Thai News Article Sentiment Classification based on User Comments on Online Social Media

Nut Kunpattanasopon

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Technology

Prince of Songkla University

2018

Copyright of Prince of Songkla University

Thai News Article Sentiment Classification based on User Comments on Online Social Media

Nut Kunpattanasopon

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Technology

Prince of Songkla University

2018

Copyright of Prince of Songkla University

Thesis Title	Thai News Article Sentiment Classification based on User Comments on Online Social Media
Author	Mr. Nut Kunpattanasopon
Major Program	Information Technology

Major Advisor

.....

(Dr. Kiyota Hashimoto)

Co-advisor

.....

(Dr. Nattapong Tongtep)

Examining Committee

..... Chairperson

(Dr. Kanjana Thongglin)

...…..... Committee

(Dr. Putthiporn Thanathamathee)

...…..... Committee

(Dr. Nattapong Tongtep)

...…..... Committee

(Dr. Kiyota Hashimoto)

The Graduate School, Prince of Songkla University, has approved this thesis as partial fulfillment of the requirements for the Master of Science Degree in Information Technology

……

(Professor Dr. Damrongsak Faroongsarng)

Dean of Graduate School

This is to certify that the work here submitted is the result of the candidate's own investigations. Due acknowledgement has been made of any assistance received.

.....Signature

(Dr. Kiyota Hashimoto)

Major Advisor

.....Signature

(Mr. Nut Kunpattanasopon)

Candidate

I hereby certify that this work has not been accepted in substance for any degree, and is not being currently submitted in candidature for any degree.

.....Signature

(Mr. Nut Kunpattanasopon)

Candidate

Thesis Title Thai News Article Sentiment Classification based on User Comments on Online Social Media

Author Mr. Nut Kunpattanasopon

Major Program Information Technology

Academic 2017

ABSTRACT

Social media have become part of our daily life and we read other people's writing including news articles and various user comments and write our own comments. Although most social media services provide a convenient way, LIKE buttons, to express a reactive feeling towards a content, its use does not precisely reflect users' feelings: users often use the button for "LIKE" even when they feel angry, sad, etc. Thus sentiment analysis of users' comments and the recommendation of contents to be presented to each user based on users' comment sentiment analysis is an important issue to improve those services. This thesis proposes a method for this purpose in the Thai language. The Thai language social media text still has difficulties in processing: difficulty in word segmentation with many spelling and other variations included, no publicly available sentiment dictionary, etc. Thus this thesis investigates a systematic noise reduction for social media text using both heuristic rule-based and Conditional Random Fields-based machine learning approaches. Based on this noise reduction method, sentiment dictionary construction based on the data is conducted by comparing different methods. Based on the constructed sentiment dictionary, the sentiment estimation of user comments to news articles is performed. This estimation is employed as labeled sentiment value data for corresponding news articles for each user. Finally news article classification, which is the foundation to recommend future news articles on the basis of each user's preference, is conducted. The final step employs Support Vector Machines as a machine learning method, and sentiment classifiers for each user are constructed. The proposed noise reduction, sentiment dictionary construction, and news article sentiment classification are evaluated with comparative experiments. The result demonstrates that the proposed methods successfully improve each process and the final classification.

Keywords: Sentiment Analysis, Sentiment Dictionary, Text Processing, Machine Learning, Natural Language Processing, Classification Algorithm, Social Media, Thai Language

ชื่อวิทยานิพนธ์ การจำแนกความรู้สึกของบทความข่าวภาษาไทยด้วยความคิดเห็นของผู้ใช้บนสื่อสังคมออนไลน์

ผู้เขียน นายณัฐ กุลพัฒนโสภณ

สาขาวิชา เทคโนโลยีสารสนเทศ

ปีการศึกษา 2560

บทคัดย่อ

ปัจจุบันสื่อสังคมออนไลน์เป็นส่วนหนึ่งในชีวิตประจำวันของเรา แม้ว่าสื่อสังคมออนไลน์ส่วนใหญ่จะให้ทางเลือกในการแสดงความรู้สึกต่อบทความด้วยการกด “LIKE” แต่วิธีดังกล่าวไม่ได้แสดงถึงความรู้สึกที่แท้จริงของผู้ใช้งาน ผู้ใช้งานมักจะกด “LIKE” แม้ว่าจะรู้สึกโกรธหรือเศร้าก็ตาม ดังนั้นการวิเคราะห์ความรู้สึกของผู้ใช้งาน ตลอดจนการนำเสนอเนื้อหาที่เหมาะสมต่อผู้ใช้งานโดยอ้างอิงจากการวิเคราะห์ความเชื่อมั่นจากความคิดเห็นของผู้ใช้งานเองจึงเป็นประเด็นสำคัญ วิทยานิพนธ์นี้ได้เสนอวิธีการแก้ปัญหาเหล่านี้กับข้อความภาษาไทย อย่างไรก็ตาม ข้อความสื่อสังคมออนไลน์ภาษาไทยยังคงมีปัญหาในการประมวลผลอยู่มาก ทั้งการแบ่งกลุ่มคำที่มีการสะกดคำและรูปแบบอื่น ๆ รวมถึงยังไม่มีพจนานุกรมความรู้สึกภาษาไทยที่เหมาะสมกับชุดข้อมูล ดังนั้น วิทยานิพนธ์นี้จึงศึกษาวิธีการลดสิ่งรบกวนสำหรับข้อความสื่อสังคมออนไลน์ โดยใช้วิธีการเรียนรู้แบบฮิวริสติก และ การเรียนรู้ของเครื่องแบบคอนดิชันนอลแรนดอมฟิลด์ (Conditional Random Fields) โดยสร้างพจนานุกรมความรู้สึกจากชุดข้อมูล จากนั้นประมาณค่าความเชื่อมั่นความคิดเห็นของผู้ใช้งานด้วยพจนานุกรมความรู้สึกที่สร้างขึ้นซึ่งใช้เป็นข้อมูลค่าความรู้สึกต่อบทความข่าวของแต่ละผู้ใช้งานเอง และจำแนกบทความข่าวสำหรับนำเสนอให้แก่ผู้ใช้งาน ด้วยการพิจารณาจากความชอบของแต่ละผู้ใช้งานโดยใช้อัลกอริทึม ซัพพอร์ทเวกเตอร์แมชชีน (Support vector machine) ในการจำแนก และเปรียบเทียบประสิทธิภาพของวิธีการ การลดสิ่งรบกวนสำหรับข้อความสื่อสังคมออนไลน์ วิธีการสร้างพจนานุกรมความรู้สึก และ การจำแนกความรู้สึกของบทความข่าว ผลการศึกษาแสดงให้เห็นว่า วิธีที่นำเสนอสามารถปรับปรุงประสิทธิภาพในแต่ละขั้นตอนได้

คำสำคัญ: การวิเคราะห์ความรู้สึก, พจนานุกรมความรู้สึก, การประมวลผลข้อความ, การเรียนรู้ของเครื่อง, การประมวลผลภาษาธรรมชาติ, อัลกอริทึมจำแนก, สังคมออนไลน์, ภาษาไทย

ACKNOWLEDGEMENT

I would Like to thank the Master of Science in Information Technology, College of Computing, Prince of Songkla University, Phuket Campus for educational opportunities and research scholarship.

In particular, I would like to express my sincere gratitude to my major advisor Prof. Dr. Kiyota Hashimoto and co-advisor Dr. Nattapong Tongtep for the continuous support of my master study and research, for his patience, immense knowledge and for guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thank the rest of thesis committee: Dr. Kanjana Thongglin and Dr. Putthiporn Thanathamathee, for their encouragement, insightful comments, question and recommendations.

Finally, I would like to thank my family, my fellow lab mates in Information Technology and all staffs for providing me with unfailing support and continuous encouragement throughout my year of study and through the process of researching and writing this thesis.

Nut Kunpattanasopon

Contents

	Page
Abstract	(5)
Abstract (Thai)	(7)
Acknowledgment	(8)
Contents	(9)
List of Tables	(13)
Table of Figures	(14)
CHAPTER 1 Introduction	1
1.1 Statement of Problem	1
1.2 Research Problems	4
1.3 Objectives	4
1.4 Research Question	5
1.5 Scopes of Research	5
1.6 Expected Outcomes	6
CHAPTER 2 Literature Review	7
2.1 Introduction	7
2.2 Thai Language	8
2.3 Thai Character Clusters	10
2.4 Natural Language in Thai Social Media	11
2.4.1 Characteristic of Social Media Texts	12
2.4.2 Characteristic of Thai Social Media Texts and how they have been treated	13
2.5 Noise Reduction	13
2.5.1 Heuristic Approach	13
2.5.2 Machine Learning Approach	14
2.6 Sentiment Analysis	14
2.7 Sentiment Dictionary Construction	15

Contents (cont.)

	Page
2.8 TF-IDF	18
2.8.1 TF	18
2.8.2 IDF	18
2.9 Classification Algorithm	20
2.9.1 Decision Tree	21
2.9.2 Naïve-Bayes	23
2.9.3 Support Vector Machine (SVM)	24
2.9.4 Conditional Random Fields	25
2.10 Recommendation	26
CHAPTER 3 Overview of the Proposed Methodology	28
3.1 Introduction	28
3.2 The Overview Framework	28
3.2.1 Data Collection	29
3.2.1.1 User Comment Collection Data	31
3.2.1.2 News Article Collection Data	32
3.2.2 Text Normalization	34
3.2.3 Sentiment Dictionary Construction	36
3.2.3.1 SO-PMI-IR algorithm with “AND” operation	37
3.2.3.2 SO-PMI-IR algorithm with “NEAR” operation	38
3.2.4 Sentiment Estimation	38
3.2.5 Sentiment Classification	40
3.3 Conclusion	41
CHAPTER 4 Sentiment Value Estimation of User Reviews	42
4.1 Introduction	42
4.2 Text Normalization	44
4.2.1 Heuristic Approach	44
4.2.1.1 Punctuation Removal	45

Contents (cont.)

	Page
4.2.1.2 Emotion and Special Character Removal	46
4.2.1.3 Repetition Removal	47
4.2.1.4 Word Segmentation	49
4.2.1.5 Word Replacement	49
4.2.2 Noise Validation	50
4.2.3 Character Based CRFs Approach	52
4.2.4 Evaluation of Character-Based CRFs Method	55
4.2.5 Evaluation with Sentiment Dictionary Construction	56
4.2.5.1 Sentiment Dictionary Construction	57
4.2.5.2 Experiment of Noise Reduction Technique	58
4.2.5.3 Result and Discussion	58
4.3 Constructed Sentiment Dictionary with Larger Data	59
4.4 Sentiment Value Estimation of User Comment	61
4.5 Conclusions	62
CHAPTER 5 Machine Learning Classification of News Articles Based on User Comment Sentiment Value	63
5.1 Introduction	63
5.2 Experiment Setting	64
5.3 Feature Extraction	66
5.4 Feature Value Calculation	67
5.4.1 TF-IDF	67
5.4.2 Public Sentiment Dictionary Feature Vector	68
5.4.3 News Article Sentiment Dictionary	68
5.4.3.1 News Article Sentiment Dictionary with “AND” Feature Vector	68
5.4.3.2 News Article Sentiment Dictionary with “NEAR” Feature Vector	69
5.5 Classification with Machine Learning	70

Contents (cont.)

	Page
5.6 Evaluation of Machine Learning Classification	71
5.6.1 Comparison of User Comment Sentiment Dictionary Creation Technique	71
5.6.2 Comparison of Feature Vector	77
5.7 Conclusion	83
CHAPTER 6 Conclusion	84
6.1 Summary	84
6.2 Future Tasks to Come	85
References	87
Appendices	101
Appendix A	102
Appendix B	111
Appendix C	132
Appendix D	134
Appendix E	167
Appendix F	187
Vitae	202

List of Table

Tables	Page
Table 2.1 Type of Character Cluster	11
Table 3.1 List of News Providers	30
Table 3.2 Summarized Data	33
Table 3.3 Type of Noise in Social Media Texts	34
Table 3.4 List of Seed Words	37
Table 4.1 Punctuation Removal	46
Table 4.2 Emotion and Special Character Removal	47
Table 4.3 Consonants Removal	48
Table 4.4 Vowel and Tone Mark Removal	48
Table 4.5 Examples of Word Segmentation	49
Table 4.6 Examples of Wrong Spelling	50
Table 4.7 Examples of Unknown Words	51
Table 4.8 Examples of Noise List	51
Table 4.9 Noise Labeling	53
Table 4.10 TCCs Training Data	53
Table 4.11 CRFs Training Data	54
Table 4.12 CRFs Compared Different Context Window Size	56
Table 4.13 Evaluation Seed Words Result	58
Table 4.14 Summarization of Sentiment Dictionary	60
Table 4.15 Class Label of User Comment to News Article	61
Table 5.1 Bag-of-Words Data Form	66
Table 5.2 TF-IDF Feature Vector	67
Table 5.3 Public Sentiment Dictionary Feature Vector	68
Table 5.4 News Article Sentiment Dictionary with “AND” Feature Vector	69
Table 5.5 News Article Sentiment Dictionary with “NEAR” Feature Vector	69
Table 5.6 All Sentiment Dictionary in This Research	71
Table 5.7 Comparison with Other Research in Thai Sentiment Classification Field	81

Table of Figures

	Page
Figure 2.1 Thai Characters	9
Figure 2.2 An Example of Thai Writing System	9
Figure 2.3 Decision Tree	21
Figure 2.4 Binary Tree	22
Figure 2.5 Support Vecter Machine (SVM) Classification	24
Figure 3.1 The Overview Framework	29
Figure 3.2 Data Collection	30
Figure 3.3 User Comment Collection	32
Figure 3.4 News Article Collection	33
Figure 3.5 Text Normalization	34
Figure 3.6 Sentiment Dictionary Construction	36
Figure 3.7 Example of SO-PMI-IR with “AND” Operation	37
Figure 3.8 Example of SO-PMI-IR with “NEAR” Operation	38
Figure 3.9 Sentiment Estimation	39
Figure 3.10 Estimate Sentiment Value of User Comments	39
Figure 3.11 Sentiment Classification	40
Figure 4.1 Overview of Text Normalization	43
Figure 4.2 Heuristic Approach	45
Figure 4.3 Noise Validation	50
Figure 4.4 Character Based CRFs Approach	52
Figure 4.5 Modified Input Template	55
Figure 4.6 Sentiment Dictionary Construction	57
Figure 5.1 Sentiment Classification	64
Figure 5.2 Random Split Data	70
Figure 5.3 Precision of User Comment Sentiment Dictionary	73
Figure 5.4 Recall of User Comment Sentiment Dictionary	74
Figure 5.5 F-measure of User Comment Sentiment Dictionary	75

Table of Figures (cont.)

	Page
Figure 5.6 Accuracy of User Comment Sentiment Dictionary	76
Figure 5.7 Precision of Feature Vector	77
Figure 5.8 Recall of Feature Vector	78
Figure 5.9 F-measure of Feature Vector	79
Figure 5.10 Accuracy of Feature Vector	80

CHAPTER 1

Introduction

1.1 Statement of Problem

More and more people have been using the Internet for communication among them with email, online chats, blogs, etc., and one of the recent communication platforms are social networking. As of 2018, social media are almost an essential part of our daily life, and many social media services such as Facebook, Twitter, Instagram, LinkedIn, Google+, etc., provide space for communication, the source of news, other people's opinions, and more of information. Different contents, services, and interfaces are provided by different social media providers, but an important common feature is the mechanism to promote interactions among users. Almost every social media has the button, so-called LIKE button, with which users easily show their feeling towards articles and users' opinions. It is also easy to share articles in each user's friend group or to the public, and users can post their comments to articles and users' opinions. Naturally, social media providers want to present to each user more articles and users' posts more intriguing to him/her. Thus, predicting which article and users' post is suitable to present to each user is an important research and development issue, which is one of the "Recommendation" issues.

Recommendation methods are roughly divided into two directions. One is collaborative filtering, with which, consulting the activities of other users' whose activity tendency is similar to a user A, the system recommends to the user A something that he has not reacted to but

other users reacted. The other is content-based recommendation, with which, consulting the contents that a user A reacted before, another content with similar characteristics is recommended to the user A. These two directions can be combined, but put it simply, if recommendation is made with the data on which articles received LIKE from a user A among other users, it is more of collaborative filtering; and if recommendation is made with the content of an article that received LIKE from user A, it is more of content-based recommendation.

One of the most popular social media in the world is Facebook (Servia-Rodriguez, et al., 2013; Balahur and Perea-Ortega, 2015). Facebook offers both open and closed communications, and the number of the users has been increasing, up to 1,500 million (Number of daily active Facebook users worldwide as of 1st quarter, 2018). Facebook has many functions, and among them are “Share” and “Like”. “Share” is to share an external link on one’s own timeline where his friends and others can see, and “Like” can be used to show his interest in other people’s article posting including shared ones. These two functions have facilitated expressing one’s interest in an article, and everything posted onto Facebook, whether it is his own article, pictures, videos, comments, shares, etc., have numbers to show how many times they are “Shared” and “Liked”. “Share” can go with a sharer’s own comment, and all the items that people can “Like” also accept comments, though “Like” and comments are not explicitly related. The trouble is that both “Share” and “Like” do not necessarily mean positive support by those who “Share” and/or “Like”. Strictly speaking, they just indicate that the item is paid attention. However, LIKE is often troublesome. From 2016, Facebook provides several types of LIKE (e.g., Interesting, Amazing, Awful, etc.), but the default one is Like, which is often chosen for convenience regardless of the actual feeling. Thus, we are not sure if indeed LIKE represent the positive feeling of a user towards a content. If the underlying sentiment, whether a person likes or dislikes, agrees or disagrees, etc., is estimated, the deeper understanding of the status of the item will be obtained. There might be some different ways for this, varying according to what is focused, and one of them is to focus on textual descriptions and comments. This is why sentiment analysis of user comment texts is important.

Textual descriptions with a shared link and textual comments to an item contain much more meaningful information, and to reveal something from those texts is conducted using text mining techniques, which are based on both natural language processing techniques and machine learning methods. In the above case, the first simple idea is to employ those techniques to analyze the texts attached to “Share” and “Like,” and then to estimate its sentiment, positive or negative. However, this simple idea will not work, mainly because an accurate analysis like that requires a large size of the text data, which are usually not available for each item, and also partly because an appropriate sentiment dictionary that is required for the above analysis is not readily available.

Thus, my research tackles these two issues, how to improve a textual sentiment analysis with “Share” and “Like” and how to prepare an appropriate sentiment dictionary for the first issue. For the first issue, not only the targeted texts but also other past texts that the same user posted will be employed to increase the data size and to estimate his preferences on “Share” and “Like”. For the second issue, the Thai language is the target and a method to construct prototypical sentiment dictionaries with other sets of textual data and estimate their appropriateness with comparison between Facebook data and the other sets of textual data.

This research focuses on Thai news providers. The language used to present articles and reviews is Thai language and the language is sometimes called an unstructured language (Yassine and Hajj, 2010; Tuckett, et al., 2014).

