques to overcome the vast number of customer reviews in natural language. Most sentiment analysis generally relies on finding the polarity of an opinionated text, containing explicit polar expression. This so-called "explicit opinion mining" is expressed as positive, negative or neutral [1, 2, 3]. In other words, most existing mining systems can only find what people like or dislike on a given product and only if the opinion contains an explicit word, so that the opinion with no polar word, or suggestive opinion, is ignored. This means that they are unable to show useful suggestion sentences which could help businesses to improve their products or services. For instance, consider the following sentence: *"I think TPBS is a good TV channel but I wish its announcers should speak grammatically correct Thai"*. Traditional opinion mining systems can only tell that this sentence expresses a positive opinion since there exists the evidence of sentiment word "good" in the first clause but they ignore the important suggestion in the last clause. Even though, approximately 20-30% of reviews have been found to contain one or more suggestive sentences [4]. These suggestions are valuable and should not be ignored as they can help businesses to understand customer thoughts and concerns.

Our study aimed to classify the suggestions given by customers in natural language form. Therefore, some techniques need to be researched to extract valuable knowledge from text. Text mining is a fundamental technique which employs text classification (TC, also known as text categorization) as a major approach. Suggestion analysis uses text mining, that is so-called "suggestion mining" and is related to earlier studies [4, 5, 6]. They presented a simple and effective task of text classification using a knowledge based approach. However, we focused on a new method that seeks to address the new challenges. We used a machine learning approach for text classification (suggestion classification); In addition, we propose t knowledge based construction with reference to natural language processing (NLP) to improve text classification performance.

This paper is organized as follows. Section II discusses previous works and techniques related to suggestion mining. Section III overviews our approach and describes the processes of knowledge based construction, data preparation, feature selection, knowledge based tagging, text representation and classification. Section IV describes the data used in the experiments, presents the experimental results, and shows some examples of the learned suggestion reviews. Finally, section V concludes our findings and discussion for future studies.

Manuscript received Jan 04, 2013; revised Jan 24, 2013.
Kanda Phawattankul. Author is with National Institute of Development Administration, Thailand. (e-mail: kanda.pha@gmail.com).
Pramote Luenam. Author is with National Institute of Development Administration, Thailand. (e-mail: promote.l@ics.nida.ac.th).

II. RELATED WORK

Much research that relies on opinion reviews have focused on finding the polarity of individual customers. They tried to address mostly explicit polar words or distinguish between positive, negative, and neutral opinions in natural language [1, 2, 3, 7]. Some methods of polar word identification include manually tagged words by experts (e.g., “good”, “terrible”) and automatically tag words by statistics or co-occurring patterns of words.

For the suggestion mining area, the researchers proposed the manual tagging method for finding relevant keywords. Two objectives for suggestion mining are (1) extracting suggestions in customer reviews [5] and (2) extracting or concluding the suggestion phrases [4]. Our work and another study [5] had the same goal. We did not conclude the reviews. However, one study [5] used the knowledge engineering approach based on manually tagged keywords and created rules for suggestion extraction. Our work differs [5] since we used the Support Vector Machine (SVM), which is a supervised machine learning technique. Actually, many machine learning methods are used for text classification, such as Decision tree (DT), k-nearest neighbor (k-NN), Naïve Bayes (NB) and Support Vector Machine (SVM). One study [8], they compared the results of text classification between NB and SVM. They found that SVM substantially outperforms NB. SVM is more suitable because it behaves robustly for different learning tasks. One remarkable property of SVM is the ability to learn can be independent of the dimensionality of the featured space. This means that we can classify text even in the presence of very many features [9, 10]. Also, it can efficiently perform non-linear classification using what is called the kernel functions.

One study [11] proposed a framework for detecting opinions in Thai news columns. Their approach is the classification of subjective (opinion) sentences from objective (fact) sentences and achieves the best result using Naïve bayes. They discussed that the performance can be improved further with various features such as POS, special symbols, and prior knowledge based items such as name entities etc., not only words.

Thus, our classifier uses various features by prior knowledge bases: (1) domain dependency (DD), (2) specific part-of-speech (POS), (3) association word lists (AW), using association rules mining technique to find relevant keywords within a corpus, and (4) domain specific word lists (DW) with finding high frequency keywords from external sources.