1.2 Research Problems

1.2.1 The total size of the tests posted with a particular item in social networking site is far below the necessary size of usual sentiment analysis.

1.2.2 An appropriate Thai sentiment dictionary for our purposes is not readily available.

1.2.3 The social media text contains a lot of noise and short text, it difficult to sentiment analysis.

1.3 Objectives

1.3.1 To develop and improve a method to estimate the sentiment of shared and liked items that are posted on social networking site.

1.3.2 To propose a method to construct a sentiment dictionary for a particular type of data using texts from different sources.

1.3.3 To propose a noise reduction method for Thai social media text.

1.4 Research Question

1.4.1 How to employ related textual data to make a finer sentiment analysis of "Share" and "Like" that are frequently used on social networking sites.

1.4.2 How to construct an appropriate sentiment dictionary for a particular purpose by using different sets of texts.

1.4.3 How to remove noise in social media text for an unofficial language and short text.

1.5 Scopes of Research

1.5.1 This research experimented with only Thai texts.

1.5.2 Facebook data is employed as a representative of social networking data that users post.

1.5.3 Estimating the value of emoticons that may be used in the textual data is basically out of scope, of it is a different research topic that may be pursued.

However, the techniques and methods to be proposed in my research are expected to work with similar types of data sets, and in this sense, this research is not limited to Facebook but to texts that are used in social networking sites in general.

1.6 Expected Outcomes

1.6.1 A method useful to estimate the sentiment of short texts is proposed.

1.6.2 A method to construct a sentiment dictionary for social networking text analysis is proposed and some samples of Thai sentiment dictionaries are obtained.

CHAPTER 2

Literature Review

2.1 Introduction

Social network or online communities as web-based services allow users to communicate and share opinions. Currently, there are many social networks providers such as: Facebook, Twitter, Google+, Flickr and Instagram. The nature and nomenclature of these connections may vary from site to site (Ellison, 2007). Each of the sites, which also has differences on the subject of information and communication tools, such as photo, video-sharing, mobile connect, API and etc.

With an increasing number of social network users, the social network is also a huge new data warehouse, that most of the information that are stored on social network are usually in text format, and this has led much research on social networks particularly in text mining tasks (Yassine and Hajj, 2010; Servia-Rodriguez, et al., 2013; Balahur and Perea-Ortega, 2015; Maynard, et al., 2017). Extracting useful information from texts has been pursued. According to Feldman, et al. (2007). There are a variety of tasks including text classification, text clustering, information extraction, author estimation, and sentiment analysis such as “A Framework for Emotion Mining from Text in Online Social Networks (Yassine and Hajj, 2010), “Social media competitive analysis and text mining: A case study in the pizza industry” (He, et al., 2013), “More than words: Social networks’ text mining for consumer brand sentiments” (Mostafa, 2013) and “Sentiment analysis

system adaptation for multilingual processing: The case of tweets" (Balahur and Perea-Ortega, 2015). As opposed to most data used in traditional data mining and data analysis, textual data is not structured in the sense that texts themselves are not in the form of relational database or other database structures. More primitively, any written text data consist of characters or letters. Thus, a series of preprocessing tasks are indispensable for text mining. Most of those preprocessing tasks are related to natural language processing with which basic linguistic units are obtained: word, sentence, and their structures and meanings (Thanangthanakij, et al., 2012). In the next section, we introduce natural language processing in Thai social media, sentiment analysis and machine learning techniques.

2.2 Thai Language

The Thai language is the official language of Thailand, which a unique writing system with its own syllabic alphabets consist of 3 parts: consonant, vowel and tone (Ager, 2017). The Thai language has 87 characters as Figure 2.1 consist of 44 consonant characters, 18 vowel symbols, 4 tone marks, 5 diacritics, 10 numerals, and 6 other symbols. For the Thai writing system, It is written from left-to-right direction, and one of the unique features is that vowel symbols are placed in front, back, above, and below a consonant character. Tone marks are also placed above a consonant, diacritics are written above or below a consonant, no word separation between words. A space used between sentences and independent phrases in a sense, but the rule of the use of spaces is not so absolute. There is no capital letters, unlike English or other Roman alphabetical writing systems. In addition, Thai verbs do not change their forms according to tense or concord, and aspect as well as tense is optionally expressed with auxiliary verbs or time adverbials (Theeramunkong and Usanavasin, 2001; Koanantakool, et al., 2009).

Consonants	44	ก ข ฃ ค ฅ ฆ ง จ ฉ ช ซ ฌ ญ ฎ ฏ ฐ ฑ ฒ ณ ด ต ถ ท ธ น บ ป ผ ฝ พ ฟ ภ ม ย ร ล ว ศ ษ ส ห ฬ อ ฮ
Vowels	18	ะ ั า ำ ิ ี ึ ื ุ ู เ แ โ ใ ไ ๅ ฤ ฦ
Tone marks	4	่ ้ ๊ ๋
Diacritics	5	ฺ ็ ์ ๎ ๎์
Numerals	10	๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
Other symbols	6	ฯ ฿ ๆ ๏ ๚ ๛
Total	87	

Figure 2.1 Thai Characters (Koanantakool, et al., 2009).

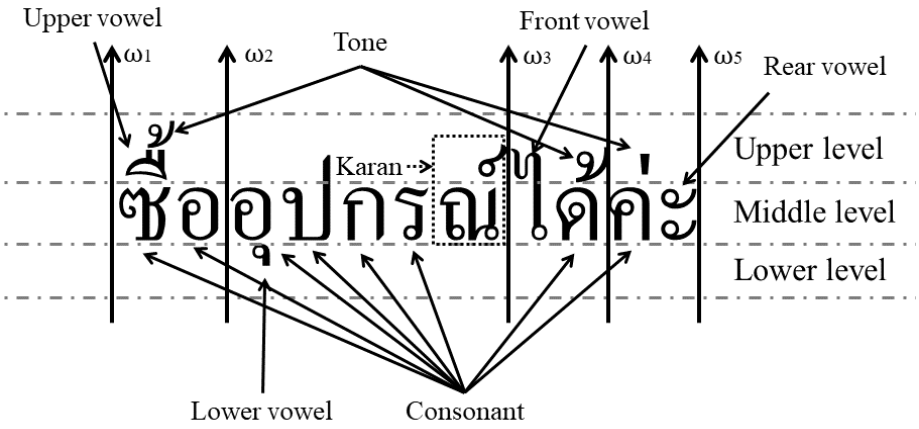


Figure 2.2 An Example of Thai Writing System

An example of the Thai writing system is shown in Figure 2.2. Usually, it is considered that the Thai writing has three levels: upper, middle, and lower levels. There are seven different types of characters, symbols, and marks consonant, upper vowel, lower vowel, front vowel, rear vowel, tone and karan (The pronunciation cancellation symbol). These different characters, symbols, and marks are configured in a way that the main consonant comes in the center and all the other symbols and marks are placed around it. ω_i indicates word boundaries (Tongtep and Theeramunkong, 2011).

2.3 Thai Character Clusters

The concept of Thai character clusters (TCCs) was first proposed by Theeramunkong, et al. (2000). Thai character clusters are the set of combinations of characters that can be regarded as inseparable units, which smaller or equal a word, bigger or equal a character, meaningful or meaningless. as usually 26 types of TCCs are employed as in Table 2.1. TCCs are descriptive rules of character combinations: e.g., a vowel is always located together a consonant (e.g. “ไป”, “มา”, “มี” and “หู”), and a tone mark always appears above a consonant (e.g. “ไม่”, “ชี้”, “เจ๊” and “อ๋อ”) (Tongtep and Theeramunkong, 2011). It is known that converting the original text using TCCs as a preprocess improves the subsequent procedure of word segmentation, and this research also employs TCCs.

Table 2.1 Type of Character Cluster (Tongtep and Theeramunkong, 2011)

Symbol	Type	Example	Symbol	Type	Example
T1	C+VF (Font vowel)	ใอ ไอ เอ แอ โอ	T1T	C+VF+T	T1+T
T2	C+VB (Right vowel)	อะ อำ อา	T2T	C+VB+T	T2+T
T3	C+VU (Upper vowel)	อั- อี๋ อ ิ อ ึ อี อื	T3T	C+VU+T	T3+T
T4	C+VL (Lower vowel)	อุ อ	T4T	C+VL+T	T4+T
TT	C+T	อ่ อ้- อ๊- อ๋-	TK	C+Karan	อ์
T6	C+VF+VB	เอะ แอะ โอะ เอาะ	T6T	C+VF+VB+T	T6+T
T7	C+VF+VU	เอิ-	T7T	C+VF+VU+T	T7+T
T8	C+VB+VU	อัวะ, อัว	T8T	C+VB+VU+T	T8+T
T9	C+VF+VB+VU	เอียะ เอือะ เอีย	T9T	C+VF+VB+VU+T	T9+T
S	Space	<space>	V	Vowel	เ แ โ ใ ไ
D	Digit	๐-๙, 0-1	P	Punctuation	@#$%&*ฯๆ
C	Consonant	ก-ฮ	T	Tone	อ อ่ อ้ อ๊ อ๋
E	English	a-z, A-Z	O	Other	

2.4 Natural Language in Thai Social Media

There are many social media services such as Twitter, Facebook, LINE, Instagram, and Google Plus, to name a few. Each of them has different foci in media and topics. For example, Instagram focuses on photo sharing; Twitter focuses mainly on short texts; LINE is more dedicated to private communication, etc. However, three of the important common features are the easy

sharing of articles and other media from both inside and out of a social media service, the use of LIKE buttons which is used to conveniently express a reactive feeling of the reader, and textual comments. As all these three common features represent sentiments, how a user feels or thinks towards something, many studies have been conducted on sentiment analysis. Among them, text sentiment analysis with articles and user comments have been pursued, as well as other text mining tasks (He, et al., 2013; Balahur and Perea-Ortega, 2015; Hirankan, et al., 2013; Kunpattanasopon, et al., 2017). This research performs text sentiment analysis with Thai textual data, and the basic process of sentiment analysis is discussed in Section 2.5.

Many sentiment analyses of social media contents have focused on LIKE buttons, because they are the most convenient way for users to express their reactive feelings towards a content. For example, Facebook currently has six different LIKE buttons: Like, Love, Laugh, Wow, Sad and Angry, although until 2016 Facebook had only one LIKE button. Even when a social media service provides different types of LIKE buttons, however, the most general LIKE button “Like” is most used. In the case of Facebook, its user interface first presents Like, and the other buttons are shown a bit later. Thus, regardless of the number of different LIKE buttons, users often use the general LIKE button. It means that users often use the general LIKE button even when their reactive feeling is more of Sad or Angry. Thus it is all the more important to analyze user comments, and this research focuses on texts.

2.4.1 Characteristics of Social Media Texts

We can separate characteristic of social media of texts in 2 types:

1) Text from news or article, the official language, there are few mistake and validation before presentation.

2) Text from user comment, it is usually written in very informal ways without careful proofreading before posting in most cases. User comments contain non-standard spellings,

slangs, word plays, and have frequent use of emoticons, onomatopoeia words, etc., to express their feelings more casually and expressively.

2.4.2 Characteristics of Thai social media texts and how they have been treated

Thai social media texts writing is different from Thai official language writing due to Thai writing complexities such as word separation. In addition, texts in social media are very short and contain many noises such as emoticons, slangs, advertisements, word plays, wrong or non-standard spelling variations and words that are deliberately written wrongly to make a sound like a written word such as “ใช่ ไช่ ช่าย” (yes) and special characters instead of Thai consonant are also used in some words. All these make word segmentation, which is the first step of text analysis, more difficult. (Aroonmanakun, 2002; Wilson, et al., 2005; Tongtep, 2013).

2.5 Noise Reduction

Noise reduction is a part of pre-processing to remove unnecessary parts in the text for sentiment analysis. This pre-processing is mandatory because social media text is written in an informal style and contain much noise such as special character, emoticon, extra repetitions of characters, etc. Some studies conduct this process rather ad hoc, but a systematic treatment is desirable. There are roughly two approaches: a heuristic approach and a machine learning approach.

2.5.1 Heuristic Approach

A heuristic approach is a noise reduction technique devised by human experience, and it usually consists of a set of rules to convert strings into others. Typically, such rules are prepared for five different types of noises: URLs, extra repetitions of characters, advertisement and

spam, special characters, and emoticons (Hemalatha, et al., 2012; and Kunpattanasopon, et al., 2017). Kouloumpis and other (Kouloumpis, et al., 2011; Wang, et al., 2012; Ortigosa, et al., 2014) detect and replace redundant characters by replacing more than two same characters with the same one or two characters. Bigi and other (Bigi, 2011; Zhang, et al., 2011; Mulkalwar and Kelkar, 2012; and Balahur and Perea-Ortega, 2015) employed a dictionary to detect and checked whether each word is contained in the dictionary or not.

2.5.2 Machine Learning Approach

A machine learning approach applies a machine learning method to noise reduction. Hirankan, et al. (2013) proposed a technique to detect wordplay tokens in Thai social network text by applying a Thai word segment tool based on Conditions Random Fields (TLex). It tries to detect repetitions of Thai consonantal character and vowel characters in wordplay tokens. Jiamthapthaksin, et al. (2016) proposed a technique to extract Thai slangs from social media posts. They applied a word *n-gram* approach to detect slang words. Kunpattanasopon , et al. (2017), applied Conditions Random Fields to detect word spelling errors after minimum heuristic noise reductions, which is part of this thesis.

2.6 Sentiment Analysis

Sentiment analysis or opinion mining is the task of analysis of identifying opinions, emotions, and evaluations into the polarity of positive and negative or more types, most are in the text classification category. (Jain and Nemade, 2010; Wilson, et al., 2005; Tsihrintzis, 2013). Few year ago, sentiment analysis have been a number of research focus on social network such as: Facebook (Ahkter and Soria, 2010; Troussas, et al., 2013; Ngoc and Yoo, 2014; Ortigosa, et al., 2014), Twitter (Kouloumpis, et al., 2011; Nielsen, 2011; Khan, et al., 2015; Nakov, 2017;

Ray and Chakrabarti, 2017; Rosenthal, et al., 2017), News article (Ahmad, 2006; Godbole, et al., 2007; Bautin, et al., 2008; Balahur, et al., 2013) and product and service (Thanangthanakij, et al., 2012; Guzman and Maalej, 2014; Fang and Zhan, 2015)

However, by the basic idea is the same: constructing a sentiment dictionary for the targeted textual domain in which each listed word has its positive or negative value, and then estimating the sentiment value of a text by summing up the frequency of the words in the sentiment dictionary that appear in the text. However, any sentiment estimation techniques are based on a sentiment dictionary, and thus the availability and reliability of a sentiment dictionary for a targeted textual domain is critical.

In Thai language, also have many research applied sentiment analysis technique with Thai social media text such as: Inrak and Sinthupinyo (2010), applied Latent Semantic Analysis (LSA) to classify emotions in Thai text by compared 3 machine learning and two models, Chirawichitchai (2014) classified Thai social media text in to six emotions which compared machine learning technique and deference feature technique, Sarakit, et al. (2015) classifying six emotion in Thai YouTube comment which three machine learning are comparison, and Vateekul and Koomsubha (2016), classify sentiment of Thai twitter data by applied two deep learning technique: Long short Term Memory (LSTM) and Dynamic Convolutional Neural Network (DCNN).

2.7 Sentiment Dictionary Construction

Sentiment dictionary or sentiment lexicons is list of words have polarity value: positive and negative use for sentiment estimation or sentiment analysis.

However, many researches proposed technique to construction dictionary such as: Steinberger, et al. (2011) proposed “Creating Sentiment Dictionaries via Triangulation”, the technique to created two sentiment dictionaries (English and Spanish) and performed the pivot

language dictionaries to other languages by the overlap of the translations. However, this technique also necessary the initial sentiment dictionary. Which the sentiment dictionary construction based on semantic orientation from association (SO-A) (Turney, 2001) which is an unsupervised learning, by calculate association between target word and reference word, there are two techniques: Semantic Orientation from Pointwise Mutual Information and Information Retrieval (SO-PMI-IR) and Semantic Orientation from Latent Semantic Analysis (SO-SA) which the SO-PMI-IR technique is simply way and suitable for a big data set (Turney and Littman, 2002)

So, in this research construct sentiment dictionary from textual comment data by applied the SO-PMI-IR algorithm proposed by Church and Hanks (1990). The PMI-IR is original approach based statistically technique to estimate the semantic orientation of word by calculated association of target word to reference word (hereafter, we will call such a pair a "seed word").

Seed word is a reference word has polarity to strong positive or strong negative that the polarity cannot be changed in any context such as "Excellent" and "Poor" or synonym, antonym and adjective. (Turney, 2002; Esuli and Sebastiani, 2005; Yuen, et al., 2004; Taboada, et al., 2006), by estimate sentiment value of word with PMI as defined in Equation 2.1 and semantic orientation (SO) defined as in Equation 2.2

$$PMI(x, y) = \log_2\left(\frac{P(x, y)}{P(x)P(y)}\right) \tag{2.1}$$

PMI is a measure of association between target word (x) and seed words (y), by defined x is target word, y is seed word, $P(x, y)$ is probability that appear both x and y co-occur in data, $P(x)$ is probability that appear x in data and $P(y)$ is probability that appear y in data. The ratio of log of $P(x, y)$ and $P(x)\ P(y)$ is a statistical dependence between (x) and (y) per amount of data that acquire about the one of words.

$$SO_{(x)} = \sum PMI(x, positive\ seed\ word) - \sum PMI(x, negative\ seed\ word) \tag{2.2}$$

SO is calculate sentiment value of target word (x) by sum PMI positive minus sum PMI negative. There are researches that applied PMI-IR algorithm in his research such as: Morpheme-based Derivation of Bipolar Semantic Orientation of Chinese Words (Yuen, et al., 2004), Methods for creating semantic orientation dictionaries (Taboada, et al., 2006) and Automatic identification of sentiment vocabulary: exploiting low association with known sentiment terms (Gamon and Aue, 2005).

In addition, SO-PMI-IR also have two popular various of "AND" operation and "NEAR" operation. For "AND" operation is total number of user comments contain both target word and seed word as Equation 2.3. For "NEAR" operation is count number of user comment contain both target word and seed word within the pre-defined distance in a comment as Equation 2.4.

$$SO_{(x)} = \log_2\left(\frac{\text{hit}(x \text{ AND } positive_list)\text{hit}(negative_list)}{\text{hit}(x \text{ AND } negative_list)\text{hit}(positive_list)}\right) \quad (2.3)$$

$$SO_{(x)} = \log_2\left(\frac{\text{hit}(x \text{ NEAR } positive_list)\text{hit}(negative_list)}{\text{hit}(x \text{ NEAR } negative_list)\text{hit}(positive_list)}\right) \quad (2.4)$$

$$Positive_list = (good, nice, \ldots, superior) \quad (2.5)$$

$$Negative_list = (bad, nasty, \ldots, inferior) \quad (2.6)$$

In addition, also have public Thai sentiment dictionary that collected in PyThaiNLP (PyThaiNLP), which is a python package for Thai language. In this research chose only the lexicon Thai of sentiment dictionary to sentiment classification in classification part describe in Section 5.6.2.