The common classifiers cannot directly process a text document in its original form. The document has to be transformed into computational structured data, which is a document representation or text representation. It is so-called the “Vector Space Model” (VSM). Bag of words is one of the basic and popular text representation methods. All words appearing in documents are represented in binary terms; ‘1’ if occur and ‘0’ otherwise, or term frequency (TF).

However, bag of words is not an appropriate term; some terms are not useful for text classification. Therefore, irrelevant terms need to be reduced, i.e., feature selection. Many techniques are proposed in the literature [9]. Examples include information gain (IG), mutual information (MI), chi-square test (CHI) and document frequency (IDF). These methods were evaluated by Yang [12] and found that IG and CHI rank highest in terms of their effectiveness. They have also found that DF is one of the simplest methods whose performance is close to IG and CHI but has the lowest cost in computation. However, a conventional approach and the most widely used in text representation techniques is term frequency with inverse document frequency (TF-IDF). TF-IDF is computed by term frequency (TF) that occurs frequently in documents and inverse document frequency (IDF) that the terms which occur in many documents are less important for classification.

The other popular method applied is association mining for finding relevant keywords [2]. We apply association mining for finding suggestion phrases then transform text into structured form by TF-IDF approach.

III. METHODOLOGIES

Our proposed method is to mine suggestions from customer reviews. The mining result is a class of review sentences which are classified as either 'suggestion' or 'non-suggestion'. Generally, customer reviews are defined in three types: (1) facts are something that have really occurred and can be proven, (2) opinions contain explicit polar word expressions such as good, excellent, or terrible and (3) suggestions which are implicit opinions (no explicit polar words). In this study, we focus on suggestions.

A suggestion is a psychological process by which one person guides the thoughts, feelings or behavior of another [13]. Consider the following sentences: (1) *“I really dislike the TV announcer”, and (2) “Is it possible to change the announcer?”* Clearly, the first one, which contains the explicit polar word 'dislike', is direct opinion. Existing opinion mining techniques are able to identify this kind of opinion. The latter is a suggestion to make an adjustment of a TV program. In our approach, suggestion sentences are extracted from the customer reviews by using suggestion word indicators (such as อยาก (want), ควร (should)), as well as terms or phrases that are the guidance on actions or recommendations (such as “เปลี่ยนพิธีกร” (*change the announcer*)) as in the example above.

An application program has been developed to classify suggestions from the other types of customer reviews. The application performs five major tasks: (1) Knowledge based construction, (2) data preparation - the process of cleaning data, correcting misspellings and synonym words grouping, (3) feature selection - the method of removing some irrelevant words, (4) text representation - the task of tagging relevant keywords using words from the knowledge based then converting text into the computational structured form, and (5) suggestion classification. The overview of our suggestion mining process is shown in Fig 1.

A. Knowledge based construction

Words are important pieces for text analysis. The larger amount of words and relevant information collected in the knowledge based, the more efficient the text analysis will be.

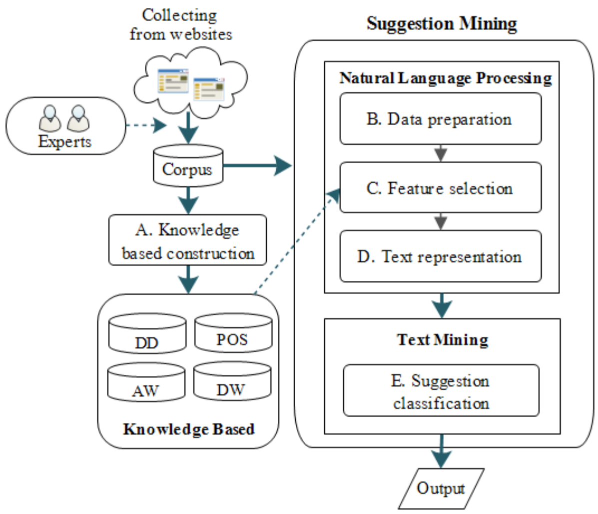

Fig. 1. The architectural overview of the suggestion mining process.

Thus, we propose the method for constructing and using the knowledge based for suggestion classification. The knowledge based consists of domain-dependency (DD), Part-of-Speech (POS) and frequent suggestion action phrase using association mining rules, so called association wordslists (AW), and, frequent noun and verb of domain retrieved from external sources, so-called domain wordlists (DW).

Domain-dependency (DD)

The knowledge based contains suggestion words and name entities which are defined by experts. The suggestion words are classified into four types: (1) explicit suggestion words group *(S_e)*, e.g., ขอ (ask), อยาก (want); (2) suggestion by asking *(S_a)*, e.g., น่าจะ (should), กรุณา, รบกวน (may, might, can, and could); (3) query suggestion *(S_q)*, e.g., ทำไม, ได้ไหม (why); and (4) condition suggestion *(S_c)*, e.g., ถ้า, หาก (if). Name entities are specific names such as people's, organization or building names. In this study, four types of the name entities are defined: (1) the name of the TV channels (OBJ0); (2) the type of TV programs (OBJ1), e.g., news, documentary; (3) the names of TV programs (OBJ2); and (4) the names of channel staffs (OBJ3).

Part-of-Speech (POS)

POS tagging provides a useful way to better understand the meaning of word. In this study, we use Lexitron -- the Thai text corpus provided by NECTEC [14] for tagging all

types of word except verbs because in suggestion mining, the verb is crucial. A suggestion is usually represented in action verbs. The Lexitron alone is not enough for suggestion action identification, as it cannot distinguish between action verbs and other type of verbs. We tag all words that refer to verbs with Thai Orchid Corpus [15]. In the Corpus, the Part-of-Speech, especially for verb, is defined in more specific ways. Verbs are classified into three subcategories: (1) action verb (VACT) indicates an action of entity, e.g., ทำงาน (work), ร้องเพลง (sing), and กิน (eat), (2) stative verb (VSTA) refers to a state and is not used in progressive aspect, e.g., เห็น (see), รู้ (know), and เป็น (be), and (3) attribute verb (VATT). that refers to an attribute of verbs, e.g., อ้วน (fat), ดี (good), สวย (beautiful). Obviously, the stative and attribute verbs do not refer to action activities, so suggestion phrases are usually represented by action verbs. After tagging, the action verbs that refer to suggestion phrases can be separated from any other type of verb.

Association wordlists (AW)

Our approach applied association mining rules in data mining to identify co-occurrences with high frequency (receiving many recommendations by many people) of nouns and verbs or a couple verbs are likely to be action phrases. Consider the following sentence: "*ขอให้เปลี่ยน พิธีกร*" *(I am asking for a change of the announcer)*, where "*พิธีกร*" *(the announcer)* is the noun, and "*เปลี่ยน*" *(change)* is the verb. We assume that the occurrence of a noun and verb pair is likely to be a suggestion action phrase.

We use the association technique for finding frequent suggestion action phrases with the minimum support of 1%, the same as in an earlier study [2]. Note that if the minimum support is too high, some relevant word pairs may be lost.

We call the frequent occurrence of pairs of words association wordlists (AW). The pairs of words in AW refer to nouns tagged with "*AW-N*" and the others refer to verbs tagged with "*AW-VACT*". All results are added to the association wordlists knowledge based. The association wordlists will be pruned, if the distance between a pair of words is greater than *k* (*k's* value is between 3 and 6). Assume that if the distance is greater than *k*, a pair of words is not likely to be relative or refer to the same aspect.

Domain wordlists (DW)

We assume that nouns and verbs which are frequently occurring together within the same domain (TV broadcasting providers) are likely to be keywords. Thus, we clawed external sources (e.g., Wikipedia) to find frequent nouns and verbs. This process has many benefits. First, we do not need a large suggestion corpus for constructing a knowledge based by experts. Second, we can get new words that are related to the experimental domain. However, all terms in external sources include irrelevant terms or noisy terms. We need to be careful when selecting keywords from frequent nouns and verbs. If the value of frequently occurring terms is too high, we may lose other relevant terms. If the value is too low, we may get noisy terms [9]. As a result, we need to choose appropriate frequent nouns and verbs which do not depreciate the performance of classification results. In our method, we used the top 100 most frequently occurring words (after removing stop words) because the top 100 have a word frequency more than 1% of all words, and it can be implied that high-frequency terms convey important information about the semantic domain of a corpus. We call the selected word in

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

this step as domain wordlists (DW). Words represented by nouns are tagged with *"DW-N"* and those represented by verbs are tagged with *"DW-VACT"*. Then all DW are added to the domain wordlists knowledge based.