2.8 TF-IDF

In text classifier, the important thing is feature of word feature, which each feature has value are deference. For the most popular technique for feature calculation is Term Frequency Inverse Document Frequency (TF-IDF) (Paltoglou and Thelwall, 2010; Xia and Chai, 2011; Breitinger, et al., 2015) which is a basic of word or term of vocabulary by count the number of each word that occurrences in whole document and bring the words frequency to compared with invest document frequency. It can measure the important of word and document in the corpus represented in matrix format, and this approach also reduce length of document (Blei, et al., 2003).

TF-IDF evolved from IDF which is proposed by Sparck Jones (Sparck Jones, 1972; Spärck Jones, 2004) which consist of two parts are Term Frequency (TF) and Inverse Document Frequency (IDF) as follows.

2.8.1 TF

TF is term frequency, the method for measure weight of each word (term) that occurred in document, it can be calculated as follow Equation 2.7.

$$W_{i,j} = TF_{i,j} \tag{2.7}$$

Where i is term and j is document, $W_{i,j}$ mean weight of term i in document j

2.8.2 IDF

IDF is inverse document frequency, the method for measure the important of word (term) relate with all document, it can be calculated as follow Equation 2.8.

$$IDF_i = log_2\left(\frac{N}{n_i}\right) + 1 \quad (2.8)$$

Where N mean the total number of documents divided by n_i the number of document that appeared term *i* that appeared at least one term and take the logarithm of them.

$$TF_{i,j} * IDF_i = TF_{i,j} * \ log_2\left(\frac{N}{n_i}\right) + 1 \quad (2.9)$$

When computing the TF-IDF those components are multiply as Equation 2.9. The result of TF-IDF show that the term occurs very often in the all document that mean this term such as “the”, “and” and “a” that word is not good represents the stop word but the term intensively appears in some documents probably represents the topic of a document category (Xia and Chai, 2011).

However, due to the TF-IDF weighting schemes is widely popular and common text classification (Wang, et al., 2014; Das and Chakraborty, 2018), there are many researches applied TF-IDF model with their research in text document field such as:

- A Fall-back Strategy for Sentiment Analysis in Hindi: a Case Study (Joshi, et al., 2010), by analysis Hindi movie review by translation the document in Hindi to English, then applied LibSNM with Term Frequency, TF-IDF, Term Presence to classification which the accuracies of TF-IDF representation are the highest.
- Automatic Mood Classification Using TF*IDF based on Lyrics (Van and Kanters, 2010), by applied TF-IDF to classification lyrics of song to automatic assignment mood for user. The result show that TF-IDF can be used effectively to identify word that describe mood of lyrics.

- A Clustering-based Approach on Sentiment Analysis (Li and Liu, 2010), by applied K-mean clustering the sentiment analysis movie review data by applied TF-IDF weight method compared with frequency method. The result is that the TF-IDF gives batter results.

- A Sentiment Analysis Model for Hotel Reviews based on Supervised Learning (Shi and Li, 2011), by analyzed sentiment of hotel review comments with support vector machine and compare between frequency and TF-IDF feature. The result show that the information of TF-IDF is more effective than frequency at the precision 85.2%, recall 89.2% and F-score 87.2%.

- Improving Native Language Identification with TF-IDF Weighting (Gebre et al., 2013), by applied TF-IDF weighting schemes and using linear classifiers support vector machines to identification Native Language and using 10-fold cross-validation which achieving the best accuracy is 84.55% come from feature that TF-IDF combined unigrams and bigrams of words.

Thus, in this research also use bag-of-word and TF-IDF model to represent the feature value describe Section 5.4.1.

2.9 Classification Algorithm

Classification Algorithm is a very necessary technique for text classification. In our study, we found that there are 3 algorithms that are popular or comparable: 3 algorithms: Decision Tree, Naïve-Bay and Support Vector Machine (SVM) (Chirawichitchai, 2014; Ortigosa, et al., 2014; Wang, et al., 2014; Sarakit, et al., 2015) and one more algorithm for labeling noise reduction is Conditional Random Fields (CRFs).

2.9.1 Decision Tree

Decision Tree is a popular classification model and easy for human to understand learned result. Decision Tree model is a flowchart like upside down tree structure consist of node and branch, when there are new data it will bring the feature of new data to compare with node in each branch in tree until last node.

Node denotes attributes used to test attributes or decisions can be divided into 3 types: 1) Root node is the topmost note in a tree and beginning node, not have branches come in to this node but have branches go out to other node 2) Internal node is node that branch come in and go out and 3) Leaf node or terminal node is last node that holds a class label. Branch is line connect the relationship between node. Which each branch represents the result of the value of the attribute as Figure 2.3.

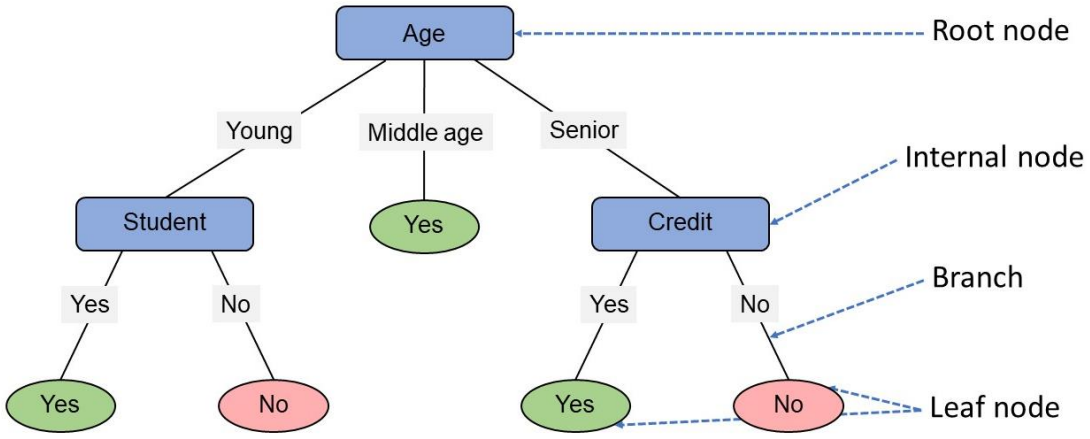


Figure 2.3 Decision Tree

In the process of creating a decision tree model, by select the attribute that influence the highest classification to root node, then find more the attribute that influence the highest classification to the next node unless the condition of decision will be true. Which the algorithm to create decision tree model as: ID3, C4.5, C5.0 and CART.

ID3, C4.5 and C5.0 algorithm proposed by Quinlan (2014) first developed the ID3 algorithm by calculate the Information Gain of each attribute to select highest value of attribute add to each node, which repeat until the data is completely categorized. Then, proposed C4.5 algorithms

to solve the bias problem in ID3 algorithm. For C5.0 algorithm, there are similar procedures C4.5 algorithms which increases performance of classification, optimized memory usage and suitable big data, but development of commercial software.

CART or Classification and Regression Trees algorithm (Breiman, 2017) is binary tree, each node consists of only two branches as Figure 2.4. by calculate Gini Index for selection each attribute to be the root node and leaf node.

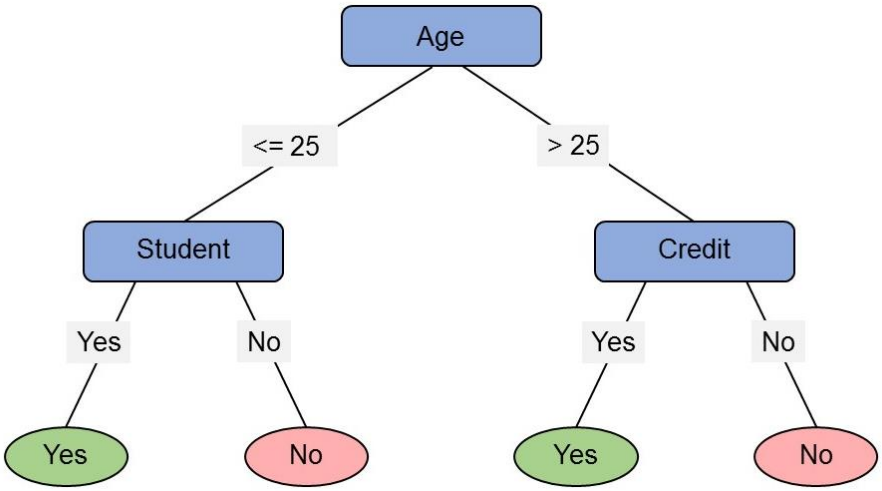


Figure 2.4 Binary Tree

However, most important thing to consider when creating decision tree is that the appropriate feature selection to be the root node and leaf node. by feature selection technique: Information gain and Gini Index.

Currently, decision tree model applied for NLP problems such as: Automatic Corpus-Based Thai Word Extraction with the C4.5 Learning Algorithm (Sornlertlamvanich, et al., 2000), Discovery of predicate-oriented relations among named entities extracted from Thai texts (Tongtep and Theeramunkong, 2012), Detecting Romanized Thai tokens in social media texts (Moknarong, et al., 2013) and Thai stock news sentiment classification using wordpair features (Netisopakul and Chattupan, 2015). In this research, we used decision tree to sentiment classification in Chapter 5

2.9.2 Naïve-Bayes

Naïve-Bayes or simple Bayesian are statistical classifiers based on applying Bayes' theorem (Bayes and Price, 1763) Which is less complicated. Can predict results and explain them. By analyze the relationship each independence variable to create the probability condition for each relationship. Theoretically, the results of Naïve-Bayes will be correct if all feature are independence. Bayes' theorem used probability calculation is called conditional probability, mean probability of the event A If there is another occurrence B, this can be calculated as follow Equation 2.10.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.10}$$

B instead of attribute used to calculate the posterior probability of A

P(A|B) is posterior probability of A conditioned on B

P(B|A) is posterior probability of B conditioned on A

P(A) is prior probability of A

P(B) is prior probability of B

Naïve- Bayes calculates the probability of each attribute affecting the class as: X is data train consist of attribute N number X = (X_1, X_2, X_3, …, X_n) and has class M (C_1, C_2, C_3, …, C_n) which can be written as Equation 2.11.

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \tag{2.11}$$

Naïve-Bayes has been applied to many researches in machine learning particular text classification such as: Thailand Tourism and conflict: Modeling sentiment from Twitter tweets using naïve Bayes and unsupervised artificial neural nets. (Claster, et al., 2010), Multi-stage annotation using pattern-based and statistical-based techniques for automatic Thai annotated corpus construction (Tongtep and Theeramunkong, 2011) and S-Sense: A sentiment analysis framework

for social media sensing (Haruechaiyasak, et al., 2013). So, we used Naïve-Bayes model to sentiment classification in Chapter 5.

2.9.3 Support Vector Machine (SVM)

SVM is a supervised learning method algorithm that proposed by Corinna Cortes and Vladimir Vapnik (1955). The SVM was designed to discriminative classification the binary class by the SVM to find and create the best linear that maximum margin to separate between two class which called hyper plane (Takeuchi and Collier, 2005; Nomponkrang and Sanrach, 2016) by focus on optimal separating hyper plane. Figure 2.5 illustrates an example of SVM, the objects belong either to class green-circle or red-square. The separating line defines a boundary on the right side of which all objects are green-circle and to the left of which all objects are red-square. SVMs find the line passing as far as possible from all points. And use a kernel function to change low dimension dataset on Input Space up to height dimension dataset in Feature Space for dividing linear data (Dumais, et al., 1998; Choksuriwong, 2012). In additional there is also developer tools for SVM such as: Joachims (1999) proposed the SVM light and Chang and Lin (2011) proposed A Library for Support Vector Machines (LIBSVM).

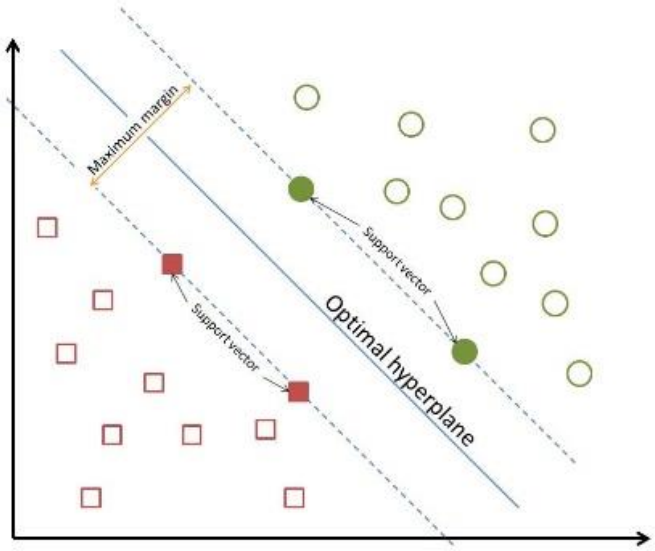


Figure 2.5 Support Vecter Machine (SVM) Classification

However, the training data mostly non-linear, which solve this problem with used kernel function. Kernel function is a function to move low dimensional input space transform to a higher dimensional space for find non-linear hyper plane. It is mostly useful for non-linear separation. Which popular kernel function of SVM as: Linear kernel, Radial basis kernel (RBF), Polynomial kernel, Sigmoid Kernel.

The SVM algorithm is very popular in many task, therefore the researcher applied SVM to many field as: An Empirical Study on Multi-Dimensional Sentiment Analysis from User Service Reviews (Thanangthanakij, et al., 2012), Sentiment analysis using support vector machine with diverse information sources (Mullen and Collier, 2004) and Twitter as a Corpus for Sentiment Analysis and Opinion Mining (Pak and Paroubek, 2010).

In this research, we used linear kernel function for our classification, based on our study, we found that the linear kernel function is a simplest kernel function and suitable for text classification (Joachims, 1998; Chirawichitchai, et al., 2010; Joshi, et al., 2010; Paltoglou and Thelwall, 2010; Mulkalwar and Kelkar, 2012; Chirawichitchai, 2014; Balahur and Perea-Ortega, 2015) which the equation of linear kernel showed that Equation 2.12.

$$K(x_i, x_j) = x_i^T x_j \tag{2.12}$$

2.9.4 Conditional Random Fields

Conditional random fields (CRFs) (Lafferty, et al., 2001) are a supervise learning based on discriminative undirected probabilistic graphical model for calculating the conditional probability. CRFs has developed from Maximal entropy markov models (MEMMs) by used advantage of MEMMs and solves the label bias problem (Klinger and Tomanek, 2007; Feng, et al., 2008; Sutton and McCallum, 2012) for the CRFs equation as Equation 2.13.

$$P(y|x) = \frac{1}{Z(x)} exp\left(\sum_{t=1}^{T}\sum_{K=1}^{K} \lambda_k f_{k(y_{t-1}, y_t, x, t)}\right) \tag{2.13}$$

Where $P(y|x)$ is conditional probability, y is a sequence of result, x is a sequence of result, K is a number of function, T is a position, λ_k is a weight of feature function, f_k is a feature function and $Z(x)$ is a normalization function can be writing in Equation 2.14.

$$Z(x) = \sum_{y} exp\left(\sum_{t=1}^{T}\sum_{K=1}^{K} \lambda_k f_{k(y_{t-1},y_t,x,t)}\right) \quad (2.14)$$

For general, CRFs used for labeling, segmenting and part-of-speech tagging such as A Supervised Learning based Chunking in Thai using Categorical Grammar (Supnithi, et al., 2010), Simultaneous character-cluster-based word segmentation and named entity recognition in Thai language (Tongtep and Theeramunkong, 2011) and LexToPlus: A Thai Lexeme Tokenization and Normalization Tool (Haruechaiyasak and Kongthon, 2013).

2.10 Recommendation

Recommendation is a frequently used technique to recommend something, whether it is a news article, a produce, or a service, to someone according to their past related activities and/or other people's activities similar to him, which many researches applied to many tasks such as citation recommendation system (Zarrinkalam and Kahani, 2012), product recommendation system (Linden, et al., 2003) and TV program recommendation (Barragáns-Martínez, et al., 2010). However, this system is work, it requires the analysis of information from user such as voting or user comment. The recommendation consists of 1) prediction phase to analyze the information about user preferences to item that used or seen to predict user preferences of new item. 2) recommendation phase to use score of new item in previous phase recommend to user (Burke, 2002).

In general, the recommendation can be divided in two approaches: content-based and collaborative filtering (Sarwar, et al., 2001). The content-based, use characteristic of item that

user used or seen to check similarity of new item to recommend user (Pazzani and Billsus, 2007; Cantador, et al., 2010). On the other hand, the collaborative filtering to recommend new item to particular purpose of user based on history of user (Linden, et al., 2003; Yu, et al., 2017). However, the collaborative filtering can be divided to two main categories: Memory-based and Item-based. The Memory-based collaborative filtering or user based is a technique to prediction from statistical by group of similarity behaves of users, known as neighbors. The Model-base collaborative filtering or item based is a technique to prediction from the history user of probability of an item being used or purchased simultaneously (Breese, et al., 1998; Sarwar, et al., 2001).

In additional, also proposed other technique to recommendation such as: Pazzani (1999) proposed Demographic Filtering by applied collaborative filtering with user personal information such as: age, gender, education, job, area code, employed and etc., and García-Crespo, et al. (2011) proposed a Sem-Fit: a semantic based expert system to provide recommendations in the tourism domain by applied fuzzy logic techniques with consumer's experience to expressed customer and hotel characteristics.

CHAPTER 3

Overview of the Proposed Methodology

3.1 Introduction

This thesis aims to propose a new method to make classification of news articles according to each user, based on the estimation of the sentiment value of the corresponding user comments by each user, which is then based on the construction of sentiment dictionaries with aggregated datasets from a social media, Facebook. The target language is the Thai language, and as already discussed in Chapter 1 and Chapter 2, social media texts need a strong noise reduction method, which is also investigated in this thesis.

3.2 The Overview Framework

The overview framework of the research in this thesis is schematized as in Figure 3.1. It consists of 5 parts:

(1) Data collection of two datasets: news articles and user comments

(2) Text Normalization (Noise reduction)

(3) Sentiment dictionary construction for both datasets

(4) Sentiment estimation of user comments

(5) Sentiment classification of news articles according to each user and recommendation

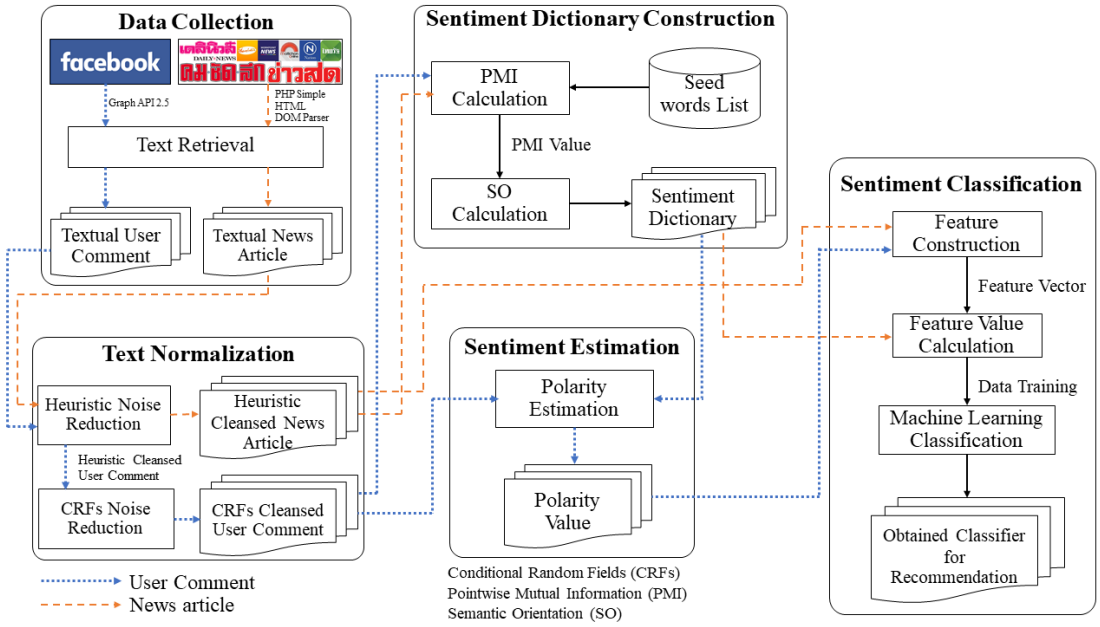


Figure 3.1 The Overview Framework

3.2.1 Data Collection

This research used data from two sources: 1) user comments from news providers' Facebook fanpage and 2) news article from news provider main websites as in Figure 3.2. They are related in a way that users can post their opinion at their Facebook fanpage to news articles at their main website.

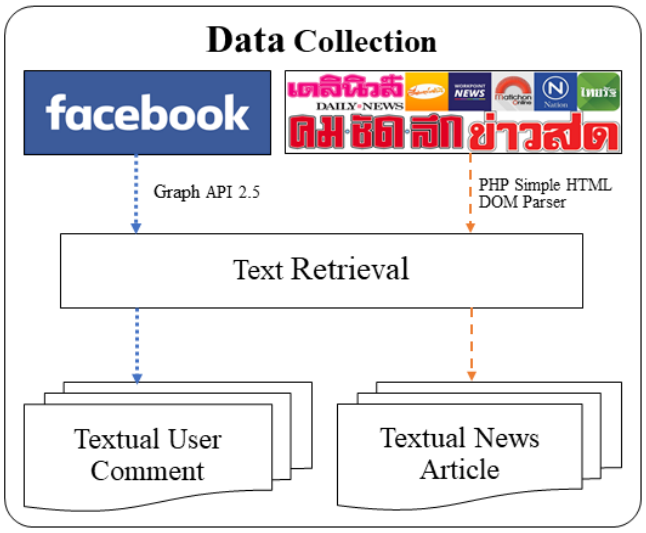


Figure 3.2 Data Collection

We choses 8 news providers from totally 13 news providers on condition that a news provider must have both main website and Facebook fanpage whose users exceeds 500,000 people. In Table 3.1, we showed the news providers: MorningNews TV3, Khaosod, Thairath, Workpoint Entertaimnent, NationTV 22, Komchadluk, Daily News Online and Matichon Online.

Table 3.1 List of News Providers

No.	News Provider	Provider Website	Number of Likes (12-Jan-2014)
1	MorningNews TV3	http://morning-news.bectero.com	9,490,592
2	Khaosod	https://www.khaosod.co.th	8,109,166
3	Thairath	https://www.thairath.co.th	5,176,584
4	Workpoint Entertaimnent	https://workpointnews.com	4,955,885
5	NationTV 22	http://www.nationtv.tv	2,409,153

Table 3.1 List of News Providers (cont.)

No.	News Provider	Provider Website	Number of Likes (12-Jan-2014)
6	Komchadluk	http://www.komchadluek.net	1,333,355
7	Daily News Online	https://www.dailynews.co.th	1,108,701
8	Matichon Online	https://www.matichon.co.th	658,269

These two types of data are scraped as explained in the next subsection.

3.2.1.1 User Comment Collection Data

We auto accumulated textual user comments by created PHP script connecting to Facebook news provider fanpage through Facebook Graph API 2.5 by collecting the URL of news articles, user “Like”, “Share” of news articles, and user comment as in Figure 3.3, then stored them in the MySQL database.

Figure 3.3 User Comment Collection

As in Figure 3.3, a post in Facebook fanpages consists of three parts: Headline and illustration, User “Like” and “Share” button and user comments. First, the headline and the illustration part show the title and the figure of the news article whose hidden URL is linked to the whole news article in the provider main website, and we used this URLs for collecting textual news article data. Second, user “Like” and “Share” button is used by Facebook users to show LIKE or share this article on their Facebook timeline. Third, the user comments part shows people’s comments and the text box to write a new comment.

3.2.1.2 News Article Collection Data

In this part, we also accumulated textual news articles by accessing to the URLs of news articles that are collected in user comment collection data part and retrieved them with PHP

Simple HTML DOM Parser, which is a PHP library to access to the topic and details to store in the MySQL database as in Figure 3.4.

Figure 3.4 News Article Collection

All data were collected over two years from January 2014 to January 2016. The amount of data that we collected is shown in Table 3.2.

Table 3.2 Summarized Data

Data	Records	Words	Characters
News Articles	172,189	≈ 85,000,000	≈ 340,000,000
User Comments	8,34,3126	≈ 110,000,000	≈ 432,000,000

3.2.2 Text Normalization

From Figure 3.5, The part of text normalization is designed to remove various noise in textual user comments and textual news articles.

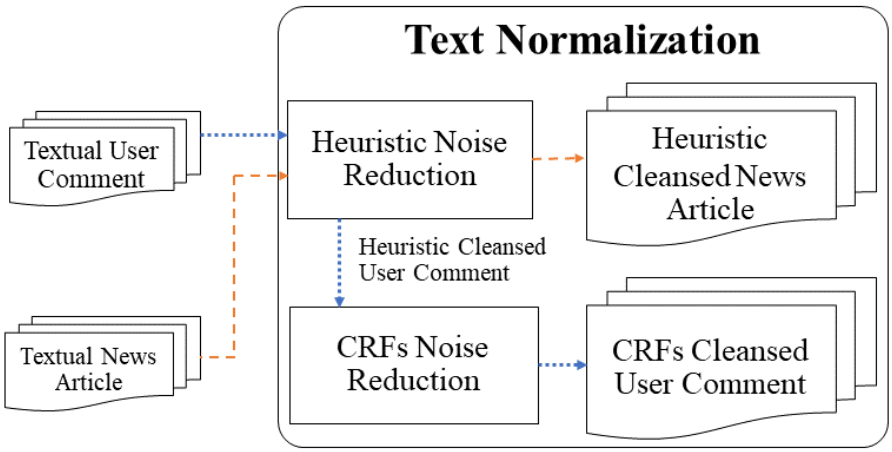


Figure 3.5 Text Normalization

From my survey of our data found that social media texts contain many noises. Typical of them are listed in Table 3.3.

Table 3.3 Type of Noise in Social Media Texts

Type of Noise	Example
Spelling errors	เครือบิน : เครื่องบิน (Airplane)
Special characters and emoticons	☯, ★, ♂ , ✿, :), =D, :P
Repetition of the same characters	ดีมากกกก (Veryyy Good)
Punctuations	_ - ( ) , . : ; ? ! ‘ ’ “ ”
Dialectal variations	บักหุ่ง ลอกอ = มะละกอ (Papaya)

Table 3.4 Type of Noise in Social Media Texts (cont.)

Type of Noise	Example
Word play	จุ๊ฟฟฟ มั๊วฟฟฟ (Kiss sound)
Slang	วัดจานบิน (Phra Dhammakaya Temple)
Repetition of the same words	ฮาฮาฮา (hahaha)
Spam advertising	ขออนุญาตประชาสัมพันธ์… (Request Information…)
URLs	http://www.thairath.co.th, https://www.matichon.co.th

However, this research have two datasets: textual user comments datasets and textual news article datasets. These two datasets have different characteristics. Textual user comment data consist of many noise types, and only the application of rules is not enough. Many past studies on the Thai language processing seem to have had manual corrections, but as my dataset is fairly large, such manual corrections are almost impossible. Thus, the application of machine learning is considered. We propose a hybrid noise reduction method using both heuristic rules-based method and a machine learning method. More specifically we propose the employment of CRFs for Thai spelling variation checks. This will be discussed in Chapter 4, with experiments to find the best noise reduction technique for our dataset.

For the textual news articles data, textual news article data need noise reduction, but the necessary noise reduction is different from user comments, because, unlike user comments, news articles are professionally written and checked, and do not contain emoticons, special characters, and various non-standard spelling variations in principle. On the other hand, noises remain coming from data scraping processes such as HTML tags, URLs, pictures and videos. These noises are the target of heuristic approaches. Thus, we used only the related heuristic noise reduction method for news article data.

3.2.3 Sentiment Dictionary Construction

In my framework, the sentiment value of words will be used in two different datasets: user comments data and new articles data. Thus, different sentiment dictionaries are constructed for each purpose. User comments sentiment dictionary is used to estimate the polarity value of user comments data for class labels in supervised data for machine learning sentiment classification part. On the other hand, news article sentiment dictionary is used to represent the polarity feature value of supervised data in sentiment classification part as described in Section 5.4.3.

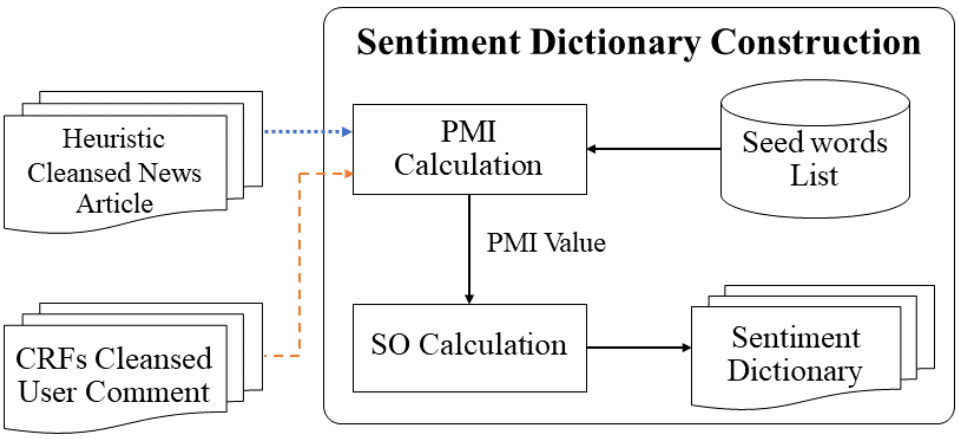


Figure 3.6 Sentiment Dictionary Construction

Although the sentiment dictionary for user comment and that for news artilecles are separately constructed, the construction method is the same as Figure 3.6, by using SO-PMI-IR algorithm to calculates association values between the target word and each seed word, which is already described in Section 2.5. We prepared initial 12 seed word, 6 words are strongly positive seed words and 6 words are strongly negative seed words as in Table 3.4 for both sentiment dictionaries.

Table 3.4 List of Seed Words (6 positive words and 6 negative words)

Positive Seed Words	Negative Seed Words
เก่ง (Good at)	เกลียด (Hate)
รัก (Love)	โง่ (Stupid)
สุข (Happy)	ชั่ว (Evil)
ชอบ (Like)	ไม่ชอบ (Dislike)
ดี (Good)	ไม่ดี (Not good)
สวย (Beautiful)	เลว (Bad)

In addition, we also adopt two settings of SO-PMI-IR algorithm with "AND" operation and SO-PMI-IR algorithm with "NEAR" operation for comparison.

3.2.3.1 SO-PMI-IR algorithm with "AND" operation

"AND" operation counts the co-occurrence of the target word and a seed word within the same user comment, and the same polarity as the seed word is given to the target word. This is the simplest way but it also has the risk to miscount accidental co-occurrences within a comment as Figure 3.7.

Figure 3.7 Example of SO-PMI-IR with "AND" Operation

3.2.3.2 SO-PMI-IR algorithm with “NEAR” operation

“NEAR” operation counts the co-occurrence of the target word and a seed word within the pre-defined distance in a comment. In this research, the distance is decided as 10 words, which means that the case that the target word co-occurs with a seed word with the distance of more than 10 words is ignored for sentiment calculation as Figure 3.8.

Figure 3.8 Example of SO-PMI-IR with “NEAR” Operation

3.2.4 Sentiment Estimation

In this part is the polarity value calculation of user comment to the class label on whether a user comment is positive or negative is not available, and thus classification techniques cannot be applied. Instead, we propose a summative method with which the summation of the sentiment value of words in the constructed sentiment dictionary is regarded as the polarity of the text.

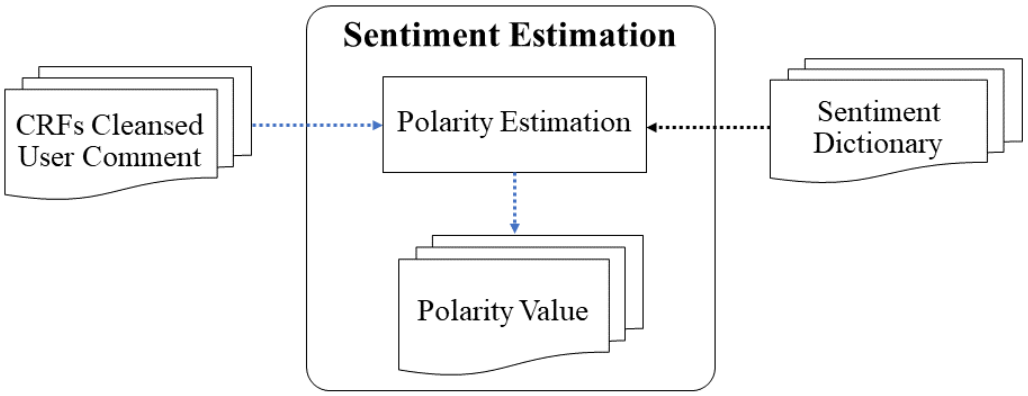


Figure 3.9 Sentiment Estimation

From Figure 3.9, we calculate the polarity value of user comments with the sentiment value of words in user comment sentiment dictionary that are constructed in Section 3.2.3. The calculation can be divided into two parts as in Figure 3.10: First, we add the sentiment values in the user comment sentiment dictionary and the sum is divided by the total number of the sentiment words. Second, we evaluate the sentiment value more than 0.3 as positive, between 0.3 and -0.3 as neutral, and less than -0.3 as negative.

|กรรม|ใด|ใคร|ก่อกรรม|นั้น|ตาม|สนอง|ใน|เร็ว|วัน|

| -0.53 |0.26| 0.34 | | 0.23 | -0.02 | -0.72 |0.32|-0.19| 0.19|

1 2 3 4 5 6 7 8 9

Positive = 0.26 + 0.34 + 0.23 + 0.32 + 0.19 = 1.34

Negative = -0.53 - 0.02 - 0.72 - 0.19 = -1.46

$$\frac{1.34-1.46}{9} = \frac{-0.12}{9} = -0.013$$

Figure 3.10 Estimate Sentiment Value of User Comments

3.2.5 Sentiment Classification

As in Figure 3.11, this part is the final classification of new articles according to each user is performed using three different machine learning methods: Decision tree, Naïve Bayes and SVM. Importantly we conducted the second experiment by applying the estimated value of the corresponding user comment that is estimated in Section 3.2.4 as the supervised class label, which enables this classification. We will discuss the effect of this method by comparing different conditions in feature vectors and machine learning method in Chapter 5.

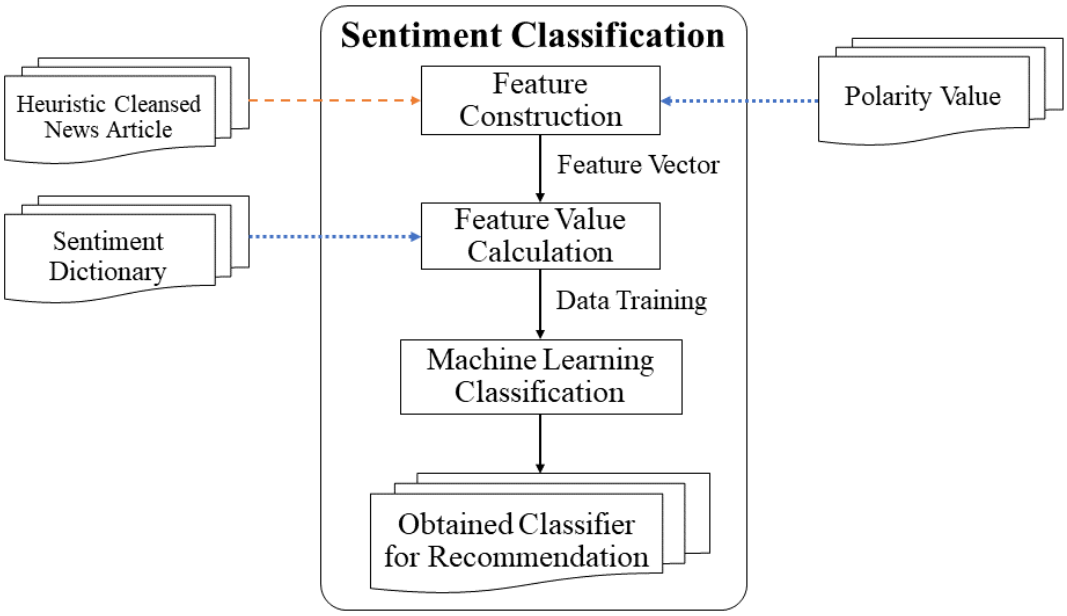


Figure 3.11 Sentiment Classification

Finally, we have the obtained classifier for recommend a new news article to each user according to the tendency of their past comments on past news articles.

3.3 Conclusion

In this chapter, we explained our research frame work that consists of 5 parts: Data collection, Data Normalization, Sentiment construction, Sentiment estimation and Sentiment classification.

CHAPTER 4

Sentiment Value Estimation of User Reviews

4.1 Introduction

In this chapter, a method to estimate the sentiment value of user comments is proposed. Before going into details, several difficulties should be pointed out.

First, user comments are written by ordinary people in a very informal manner and contain a great many of non-standard language use including non-standard or wrong spellings, slangs and other colloquial expressions, as well as HTML tags and other noises usually found in scraped data. Thus, a systematic way of strong noise reduction has to be devised. This is discussed from Section 4.2 Text Normalization.