B. Data Preparation

Each review is broken up into a sequence of sentences and token words using JavaScript API BreakIterator. The data preparation processes include cleaning, misspelling correction, and grouping synonymous words. We assume that each document consists of reviews. Each document *d* contains the sequence of sentences *s*. The document can be represented as $d = (s_1, s_2, s_3 \ldots s_n)$, where *n* is the amount of sentences in *d*. Each sentence *s* consists of a sequence of words w_{ij}. The sentence can be represented as, $s_i = (w_{i1}, w_{i2}, w_{i3}, .. w_{im})$, where *m* is the amount of selected keywords.

C. Feature selection

Some words are common words which can be frequently found in documents but are not useful for text classification. They can be discarded with no harm to the classifier performance and may even result in its improvement [16]. Most text classification studies at least remove the stop words. POS of the stop words are discarded in our experiment including ending words (e.g., ค่ะ ครับ), conjunction (e.g., และ (and), หรือ (or)), preposition (e.g., บน (on), ของ (of)), classifier (used in Thai only, e.g., ตัว อัน), determiner (e.g., ทั้งหมด (all), นี้ (this)), interjection (e.g., โอ้ย (ouch)), negator (e.g., ไม่ (not), pronoun (e.g., คุณ (you)), number, and symbol.

D. Text Representation

This process includes two tasks: knowledge based tagging and text representation.

Knowledge based tagging

In this step, we add some more relevant keywords into the knowledge based. Five different experiments of tagging words in the knowledge based are carried out. The experiments are as follows:

(I) Analysis from words

(II) Analysis from words with Part-of-Speech (POS) tagging

In this experiment, we assign the Part-of-Speech (POS) tag to each word in a sentence. A word can also rely on many different tags with many meanings. In the experiment, suggestion words, including the four types of suggestions and the name entities (e.g., "S_e", "*OBJ1*"), are tagged.

(III) Analysis from words with specific POS tagging

We assume that suggestion phrases are usually represented in action phrases. Therefore, the word that is categorized as an action verb in the Thai Orchid Corpus is tagged with *"VACT"*.

(IV) Analysis from words with POS, specific POS and AW tagging

AW from the association wordlists is used for the third step tagging. This is the process of tagging a pair of noun (n_i) and verb (v_i) that appears together in the same sentences and the distance between the pair is not greater than *k* because if the distance is more than *k*, it is not likely to be relative or refer to the same aspect. The *k's* value is set between 3 and 6 for examine and find an appropriate *k* value.

(V) Analysis from words with specific POS, AW and DW tagging

In this experiment, DW from the domain wordlists knowledge based such as subjects, people's named etc. are tagged to each word.

Text representation

Typically, the classifiers cannot directly process the text document in the original form. Therefore, prior to performing the classification, texts are converted into a more manageable representation. The texts are represented by feature vectors that calculate term frequency with inverse document frequency (TF-IDF).

E. Suggestion Classification

The objective of this process is to classify a class of customer reviews as 'suggestion' or 'non-suggestion'. To do so, we construct a suggestion classification model from a training set and use the model to classify a test set.

To find the best methodology for suggestion classification, the performance results of the experiments -- including Precision and Recall are compared and evaluated.

IV. EXPERIMENTS

The data used in the experiments are television watcher reviews obtained from several sources including the Thai Public Broadcasting Service (TPBS) website, facebook.com, and some other related websites. These reviews are subjective opinions of television watchers regarding programs and broadcasting services. The data set consists of 2,561 sentences which are manually divided into two groups: (1) 1,757 opinion sentences, and (2) 804 suggestion sentences

The data set is randomly partitioned into two parts: a training and a test set. Subsequently, 814 sentences are assigned as the training set and 815 records are assigned as the test set. These data go through learning algorithms with 10 fold-cross validations.

In our approach, three knowledge repositories have been constructed: (1) domain-dependency (DD), (2) part-of-speech (POS), (3) association wordlist (AW), and (4) domain wordlist (DW).

The first step of suggestion mining framework is the data preparation process which includes cleaning, correcting misspellings, and synonym words grouping. Then the stop words are dropped in the feature selection process and the knowledge based is used for tagging the relevant words. Examples results of tagging processes are as follows:

Examples suggestion sentences as follows:

s_1 = "*น่าจะออกอากาศใหม่เป็นเวลาหัวค่ำนะค่ะ*

(It should have been rerun in evening.)

s_2 = "*ผมว่าข่าวหัวค่ำน่าจะเปลี่ยนพิธีกรข่าวใหม่*"

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

(I think the evening news should change the TV announcer.)

s_3 = "*สถานีควรเปลี่ยนพิธีกรข่าวหัวค่ำ*"

(The channel could have changed the TV announcer.)

s_4 = "*ชอบพิธีกรคนใหม่จัง*"

(I like the new TV announcer.)

Examples results of tagging processes in method-V are as follows:

s_1 = *"น่าจะ (should have been)<S_a>|ออกอากาศ (rerun)<AW-VACT>|ใหม่ (again)<ADV>| เป็น (be)<VSTA>| เวลา (time)<AW-N>| หัวค่ำ (evening)<N>| นะ <END>| ค่ะ <END>"*

s_2 = *"ผม (I) <PRON> | ว่า (that)<AUX> | ข่าว (news) <OBJ1> | หัวค่ำ (evening) <N> | น่าจะ (should) <S_a> | เปลี่ยน (change) <AW-VACT> | พิธีกร (TV announcer) <AW-N> | ข่าว (news) <OBJ1> | ใหม่ (newly)<ADV>"*

s_3 = *"สถานี (channel) <DW-N> | ควร (could) <S_a> | เปลี่ยน (change) <AW-VACT> | พิธีกร (TV announcer) <AW-N> | ข่าว (news) <OBJ1> | หัวค่ำ (evening) <N> |"*

s_4 = *"ชอบ (like) <VACT> | พิธีกร (TV announcer) <N> | คน (man) <N> | ใหม่ (new) <ADJ>| จัง <END>"*

For preparing input data to VSM, all words and tags are transformed into computational structured data. Consequently, each sentence is converted into two feature vectors: (1) a weights vector of selected word and (2) a weights vector of DD, POS, AW and DW. The feature representations of the VSM, represented by a matrix, are shown in Table I.

TABLE I
THE EXAMPLE OF TEXT REPRESENTATIONS OF THE VECTOR SPACE MODEL

TF-IDF	s_1	s_2	s_3	s_4
ข่าว(news)	*0*	*0.78*	*0.31*	*0*
ควร(could)	*0*	*0*	*0.62*	*0*
ชอบ (like)	*0*	*0*	*0*	*0.95*
น่าจะ(should)	*0. 23*	*0.39*	*0*	*0*
พิธีกร(announcer)	*0*	*0.16*	*0.12*	*0.19*
สถานี(channel)	*0*	*0*	*0.62*	*0*
หัวค่ำ(evening)	*0.13*	*0.16*	*0.12*	*0*
ออกอากาศ(rerun)	*0.65*	*0*	*0*	*0*
เปลี่ยน(change)	*0*	*0.39*	*0.31*	*0*
เวลา(time)	*0.65*	*0*	*0*	*0*
ใหม่(newly,again)	*0.13*	*0.16*	*0.19*	*0*
ADJ	*0*	*0*	*0*	*0.70*
ADV	*0.18*	*0.42*	*0*	*0*
AW-N	*0.33*	*0.17*	*0.17*	*0*
AW-VACT	*0.33*	*0.17*	*0.17*	*0*
DW-N	*0*	*0*	*0.85*	*0*
N	*0*	*0*	*0*	*0*
OBJ1	*0*	*0.85*	*0.42*	*0*
S_a	*0.33*	*0.17*	*0.17*	*0*
VACT	*0*	*0*	*0*	*0.7*

For generating the SVM classifier, Rapidminer, an open source system for data mining, is used. In the classifier training process, the model is trained using a linear kernel. Various types of input features (method I-V) are examined and compared in order to find the best results for text classification. In details, five different experiments are carried out: (I) words only, (II) words with POS, (III) words with specific POS, (IV) words with POS, specific POS and AW, and (V) words with POS, specific POS, AW and DW.