Second, my dataset of user comments naturally does not contain the class label whether each comment is positive or negative, and due to the large size of dataset, manual tagging is impossible. Thus, the summative estimation of the sentiment value is the only applicable method, which requires a sentiment dictionary with which sentiment words are given their sentiment values. For other languages, there are several publicly available well-edited sentiment dictionaries, but as far as we know there is only one publicly available sentiment dictionary from PyThaiNLP (PyThaiNLP), but the number of entry words is approximately 930 (465 positive words and 465 negative words). As each user comment is usually very short, so if the number of entry words in the sentiment dictionary is small, it is likely that many user comments will not contain any sentiment words, which lead to the failure of sentiment calculation. Thus, a sentiment dictionary should be constructed by my own dataset, which is discussed in Section 4.3.

Third, based on the constructed sentiment dictionary, each user comment should be assigned with a sentiment value. This is discussed in Section 4.4.

The overview of this chapter is shown in Figure 4.1. Thus, the main aim of this chapter is to assign a sentiment value to each user comment, but the main discussion goes to how to remove noise more effectively, caused by the Thai writing difficulties.

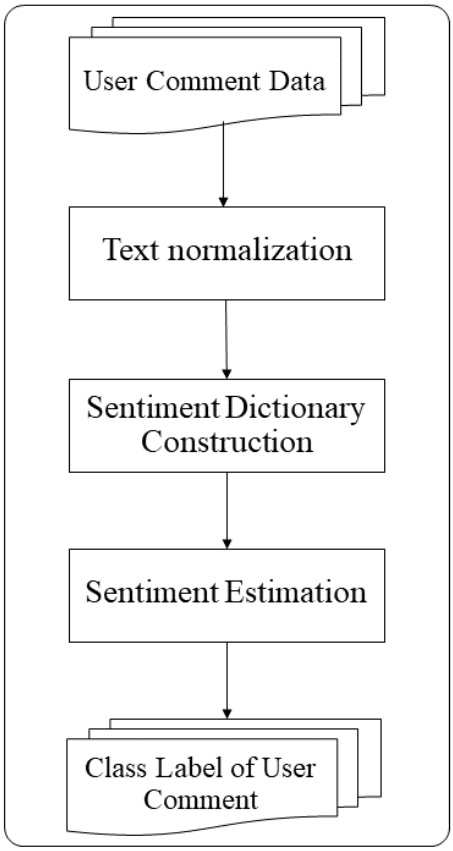


Figure 4.1 Overview of Text Normalization

From Figure 4.1, we used only user comment data to remove general noises by a heuristic approach, then we used a machine learning approach using CRFs to detect and remove other noises that are not correctly removed by the heuristic approach.

4.2 Text Normalization

In this research, the whole size of user comment dataset is shown in Table 3.2. The size of these dataset makes it almost impossible even just to check noise reduction done by computer. On the other hand, as mentioned in the previous section, user comment texts are very different from news article texts, and much more varieties of noises are found, which should be removed as much as possible for later processes. In former studies of Thai language processing, rule-based noise reduction has been applied, but the set of rules vary among studies, and they also manually corrected the data. Thus, in the following three sections, we will propose a hybrid approach to noise reduction using both heuristic rule-based corrections and machine learning-based CRFs noise detection. For this proposal, in order to make more intensive observations, we randomly chose 3,000 user comments for evaluation, which is discussed in Section 4.2.5. In this experiment, we focus on the textual part, so we removed the attached pictures, videos, URLs and spamming advertisements in these initial preprocesses. However, there is still a variety of noises which are almost 11.5% of the dataset: emoticons, special characters, character repetitions and other spelling variations and noises. All the subsequent tasks of this research—sentiment dictionary constructions, sentiment polarity estimation of user comments, and the sentiment classification of news articles—are conducted only with these noise reduction techniques and no manual corrections are conducted.

4.2.1 Heuristic Approach

The heuristic approach is the first part of my proposal for noise reduction, in which a set of rules are systematically designed according to human knowledge and experiences. After an extensive survey on what types of rules are employed in the past studies of Thai language processing, we designed rules for noise removal as discussed in the rest of this subsection. The procedure of heuristic noise reduction is schematized in Figure 4.2 consist of punctuation

removal, emotion and special character removal, repetition removal, word replacement and words replacement.

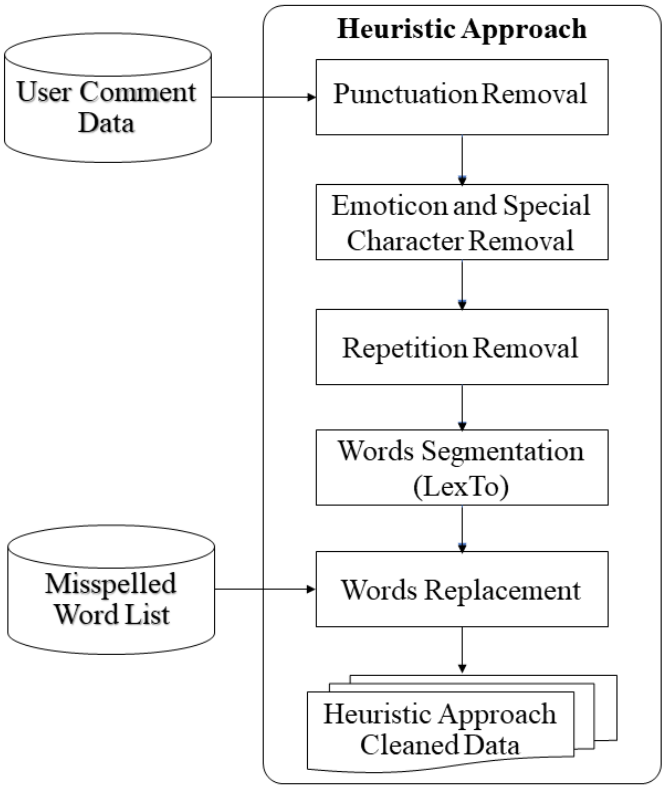


Figure 4.2 Heuristic Approach

4.2.1.1 Punctuation Removal

Punctuation removal is the first part to remove all Thai punctuation marks, Thai number and math symbol out of data (see all in Appendix A), as some examples are shown in Table 4.1.

Table 4.1 Punctuation Removal

Text with Punctuation	Punctuation Removal
น่าเบื่อมากๆๆๆๆๆๆๆๆๆๆ	น่าเบื่อมาก
โคตรขำ๕๕๕๕๕๕๕๕๕	โคตรขำ
คำเดียว "ริษยา"	คำเดียว ริษยา
ไล่ตามกันเป็นแถว555++	ไล่ตามกันเป็นแถว555
……เอิ่ม.....	เอิ่ม

4.2.1.2 Emoticon and Special Character Removal

Emoticons are the special types of text strings which are created to convey various feelings in a convenient way, whose examples are shown in Table 4.2. For sentiment analysis of social media, emoticons often play a crucial role in estimating the sentiment value of a text. For example, “It’s good :)” contains an emoticon “:)” and its feeling to be conveyed matches the positive sentiment word “good”. In this case, the use of “:)” just strengthens the positive value of this sentence. However, “It’s good (T_T)” contains an emoticon “(T_T)” whose sentiment is negative. In this case, the positive sentiment of “It’s good” is somehow canceled, the exact reason of which may vary according to the context. Thus, some studies focused on emoticons. However, in this research, my aim is to focus on text parts, and all emoticons and special symbols are removed, as in Table 4.2. Consideration of emoticons is a future issue. Thus, we remove emotion and special character in clouding the HTML character entities that we list in Appendix A out of datasets.

Table 4.2 Emotion and Special Character Removal

Text with Emotion and Special Character	Emotion and Special Character Removal
เขิลเลย(><)	เขิลเลย
❤ ❤ ❤ รักกัน ❤ ❤ ❤	รักกัน
✗ ✓ ว่ากันที่ศาล	ว่ากันที่ศาล
พูดแล้วอยากจะT-T	พูดแล้วอยากจะ
สาธุ™	สาธุ

4.2.1.3 Repetition Removal

As the Thai language writing system has no distinction between capital and small letters, people often duplicate, or repeat, some characters to convey different degrees of feeling. A typical example is “มาากก” for “มาก” or “ไปปปปปป” for “ไป”. In social media texts, this type of non-standard reduplication of characters are widely observed. In this subsection, all the types of this type to be removed are discussed according to my survey.

1) Consonant Repetition Removal

We replaced the same consonants that the character repeated more than two consonants to the same two characters, because the Thai language has many words that are spelled with the same consonant repeated, but such duplication is no more than two. Note however that the initial removal of this type still leaves some wrongly spelled words, but a set of simple rules cannot remove all these errors, so some of them are expected to be removed with the later step of machine learning noise reduction. Some examples of this type of noise reduction are shown in Table 4.3.

Table 4.3 Consonants Removal

Standard Spelling	Redundant	Repetition
กินนน้ำร้อน	กินนนนน้ำร้อน	กินน้ำร้อน
คนมีเงิน	คนมีเงินนนนนนนนนนน	คนมีเงินน
นายกรัฐมนตรี	นายยยยยยยยกรัฐมนตรี	นายยกรัฐมนตรี
เบื่อ	เบื่อออออออ	เบื่ออ
บอกก่อน	บบบบบบบบบบบบบอกก่อน	บบอกก่อน
มากกว่า	มากกกกกกกกว่า	มากกว่า

2) Vowel and Tone Mark Repetition Removal

Vowel and tone mark should be handled differently from consonants because, unlike consonants, they are never repeated in the Thai writing system. However, the actual data contain many repetitions of them, partly because of emphatic intention and partly because of misspelling. Some examples are shown in Table 4.4.

Table 4.4 Vowel and Tone Mark Removal

Standard Spelling	Redundant	Repetition
ไม่ไป	ไไไไไม่ไป	ไม่ไป
ค่าเงินบาท	ค่าเงิินบาท	ค่าเงินบาท
คุณครู	คุณครูู	คุณครู
คร๊า	คร๊าาาาาาาาาาา	คร๊า
ไม่	ไม่่	ไม่
กินข้าว	กินข้้าว	กินข้าว
สงกรานต์	สงกรานต์์	สงกรานต์

4.2.1.4 Word Segmentation

Some words are commonly misspelled, partly due to the Thai keyboard configuration. We cannot correct all the misspelled words by rules, and too strong a rule set will also produce wrong corrections. However, common misspelling is corrected in a mild way. In this part, we employed LexTo “Thai lexeme Tokenizer” (National Electronics and Computer Technology Center: NECTEC, 2016) to tokenize each word in our dataset as examples are shown in Table 4.5.

Table 4.5 Examples of Word Segmentation

Original Words	Segmented Words
เก่งครับ	|เก่ง|ครับ|
ได้คืบจะเอาศอกแล้วบอกว่าพอเพียง	|ได้คืบจะเอาศอก|แล้ว|บอก|ว่า|พอเพียง|
มีการสอบด้วยเหรอ	|มี|การ|สอบ|ด้วย|เหรอ|
ไม่คุ้มกับชีวิต	|ไม่คุ้ม|กับ|ชีวิต|
สงครามศาสนาย่อม	|สงครามศาสนา|ย่อม|
สักคนเดียวก็ไม่ได้ มึงอยากได้ ต้องเลือกตั้ง	|สัก|คนเดียว|ก็ไม่ได้| |มึง|อยากได้| |ต้อง|เลือกตั้ง|

4.2.1.5 Word Replacement

In this part, we manually created list of frequently misspelled words as Appendix B is used to find and replace them to the correct spelling, as in Table 4.6. Some research tried to add more rules to remove noise but it needs many rules due to the variety of spelling found in our data. So, instead of preparing a lengthy list of commonly misspelled words, misspelling replacement is done mildly and the rest will be treated by the machine learning approach that follows.

Table 4.6 Examples of Wrong Spelling

Wrong Spelling	Standard Spelling
ก้อ	ก็
รสชาด, รดชาด	รสชาติ
ปรากฏ, ปรากด, ปากด	ปรากฏ
เวนกำ	เวรกรรม
ม้ายช่าย, ไม่ไช่	ไม่ใช่

4.2.2 Noise Validation

Before going to the second step of the machine learning approach, we need the correct labeling of our dataset, because my proposal of machine learning is supervised learning, which needs the correct labeling. So, we create noise list for labeled noise data as Figure 4.3

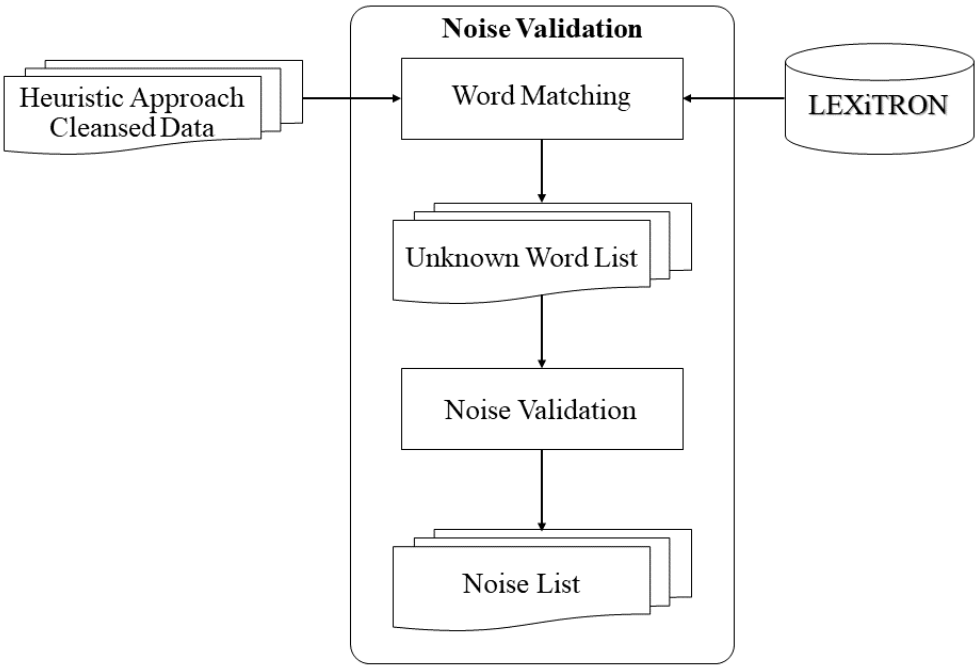


Figure 4.3 Noise Validation

As shown in Figure 4.3, we made matching and tagging of words with LEXiTRON dictionary (LEXiTRON, 2016) for words that are not found in the dictionary. Some examples are shown in Table 4.7.

Table 4.7 Examples of Unknown Words

กลุ่มบุคคล	การบวช	กาสาวพัสตร์	ก้าม๊ง
ครัง	คว่ม	คุณ ทักษิณ	คู่
งม้น	จริงจริ๊ง	จะเป็นจะตาย	ชกำ
ชี้าว	ต้าน	ตำ่	ตายแล้ว
ทธิ์	มีเชี้	มี้ย	ยงั้น
หมาเน่า	หรือหว่	ห์	อย่าคิดว่า
เป็นจจขจจ่	เผด็	เฟ็	เอาสีข้างเข้าถู

Then, we manually checked the unknown word list to judge whether they are noises or not. Because some words that are not found in LEXiTRON dictionary are words that newly appeared and have not been included in the dictionary yet. Examples of this type are shown in Table 4.8.

Table 4.8 Examples of Noise List

กก่	ขั่า	คู่	ค์ะ
คว่ม	งงาแบ	งม้น	งอี
จน์	ชี้าว	ชกำ	ด้
ต้าน	ตลุะง	ตำ่	ตอ้ง
ทธิ์	นำ้	เป็นจจขจจ่	เฟีย่ว
มีเชี้	มี้ย	มุะง	ยงั้น
ไาส	ไห้วแลว้	ห์า	หรือหว่

4.2.3 Character Based CRFs Approach

This section, we used heuristic approach cleaned dataset to detected noise with machine learning approach, because our dataset still remains various noise mostly spelling errors. So, this study we apply the character-based Conditional Random Fields (CRFs) to detect noises without dictionary as Figure 4.4.

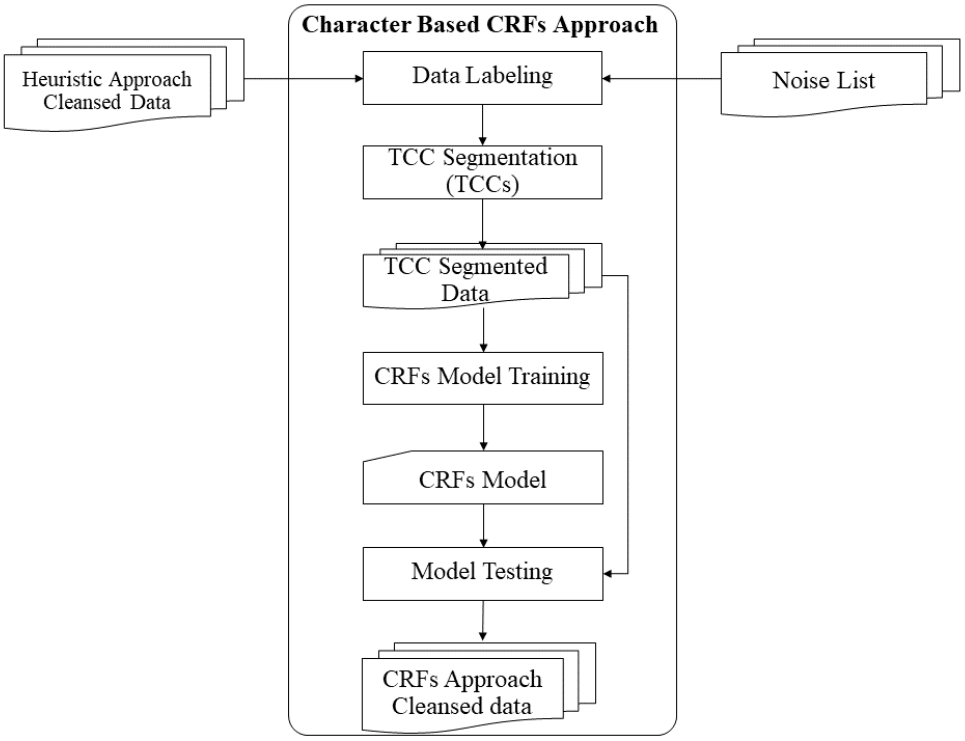


Figure 4.4 Character Based CRFs Approach

First step of noise reduction, we construct a model for tagging noise word from the noise list in the previous section to user comments that are cleaned with a heuristic approach. The parts that are judged as noises are tagged with <N> and </N> as shown in Table 4.9.

Table 4.9 Noise Labeling

ไม่มีส่วน<N>เ◌ี้กี่ยว</N>ข้อง
มี<N>เงน</N>มากกว่าเยอะ
<N>กิน่</N>ข้าวบ้างน่ะ
คุณยายน่ารักจัง<N>า</N>เลย
เล่นน้ำ<N>สงกราต์</N>
<N>สาฐู</N>
หน่วยงาน<N>รัดบานะ</N>

For the second step of noise reduction, we propose the use of Conditional Random Fields (CRFs) as a machine learning method because CRFs are designed for sequential labeling and noise detection with machine learning can be regarded as sequential labeling to find the error labels. The mechanism of CRFs is already described in Section 2.9.4.

As CRFs are one of the supervised learning, so the correct sequential labeling must be prepared. From the procedure of the last subsection, we first employed Thai character cluster (TCCs) that described in Section 2.3 to tokenized the noise labeled data as shown in Table 4.10 and is expected to improve the performance of this method.

Table 4.10 TCCs Training Data

|ไม่|มี|ส่|ว|น|<N>|เ|◌ี้|กี่|ย|ว|</N>|ข้|อ|ง|
|มี|<N>|เง|น|</N>|มา|ก|ก|ว่า|เย|อะ|
<N>|กิ|น่|</N>|ข้า|ว|บ้า|ง|น่ะ|
|คุ|ณ|ยา|ย|น่า|รัก|จัง|<N>|า|</N>|เล|ย|
|เล่|น|น้ำ|<N>ส|ง|ก|รา|ต์|</N>
<N>|สา|ฐู|</N>
|ห|น่|ว|ย|งา|น|<N>|รัด|บา|นะ|</N>

Then, we convert the data as shown in Table 4.10 into the form of the training data as shown in Table 4.11.

Table 4.11 CRFs Training Data

Spelling with TCCs	W (Correct) or N (Noise)	Location within Words
ไม่	W	B-W
มี	W	I-W
ส่	W	B-W
ว	W	I-W
น	W	I-W
เ้	N	B-N
กี่	N	I-N
ย	N	I-N
บ	N	I-N
ข้	W	B-W
อ	W	I-W
ง	W	I-W
.	.	O

In Table 4.11, the first column is TCC-based expressions, the second column is the actual class: “W” (correct word) or “N” (noise), and the third column is the location within words: “B-W,” the beginning character of a correct word, “I-W,” an intermediate character in a correct word, “B-N,” the beginning character of a noise part, “I-N,” an intermediate character in a noise part, and “O,” the end of a sequence.

In this research, we used CRFsuite, developed by Okazaki (2007) for construct CRF model. Before constructing a CRFs model, we also reformat the training data into the data

format for CRFsuite including the window size of characters to be processed, the example of 3 window size as shown in Figure 4.5, where W is the target character and the number specifies the number of the characters to be considered. The window size is set from 3 to 15 characters. Finally, we constructed CRF noise detection models and trained this model with 5-fold cross-validation.

```
templates = (
    (('w', -2), ),
    (('w', -1), ),
    (('w',  0), ),
    (('w',  1), ),
    (('w',  2), ),
    (('w', -1), ('w',  0)),
    (('w',  0), ('w',  1)),
    (('w', -2), ('w', -1), ('w',  0)),
    (('w', -1), ('w',  0), ('w',  1)),
    (('w',  0), ('w',  1), ('w',  2)),
    (('w', -2), ('w', -1), ('w',  0), ('w',  1)),
    (('w', -1), ('w',  0), ('w',  1), ('w',  2)),
    (('w', -2), ('w', -1), ('w',  0), ('w',  1), ('w',  2)),
)
```

Figure 4.5 Modified Input Template

4.2.4 Evaluation of Character-Based CRFs Method

For the evaluation of the proposed application of CRFs to noise detection, experiments are conducted with the context window sizes from 3 to 15 characters, and the result is shown in Table 4.12.

Table 4.12 CRFs Compared Different Context Window Size

Context Window Size	Precision	Recall	F1 measure
3	0.9548	**0.7432**	**0.7882**
5	0.9608	0.7268	0.7694
7	0.9624	0.7120	0.7503
9	**0.9625**	0.7038	0.7396
11	0.9602	0.6976	0.7313
13	0.9584	0.6929	0.7250
15	0.9567	0.6883	0.7187

As of precision, the window size of 9 characters achieved the best while, the window size of 3 characters achieved the best as of recall and F1 measure. One possible reason for this is that mostly noises that appear in our dataset come from the wrong position of just one character, because I already removed character repetitions by the heuristic approach that precedes the machine learning step. Our investigation found that most errors occurred with the wrong position of vowel and tone marks.

The effect of the proposed noise reduction should be evaluated from the purpose of the task because different tasks require different levels of correctness of the previous procedure. In this research the purpose of noise reduction is to find more sentiment words both in sentiment dictionary construction and sentiment estimation. In the next section, I made another evaluation from this perspective.

4.2.5 Evaluation with Sentiment Dictionary Construction

As mentioned in the last section, even when we achieve a better noise reduction, it may not be appreciated if the subsequent procedure does not receive benefits from it. In this research, the subsequent procedure is sentiment dictionary construction and sentiment estimation, particularly for user comments. So, I conducted another experiment to evaluate the usefulness of

my noise reduction by constructing sentiment dictionaries with 3,000 comments under different settings.

4.2.5.1 Sentiment Dictionary Construction

As described in Section 2.7, my sentiment dictionary construction is based on SO-PMI-IR. The procedure is schematized in Figure 4.6. First, the sentiment value of target words is calculated with their co-occurrences with each seed word with PMI and SO algorithm. Second, the most frequent word of positive and negative values are chosen to add in the seed word list. In this research, this process is itinerated to produce more sentiment words to be included in the sentiment dictionary. So, the calculation is repeated for 15 rounds. The result of the choices of seed words are used to evaluate the usefulness of noise reduction.

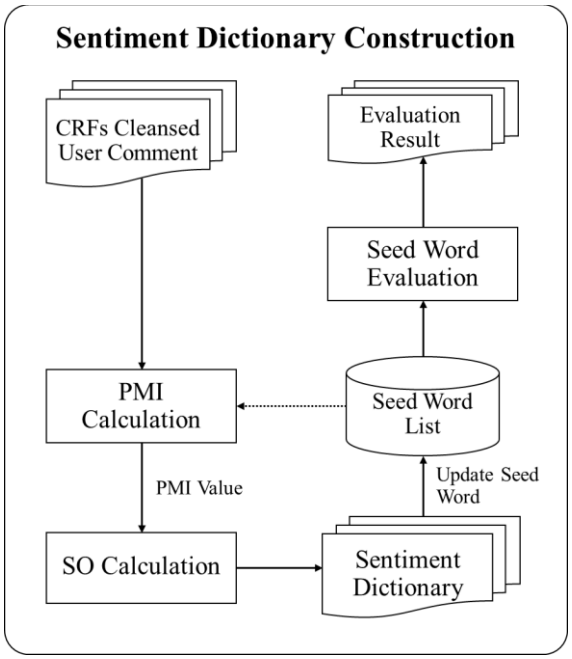


Figure 4.6 Sentiment Dictionary Construction

4.2.5.2 Experiment of Noise Reduction Technique

Three user comment sentiment dictionaries are constructed for evaluation with differently preprocessed dataset: 1) Dataset A: data without noise reduction. 2) Dataset B data cleaned only with a heuristic approach. and 3) Dataset C: data cleaned with both heuristic and CRF approaches. As already mentioned, the calculation is itinerated for 15 times, which means that 30 words are chosen for additional seed words, which are compared to the manually tagged sentiment value. The reason to compare only 30 words instead of all the sentiment words to be included in the sentiment dictionary is that as the dataset for this experiment is a small dataset of 3,000 user comments, and that it is not enough to create a sentiment dictionary of high quality, together with a practical reason that it is almost impossible to manually tag the sentiment value of thousands of words.

4.2.5.3 Result and Discussion

As described in the previous subsection, the evaluation is conducted with only 30 words chosen at the iterative rounds of calculation of PMI and SO algorithm.

Table 4.13 Evaluation Seed Words Result

	P(P)	**P(N)**	**R(P)**	**R(N)**	**F(P)**	**F(N)**	**Accuracy**
Dataset A	0.3467	0.8800	0.9866	0.3099	0.5000	0.4583	0.4800
Dataset B	0.4182	0.8667	0.7931	0.5493	0.5476	0.6724	0.6200
Dataset C	**0.4561**	**0.9302**	**0.8966**	**0.5634**	**0.6047**	**0.7018**	**0.6600**

The result is shown in Table 4.13:

- P(P) is precision of positive words
- P(N) is precision of negative words
- R(P) is recall of positive words
- R(N) is recall of negative words

- F(P) is F-measure of positive words
- F(N) is F-measure of negative words
- Accuracy.

For all the correctness values, Dataset C with a hybrid approach of noise reduction achieved the best result, which means that a stronger noise reduction indeed improves the sentiment dictionary construction. On the other hand, the precision of positive words is much lower than the precision of negative words under all conditions. This is caused by the data imbalance: most of the comments in the dataset were negative, and thus the occurrences of positive words are much lower than expected. However, it does not mean that the sentiment dictionary construction using the whole dataset also suffers from the same issue. For the sentiment dictionary construction described in the next section, the whole set of user comments is used. In principle, data imbalance is observed as in the small dataset in this section, but the whole dataset is much larger and it is expected, and indeed it turned out to be the case, that the large size of the dataset produced more reliable sentiment values for many words.

4.3 Constructed Sentiment Dictionary with Larger Data

In this section, we construct the sentiment dictionary from our whole user comments data with noises removed by my hybrid noise reduction. As already mentioned, this research employs SO-PMI-IR algorithm to calculate the sentiment value of each word, and it allows several different settings: SO-PMI-IR with “AND” operation and SO-PMI-IR with “NEAR” operation

Finally, two sentiment dictionaries are constructed for user comments based on the whole set of user comments that use for class label in sentiment classification part, and another set of two news article sentiment dictionaries are also constructed for news articles based

on the whole set of news articles use to represent the feature value, which is described in Section 5.4.3. In Table 4.14, the organization of these sentiment dictionaries are shown: a user comment sentiment dictionary constructed with "AND" operation for user comments, a user comment sentiment dictionary constructed with "NEAR" operation for user comments, a news article sentiment dictionary constructed with "AND" operation for news articles, and a news article sentiment dictionary constructed with "NEAR" operation for news articles. Note that the total number of words is slightly higher than the sum of the numbers of positive and negative words because we remove the words that appear in less than 3% of the user comments in the dataset, because they are very infrequent and thus rarely contribute to sentiment classification. Some of them are noises that are missed by noise reduction steps. Some others are dialectal or slangs that are not often used.

Table 4.14 Summarization of Sentiment Dictionary

Sentiment Dictionary	**Total Number of Words**	**Number of Positive Words**	**Number of Negative Words**
User Comment Sentiment Dictionary **(AND)**	2626	2353	97
User Comment Sentiment Dictionary **(NEAR)**	2626	1503	153
News Article Sentiment Dictionary **(AND)**	1506	258	163
News Article Sentiment Dictionary **(NEAR)**	1506	62	1353

4.4 Sentiment Value Estimation of User Comment

In this section, we calculate the polarity value of user comments with both the user comment sentiment dictionary (AND) and the user comment sentiment dictionary (NEAR). By summing up the sentiment values of the sentiment words that appear in a user comment and dividing it by the total number of sentiment words in the user comment, the polarity of a user comment is estimated as in Section 3.2.4. Then, we bring the result of positive, negative and neutral to create the class label of each user comment to a news article as in Table 4.15 (P: Positive, N: Negative and NE: Neutral).

Table 4.15 Class Label of User Comment to News Article

	$Article_1$	$Article_2$	$Article_3$	$Article_4$	$Article_5$	…	$Article_n$
$User_1$	-	P	NE	N	P	…	N
$User_2$	N	P	-	P	P	…	-
$User_3$	-	NE	N	P	-	…	NE
$User_4$	N	-	P	-		…	N
$User_5$	P	N	NE	N	N	…	-
$User_6$	-	P	NE	P	P	…	P
$User_7$	NE	-	-	NE	-	…	N
…	…	…	…	…	…	…	…
$User_m$	P	NE	NE	-	NE	…	P

4.5 Conclusions

In this section, we do the first experiment for compared the effectiveness of each noise reduction technique, which the proposed hybrid noise reduction technique combiining a heuristic approach and a CRF-based machine learning approach is the best technique. Then, we applied this hybrid noise reduction technique with the whole user comment data and constructed two user sentiment dictionaries with SO-PMI-IR algorithm with “AND” operation and SO-PMI-IR algorithm with “NEAR” operation, which used for sentiment value calculation of user comment for class label in sentiment classification part in Chapter 5. In addition, we also construct two news article sentiment dictionaries by the same technique with a user comment sentiment dictionary, which is used to construct feature values in Chapter 5.

CHAPTER 5

Machine Learning Classification of News Articles Based on User Comment Sentiment Value

5.1 Introduction

In this chapter, we will conduct machine learning classification of news articles according to each user, based on the estimation of the sentiment value of user comments to them, which was presented in the previous chapter. As this news article classification is the basis for customized recommendation of future news articles to each user, classification is conducted for each user. We compare three machine learning methods: Decision Tree, Naïve Bayes, and Support Vector Machines, and the average accuracy, precision, recall, and F1-measure among users are compared among these machine learning methods. We also conduct experiments to compare four feature vectors of training data: word sentiment value with a public sentiment dictionary, two versions of word news article sentiment values with my constructed sentiment dictionary based on the data, and tf-idf. For my constructed sentiment dictionaries, two versions are constructed based on neighboring co-occurrences with sentiment seed words ("NEAR") and on simple co-occurrences with sentiment seed words within each data entry ("AND").

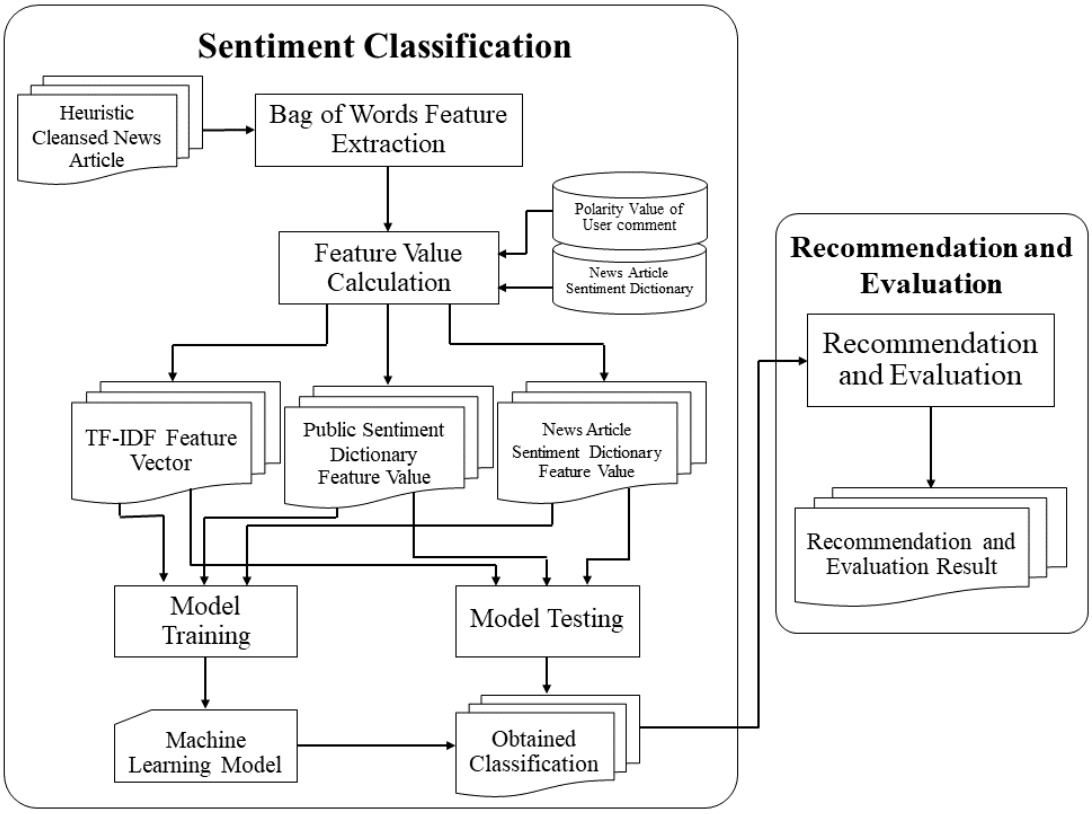


Figure 5.1 Sentiment Classification

5.2 Experiment Setting

The experimental data are constructed with two sources. The main part of the data for classification is the news articles, and their supervised label is the estimation of the user's sentiment value to them, which are described in the previous chapter. News articles need noise reduction, but the necessary noise reduction is different from user comments, described in Section 4.2, because, unlike user comments, news articles are professionally written and checked, and do not contain emoticons, special characters, and various non-standard spelling variations in principle. On the other hand, noises remain coming from data scraping processes such as HTML tags, URLs, pictures and videos. These noises are the target of heuristic approaches described in

Section 4.2.1, and thus we used only the related heuristic noise reduction but not CRF-based noise reduction for news article data.

I have to select the target users for experiments. However, some consideration is necessary. First, the target users must have enough number of comments to news articles for machine learning training. Second, from manual observations, positive sentiments are on average found much more than negative sentiments in user comments, which causes data imbalance issues for machine learning. Thus, we randomly choose 25 Facebook users whose comment posting exceeds 50 even after balancing the proportion of positive and negative comments within 20% difference, in other words, 60% v. 40% at most. More concretely, if the total number of comments of User A is 400 but with 360 positive and 40 negative comments, for example, at least 300 positive comments are randomly ignored for data balancing, and finally only 100 comments (60 positives and 40 negatives) are to be selected.

For these 25 Facebook users, the sentiment value of each comment is estimated in the way described in Section 4.4. For this estimation, two sentiment dictionaries are constructed for comparison: construction based on neighboring co-occurrences with sentiment seed words ("NEAR") and on simple co-occurrences with sentiment seed words within each data entry ("AND"). Note that the sentiment dictionary employed is constructed from all the user comments in my data and no customization is made according to each user. Although we admit that, according to sex, age, etc., the sentiment value of each word may vary among users, we assume that such difference is small enough to be ignored. We also ignore the possibility that the sentiment value of a word has a variation according to topics. As far as we know, these two ignorances are usually adopted in other research, and a pursuit of finer treatment of these two issues will be investigated in future.

For each target user, the corresponding news articles to their user comments are selected for classification experiments. Thus the total number of news articles is different among target users, but not less than 50 articles.

5.3 Feature Extraction

Each news article, after noise reduction preprocess to remove HTML tags and others, is processed for word segmentation with LexTo, and is converted into a bag-of-words vector, as shown in Table 5.1. The size of vector is 40,005; in other words, the total number of words in the whole set of news articles for all target users is 40,005 words.