For evaluation of the model, results from the classifier are compared with those obtained by experts. In this step, the reviews are manually processed sentence by sentence. Each sentence is then manually labeled as either 'suggestion' or 'non-suggestion'. Results, as illustrated in Table II, show that experiment-IV with k=3 was the best among all experiments.

TABLE II
THE PRECISION, RECALL AND F-MEASURE AT EACH EXPERIMENT OF SUGGESTION MINING.

Experiment	Suggestion mining		
	Recall	Precision	F-measure
1.words	77.98%	86.73%	82.12%
2.words with POS tagging	81.51%	93.61%	87.14%
3.words with spec. POS tagging	82.50%	92.84%	87.36%
4.words with spec. POS and AW tagging			
k = 3	83.41%	93.86%	88.33%
k = 4	83.22%	93.86%	88.22%
k = 5	83.14%	93.35%	87.95%
k = 6	83.26%	92.84%	87.79%
5.words with spec. POS, AW, DW tagging	83.33%	92.07%	87.48%

V. CONCLUSION

In this paper, we propose the method, based on data mining and natural language processing techniques, to extract suggestion sentences from television watcher reviews. Experimental results show that our proposed method had good performance for disambiguation of suggestion sentences. We believe that our method can seek out people's thoughts and feelings. Businesses may use these important pieces of information to increase their competitiveness and market success.

One major limitation of our study is due to inadequate information in both Thai text corpus and Thai WordNet as they are in early stages of development. Consequently, we can analyze the reviews only at the syntactic analysis level. In our future work, we will further focus on dealing with ambiguities and vagueness appearing in natural language. In order to do so effectively, more human-like natural language processing is required, at least at the semantic level.

REFERENCES

[1] P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Review," *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics,* Philadelphia, 2002, pp. 417-424.

[2] B. Liu, "Mining Opinion Features in Customer Review," *American Association for Artificial Intelligence,* 2004, pp. 755-760.

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[3] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” *Foundation and Trends in Information Retrieval*, Vol.2 Nos.1-2, 2008, pp. 1-135.

[4] A. Viswanathan, P. Venkatesh, B. Vasudevan, R. Balakrishnan and L. Shastri, “Suggestion Mining from Customer Reviews,” *Proceedings of the 17th Americas Conference on Information Systems*, Michigan, 2011, pp. 1-9.

[5] J. Vishwanath, S. Aishwarya. “User Suggestion Extraction from customer Reviews,” *International Journal on Computer Science and Engineering*, 2011, pp. 1203-1206.

[6] Z. Jingbo and Y. Tianshun, “A Knowledge-based Approach to Text Classification”, *Proceedings of the first SIGHAN workshop on Chinese language processing*, Association for Computational Linguistics Stroudsburg, PA, USA, 2002, pp.1-5.

[7] K. Alisa, A. Niran, S Cathchwaj, Pornpimon,Chartchw and H Choochart, “”,2010, National Electronics and Computer Technology Center (NECTEC)

[8] M. Rennie and R. Rifkin, “Improving Multiclass Text Classification with the Support Vector Machine”, 2002, pp.1-14.

[9] A. Nuntiyagul, “Text Categorization & Retrieval for Thai Item Bank using Patterned Keyword in phrase(PKIP)”, *Mahidol University*, 2006.

[10] T. Joachims, “Text Categorization with Support Vector Machine: Learning with Many Relevant Features”, *In European Conference on Machine Learning (ECML)*, 1998

[11] K. Sukhum, S Nitsuwat, “Opinion Detection in Thai Political News Columns Based on Subjectivity Analysis”, *International Conference on Computing and Information Technology*, 2011, pp. 27-31.

[12] O. Pedersen, Y Yang, “A Comparative Study on Feature Selection in Text Categorization”, *Proceedings of the Fourteenth International Conference on Machine Learning*, 1997, pp. 412-420.

[13] Meaning of suggestion. [Online]. Available: http://en.wikipedia.org/wiki/Suggestion, access 3 Sep 2012

[14] Lexitron, The National Electronics and Computer Technology Center (NECTEC). [Online]. Available: http://lexitron.nectec.or.th.

[15] http://culturelab.in.th/files/orchid.html, The National Electronics and Computer Technology Center (NECTEC).

[16] *The Text Mining Handbook*, R. Feldman and J. Sanger, Cambridge University Press, United States of America, 2007.