Table 5.1 Bag-of-Words Data Form

(The number in the table represent the total number of occurrences in an article, just for example)

	กงจักร	กงล้อ	กงหรา	กฎกติกา	กฎหมู่	…	ไฮโล
Article$_1$	13	3	0	0	0	…	0
Article$_2$	0	0	4	0	0	…	0
Article$_3$	0	0	0	0	8	…	10
Article$_4$	2	0	0	0	0	…	0
Article$_5$	0	1	0	3	2	…	0
…	…	…	…	…	…	…	…
Article$_n$	0	0	0	17	5	…	0

Under this bag-of-words model, we employ three different features for comparison: TF-IDF, the sentiment value in a publicly available sentiment dictionary, and the sentiment value in my constructed news article sentiment dictionary based on news article data.

5.4 Feature Value Calculation

In this experiment, we compare three feature vectors to machine learning classification. So, I calculate feature value form 3 techniques as follows below.

5.4.1 TF-IDF

We calculated TF-IDF of words that appeared in each news article and we calculated IDF value from whole news article example as Table 5.2. As "W_i" is each word that appeared in news article, "A_j" is a news article in the dataset for a target user, and the float number is a TF-IDF value. The column "Class" is the class label of its polarity value that is estimated based on the corresponding user comments. Thus importantly, the value of "Class" of a news article will be different among target users, which is why classification is conducted according to each target user, while TF-IDF values are the same among target users because TF-IDF values are calculated based on the whole dataset, but not on the dataset of each target user.

Table 5.2 TF-IDF Feature Vector

	W_1	W_2	W_3	W_4	W_5	W_6	…	W_n	Class
A_1	0.0959	0.0140	0.0988	0.0184	0.534	0.1063	…	0.0708	N
A_2	0.1034	0	0.0533	0.0458	0.1308	0	…	0.0332	P
A_4	0.0421	0.1027	0.0537	0.0771	0.0682	0.0734	…	0	N
A_5	0.0232	0.748	0.0189	0	0.0251	0.0111	…	0.1529	P
…	…	…	…	…	…	…	…	…	…
A_m	0.1111	0.1230	0.0685	0	0.0298	0.0956	…	0.0256	P

5.4.2 Public Sentiment Dictionary Feature Vector

The second feature is based on a public sentiment dictionary (PyThaiNLP) This is a sentiment dictionary for approximately 934 Thai words with their polarity value (467 positive words and 467 negative words). The feature vector of each news article is then constructed with these values, as in Table 5.3.

Table 5.3 Public Sentiment Dictionary Feature Vector

	W_1	W_2	W_3	W_4	W_5	W_6	…	W_n	Class
A_1	N	P	N	-	P	N	…	N	**N**
A_2	-	P	-	P	P	-	…	-	**P**
A_4	-	P	N	-	P	N	…	N	**N**
A_5	-	P	N	-	-	-	…	-	**P**
A_6	N	-	N	P	P	N	…	N	**P**
…	…	…	…	…	…	…	…	…	…
A_m	-	P	N	P	-	P	…	N	**P**

5.4.3 News Article Sentiment Dictionary

Two news article sentiment dictionaries constructed as in Section 4.3, each of which is constructed with the whole news article dataset with SO-PMI-IR technique with different operations: “AND” operation and “NEAR” operation.

5.4.3.1 News Article Sentiment Dictionary with “AND” Feature Vector

We represented the positive and negative values as P and N, respectively, and neutral values are not counted. Thus, each news article is showed as in Table 5.4.

Table 5.4 News Article Sentiment Dictionary with "AND" Feature Vector

	W_1	W_2	W_3	W_4	W_5	W_6	…	W_n	Class
A_1	N	P	P	-	-	P	…	P	**N**
A_2	-	P	-	P	-	-	…	P	**P**
A_4	N	P	P	-	N	P	…	P	**N**
A_5	-	P	P	P	-	P	…	-	**P**
…	…	…	…	…	…	…	…	…	…
A_m	-	P	P	P	-	P	…	-	**P**

5.4.3.2 News Article Sentiment Dictionary with "NEAR" Feature Vector

We represented the positive and negative values as P and N, respectively, and neutral values are not counted. Thus, each news article is showed as in Table 5.5.

Table 5.5 News Article Sentiment Dictionary with "NEAR" Feature Vector

	W_1	W_2	W_3	W_4	W_5	W_6	…	W_n	Class
A_1	N	P	N	-	P	N	…	P	**N**
A_2	-	P	-	N	P	-	…	P	**P**
A_4	N	P	N	-	P	N	…	P	**N**
A_5	-	P	N	N	-	N	…	-	**P**
…	…	…	…	…	…	…	…	…	…
A_m	-	P	N	N	-	N	…	-	**P**

5.5 Classification with Machine Learning

In this experiment, we chose 3 machine learning methods for evaluation and comparison: Decision Tree, Naïve-Bayes, and SVM. For **decision tree**, we used CART (Classification and Regression Trees) algorithm for decision tree classification and Gini impurity for measures of node impurity with 3 feature vectors. For Naïve-Bayes, we used Multinomial model with the training dataset with TF-IDF and used Bernoulli models with the training data set with a public sentiment dictionary and a news article sentiment dictionary because Multinomial model is suitable to fractional count feature or TF-IDF, while Bernoulli models is suitable to binary or Boolean features. And SVM, we used linear for kernel function.

Then, we made a random separation of the supervised data into two parts: 80% for training set and 20% for testing set as Figure 5.2. Then, we bring this data to classification with three machine learning methods, whose results is explained in Section 5.5.

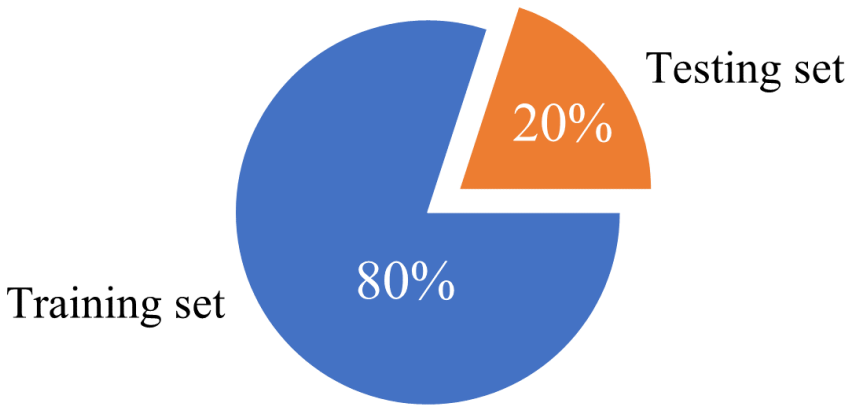


Figure 5.2 Random Split Data

5.6 Evaluation of Machine Learning Classification

In this section, the evaluation result of classification is shown with 3 machine learning models with four different feature vector sets. We used the confusion matrix table to calculate the effectiveness of experiment: Precision, Recall. F-measure and Accuracy.

In this research, we proposed two sentiment dictionary construction techniques by calculating the SO-PMI-IR equation with "AND" operation and "NEAR" operation shown in Table 5.6, and we constructed two user comment sentiment dictionaries from user comment data and two news article sentiment dictionaries from news article data with both "AND" operation and "NEAR" operation explained in Section 3.2.3.

Table 5.6 All Sentiment Dictionary in This Research

	User Comment Sentiment Dictionaries	**News Article Sentiment Dictionaries**
Dataset	User comment data	News article data
Calculation Methods	(1) SO-PMI-IR with "AND" operation (2) SO-PMI-IR with "NEAR" operation	(1) SO-PMI-IR with "AND" operation (2) SO-PMI-IR with "NEAR" operation
Purpose	Estimation polarity value of user comment value to be used as class labels of corresponding news articles	As features of news articles for classification of news articles (according to each user)

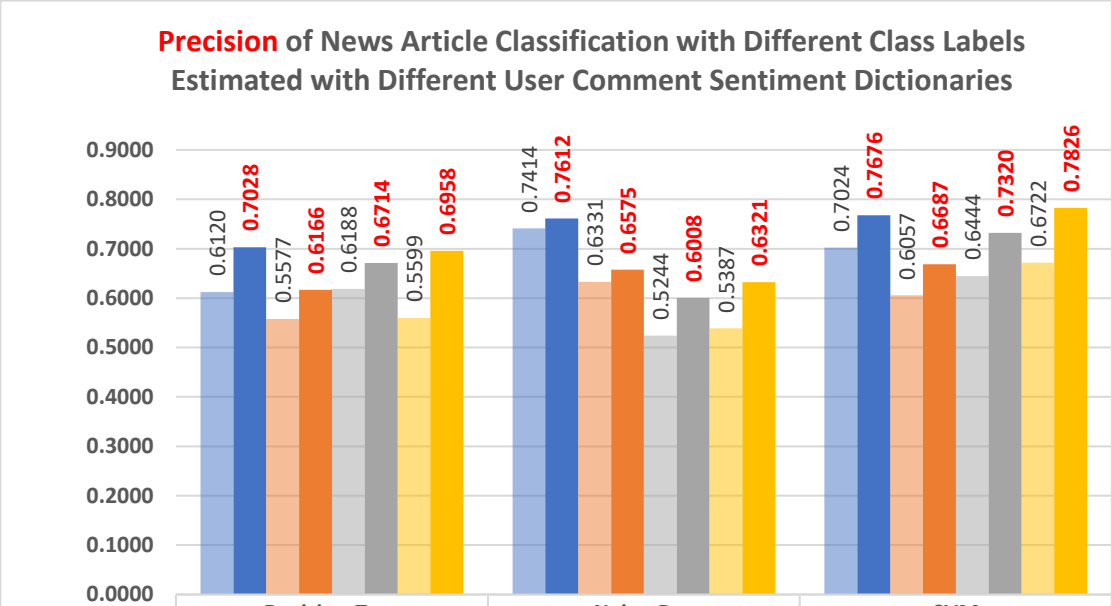
However, we used both user comment sentiment dictionaries to estimate polarity of comments for class label in this experiment. So, in this classification, we have two class labels for compared result as follow.

5.6.1 Comparison of User Comment Sentiment Dictionary Creation Technique

In this section, we compared the user comment sentiment creation technique: SO-PMI-IR with two operations ("AND" and "NEAR"). So, in this experiment, we used both user

comment sentiment dictionaries to estimate the polarity value of user comments that expressed to class label. The experiment result of both user comment sentiment dictionary is shown below:

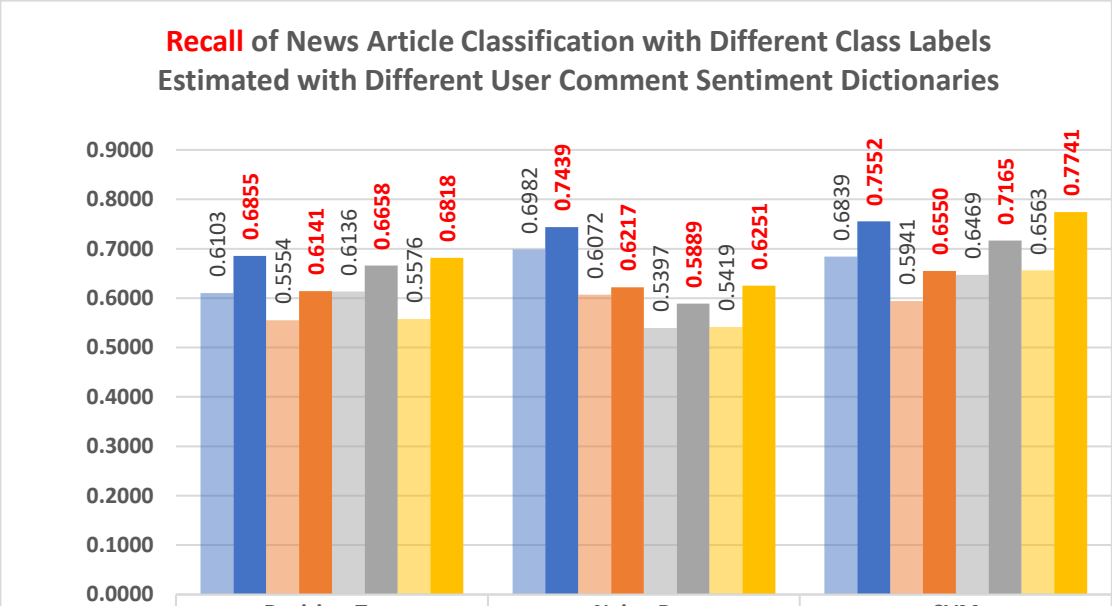
- TF-IDF (A): TF-IDF feature vector with class label of user comment sentiment dictionary “AND” operation.
- TF-IDF (N): TF-IDF feature vector with class label of user comment sentiment dictionary “NEAR” operation.
- Public (A): Public sentiment dictionary feature vector feature vector with the class label of the user comment sentiment dictionary “AND” operation.
- Public (N): Public sentiment dictionary feature vector feature vector with the class label of the user comment sentiment dictionary “NEAR” operation.
- News1 (A): News article sentiment dictionary (AND) feature vector with the class label of the user comment sentiment dictionary “AND” operation.
- News1 (N): News article sentiment dictionary (AND) feature vector with the class label of the user comment sentiment dictionary “NEAR” operation.
- News2 (A): News article sentiment dictionary (NEAR) feature vector with the class label of the user comment sentiment dictionary “AND” operation.
- News2 (B): News article sentiment dictionary (NEAR) feature vector with the class label of the user comment sentiment dictionary “NEAR” operation.



	Decision-Tree	Naive-Bay	SVM
TF-IDF (A)	0.6120	0.7414	0.7024
TF-IDF (N)	0.7028	0.7612	0.7676
Public (A)	0.5577	0.6331	0.6057
Public (N)	0.6166	0.6575	0.6687
News 1 (A)	0.6188	0.5244	0.6444
News 1 (N)	0.6714	0.6008	0.7320
News 2 (A)	0.5599	0.5387	0.6722
News 2 (N)	0.6958	0.6321	0.7826

Figure 5.3 Precision of User Comment Sentiment Dictionary

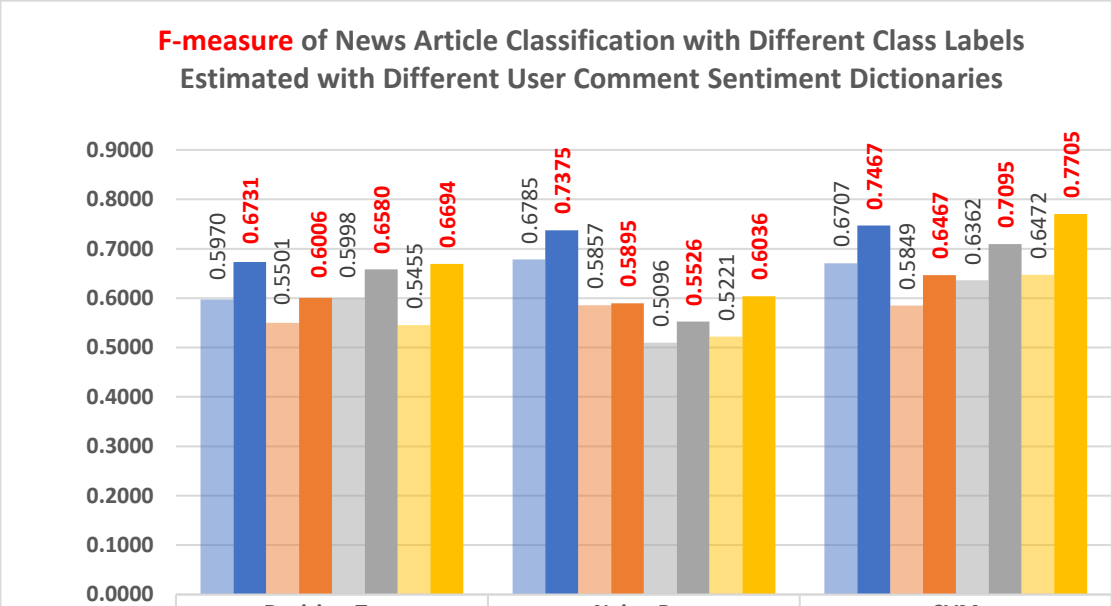
As in Figure 5.3, the result showed the precision of news article classification with class labels estimated with user comment sentiment dictionaries “NEAR” operation is higher than class labels estimated with user comment sentiment dictionaries “AND” operation in all machine learning methods and with all feature vectors.



	Decision-Tree	Naive-Bay	SVM
TF-IDF (A)	0.6103	0.6982	0.6839
TF-IDF (N)	0.6855	0.7439	0.7552
Public (A)	0.5554	0.6072	0.5941
Public (N)	0.6141	0.6217	0.6550
News 1 (A)	0.6136	0.5397	0.6469
News 1 (N)	0.6658	0.5889	0.7165
News 2 (A)	0.5576	0.5419	0.6563
News 2 (N)	0.6818	0.6251	0.7741

Figure 5.4 Recall of User Comment Sentiment Dictionary

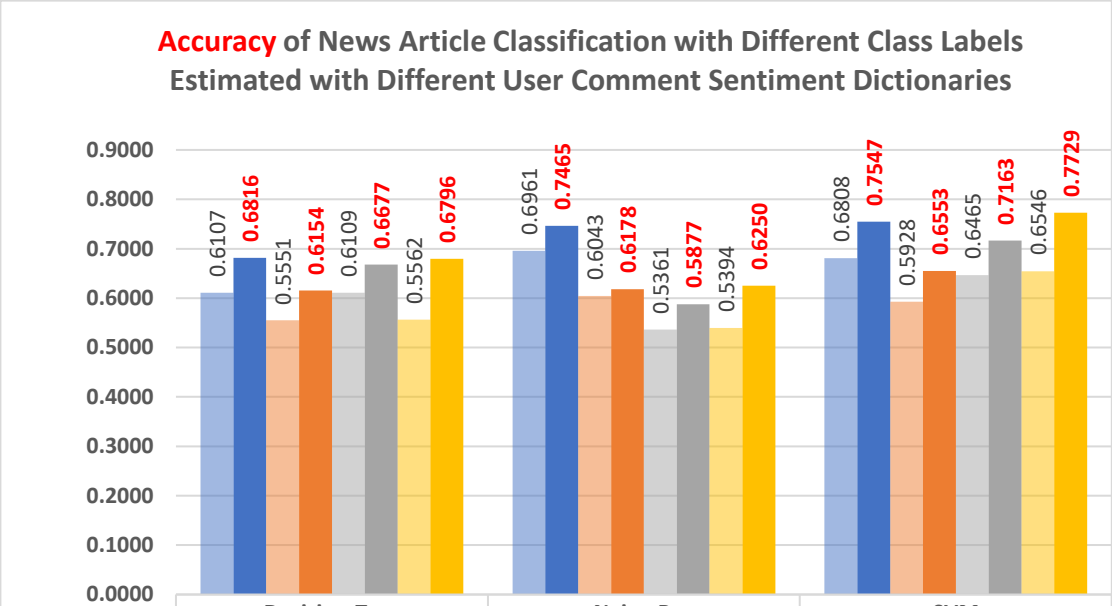
As in Figure 5.4, the result showed the recall of news article classification with class labels estimated with user comment sentiment dictionaries “NEAR” operation is higher than class labels estimated with user comment sentiment dictionaries “AND” operation in all machine learning methods and with all feature vectors.



	Decision-Tree	Naive-Bay	SVM
TF-IDF (A)	0.5970	0.6785	0.6707
TF-IDF (N)	0.6731	0.7375	0.7467
Public (A)	0.5501	0.5857	0.5849
Public (N)	0.6006	0.5895	0.6467
News 1 (A)	0.5998	0.5096	0.6362
News 1 (N)	0.6580	0.5526	0.7095
News 2 (A)	0.5455	0.5221	0.6472
News 2 (N)	0.6694	0.6036	0.7705

Figure 5.5 F-measure of User Comment Sentiment Dictionary

As in Figure 5.5, the result showed the F-measure of news article classification with class labels estimated with user comment sentiment dictionaries “NEAR” operation is higher than class labels estimated with user comment sentiment dictionaries “AND” operation in all machine learning methods and with all feature vectors.



	Decision-Tree	Naive-Bay	SVM
TF-IDF (A)	0.6107	0.6961	0.6808
TF-IDF (N)	0.6816	0.7465	0.7547
Public (A)	0.5551	0.6043	0.5928
Public (N)	0.6154	0.6178	0.6553
News 1 (A)	0.6109	0.5361	0.6465
News 1 (N)	0.6677	0.5877	0.7163
News 2 (A)	0.5562	0.5394	0.6546
News 2 (N)	0.6796	0.6250	0.7729

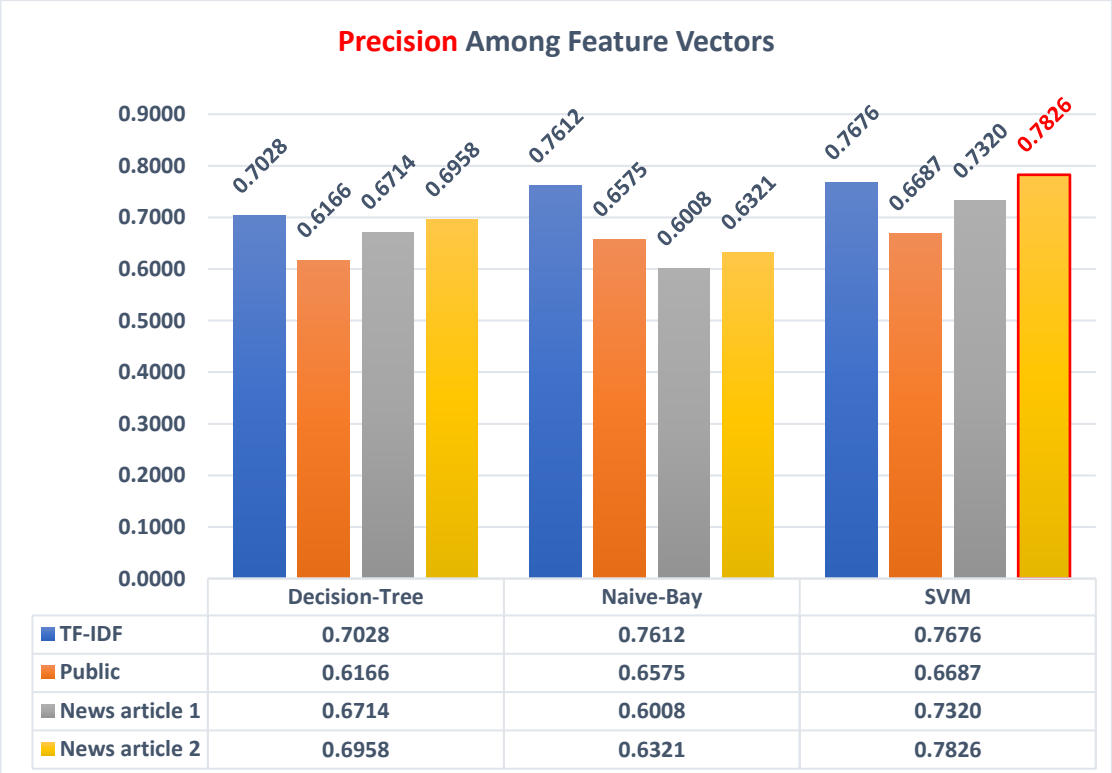
Figure 5.6 Accuracy of User Comment Sentiment Dictionary

As in Figure 5.6, the accuracy showed the F-measure of news article classification with class labels estimated with user comment sentiment dictionaries “NEAR” operation is higher than class labels estimated with user comment sentiment dictionaries “AND” operation in each machine learning and each feature vector.

From all result we obtained, we can conclude that the sentiment dictionary of user comments with operation with “NEAR” achieved the best results with all three machine learning methods. Thus, we conclude that for class label estimated with user comment sentiment dictionaries “NEAR” operation is the best method for this purpose.

5.6.2 Comparison of Feature Vector

In this subsection, the classification of 3 machine learning with 4 feature vectors is shown: TF-IDF, feature vector of public sentiment dictionary, feature vector of news article sentiment dictionary "AND" operation and feature vector of news article sentiment dictionary "NEAR" operation as follows.



	Decision-Tree	Naive-Bay	SVM
TF-IDF	0.7028	0.7612	0.7676
Public	0.6166	0.6575	0.6687
News article 1	0.6714	0.6008	0.7320
News article 2	0.6958	0.6321	0.7826

Figure 5.7 Precision of Feature Vector

As in Figure 5.7, the result showed the precision of news article sentiment dictionary "NEAR" operation feature vector with SVM is highest score at 0.7826.

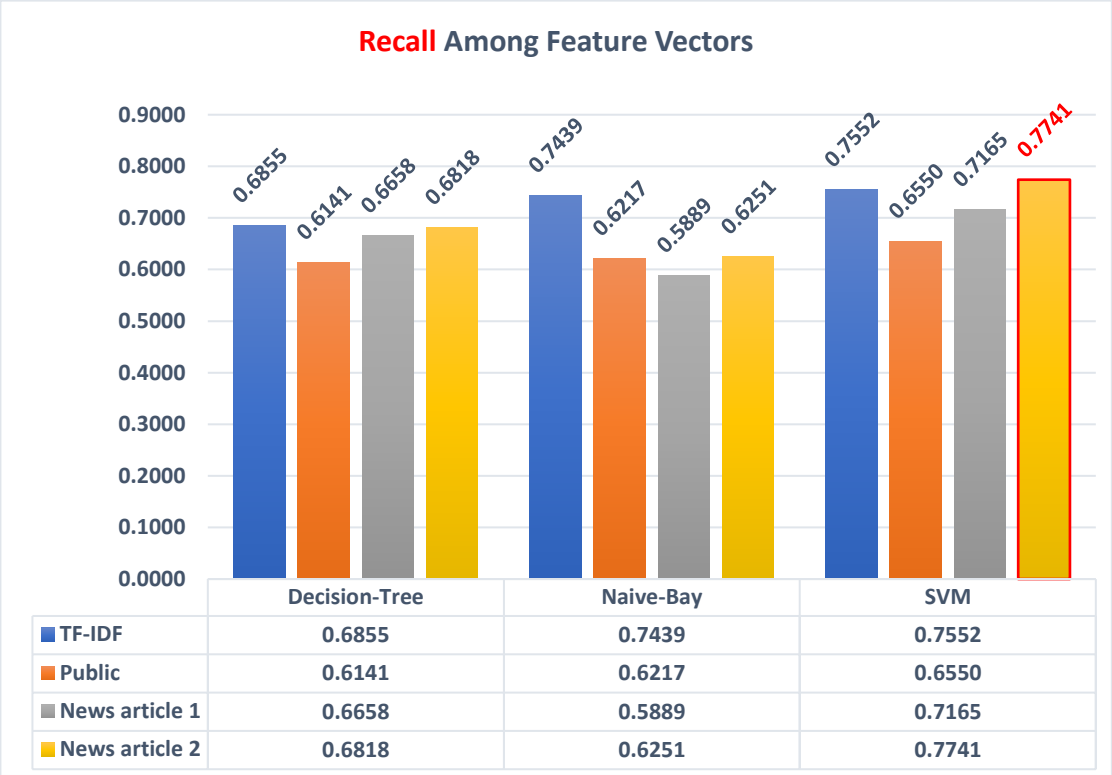


Figure 5.8 Recall of Feature Vector

As in Figure 5.8, the result showed the recall of news article sentiment dictionary "NEAR" operation feature vector with SVM is highest score at 0.7741.

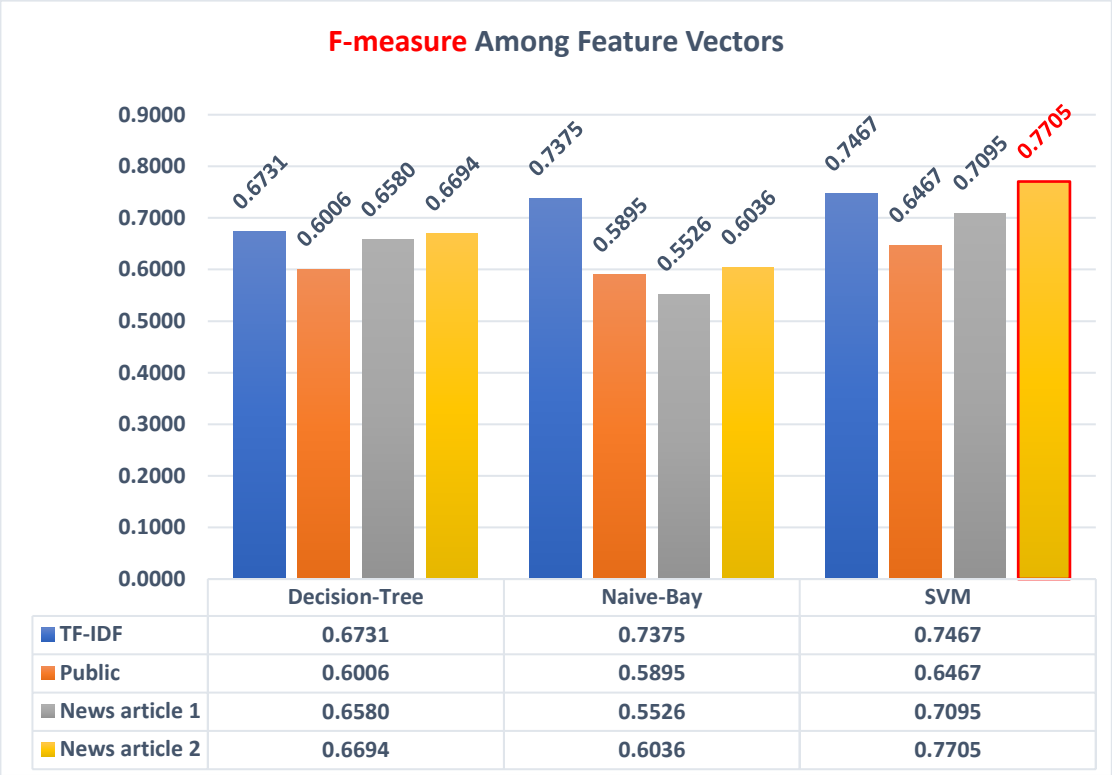


Figure 5.9 F-measure of Feature Vector

As in Figure 5.9, the result showed the F-measure of news article sentiment dictionary “NEAR” operation feature vector with SVM is highest score at 0.7705.

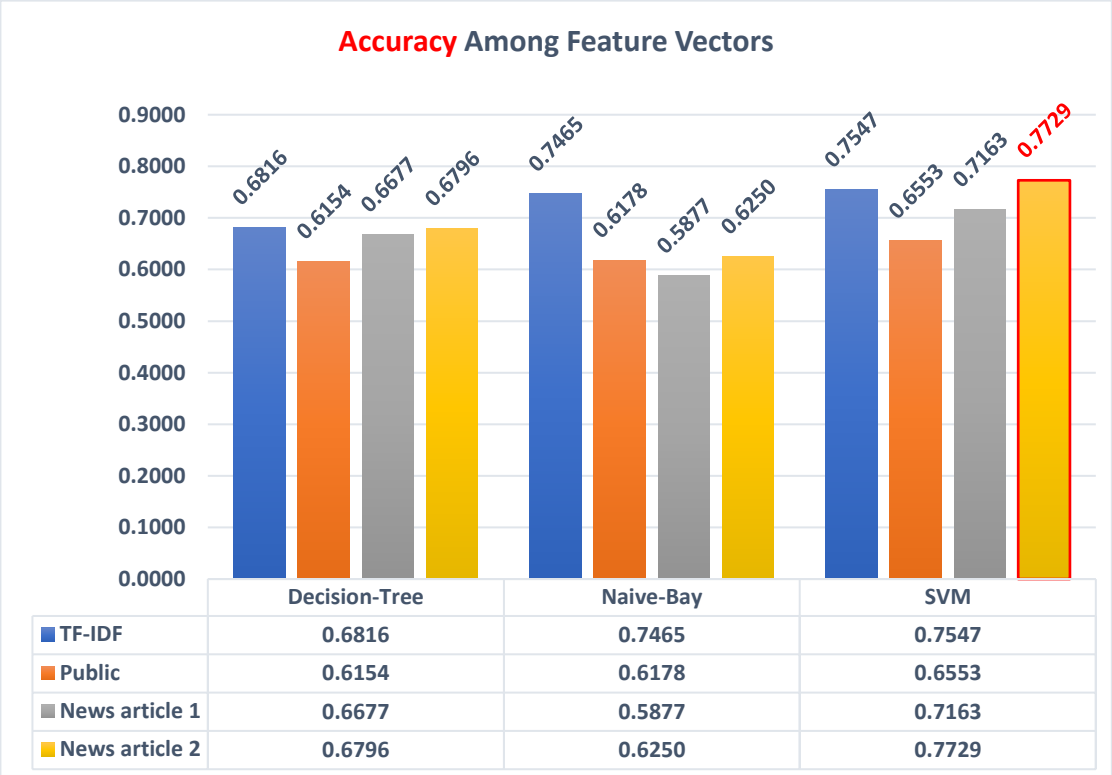


Figure 5.10 Accuracy of Feature Vector

As in Figure 5.10, the result showed the accuracy of news article sentiment dictionary "NEAR" operation feature vector with SVM is highest score at 0.7729.

From these results, we can conclude that SVM is the best machine learning model with the feature vector of news article sentiment dictionary "NEAR" operation for our purpose

Table 5.7 Comparison with Other Research in Thai Sentiment Classification Field

	Sukhum, et al. (2011)	**Haruechaiyasak, et al. (2013)**	**This research**
Purposes	Propose an opinion mining framework for detecting opinions in Thai political news columns	Proposed a framework for developing a social media analyzing tool	Thai news article classification according each user to recommend new news to user
Dataset	117 political news columns segmented to 2,539 sentence and totally around 62,000 words	Thai text from Twitter and Pantip in mobile service domain 1,768 sentiment corpus (156 Positive/ 1,612Negative)	Textual from news article based user comments 172,189 news articles 85 million words
Pre-Processing	Remove general noise, manually annotation and manually sentence segmentation, word segmentation with LexToPlus, stop word removal	Sentence segmentation, tokenization, cleaning repeated character	Noise reduction (describe in section 5.2), Word segmentation, stop word removal
Feature selection	- Unigram with TF weighting to fine the best machine learning - Prior- Knowledge - POS - Numbers	General lexicon and General lexicon + Clur lexicon	TF-IDF, Public sentiment dictionary and 2 Sentiment dictionary constructed from our data

Table 5.7 Comparison with Other Research in Thai Sentiment Classification Field (cont.)

	Sukhum, et al. (2011)	**Haruechaiyasak, et al. (2013)**	**This research**
Machine Learning Algorithm	Naïve-Bayes, K-NN and SVM	Naïve-Bayes compared between 2 features	Decision-Tree, Naïve-Bayes and SVM
Classification	Classification text in to 2 classes (Subjective and Objective)	Classification text in to 2 classes (Positive and Negative)	Balance polarity data and Classification news article in to 2 classes (Positive and Negative)
Result	80.30%	91.64%	77.29%

Table 5.7 shows the summary of comparison with other studies. When we look at accuracy, we have to say that our result is the worst, but this may not be such, considering the way of annotation, data sizes, and balance of the polarities.

First, the other studies' classification did not handle differences among users and only one polarity label is estimated for each news article. On the other hand, this research estimates different results according to each user's past comments. Thus, a simple comparison is not appropriate.

Sukhum, et al. (2011) made noise reduction and annotation mostly manually, which was possible because their dataset is small and contains 2,539 sentences. However, our data size is roughly seven times larger, and thus noise reduction used machine learning techniques and annotation of the class label was estimated from the corresponding user's comments. Considering these points, our result is considered to be high enough. The best accuracy was reported in Haruechaiyasak, et al. (2013). However, their dataset was highly imbalanced: almost 90% of them were negative and only 10% was positive. In such a case, it is well known that accuracy is not trustworthy because 100% of false positives can result in 90% of accuracy. They did not report precision, recall, and F1-measure. On the other hand, our dataset was designed balanced, whose result is trustworthy enough. It is also noted that the results shown in Table 5.7 came from the

average accuracy of 25 users, and that the best result of each user was 93.75%, which is higher than their result.

5.7 Conclusion

In this chapter, I classified the sentiment of news article with machine learning and the effective by created second experiment. The effectiveness of three classification algorithm: Decision-Tree, Naïve-Bayes and Support Vector Machine and the effectiveness of three feature vector: TF-IDF, feature vector of public sentiment dictionary, feature vector of news article sentiment dictionary “AND” operation and feature vector of news article sentiment dictionary “NEAR” operation. The result showed that Support Vector Machine and feature vector of news article sentiment dictionary constructed with operation “NEAR” achieved the best result.

CHAPTER 6

Conclusion

6.1 Summary

The purpose of this research is to propose a recommendation method of new news articles according to the taste of each user, based on their past user comments to past news articles on social media. For this purpose, the research framework consists of 5 parts: Data Collection, Text Normalization, Sentiment Dictionary Construction, Sentiment Estimation and Sentiment Classification. As the target language is Thai, there are difficulties in most parts. The most serious difficulty is the language characteristics of user comments: they contain many different types of non-standard spelling, words, and other noises, which seriously harm almost every step of the procedure. The second serious issue is lack of trustworthy sentiment dictionaries available in the Thai language, and sentiment dictionaries should be constructed with the dataset for this research. The third critical issue is lack of class labels for news article sentiment classification in terms of each user.

For these three issues, this thesis proposes the followings:

(0) A large set of user comments and corresponding news articles were collected from Facebook.

(1) A hybrid noise reduction method using heuristic rules and CRFs-based machine learning noise detection (in Chapter 4).

(2) Sentiment dictionary construction with SO-PMI-IR method (in Chapter 4).

(3) Sentiment estimation of user comments that are employed as class labels for news articles in terms of each user (in Chapter 5).

The large set of data, about 110 million words of user comments and about 85 million words of news articles enabled the reliable sentiment dictionary constructions as well as conducting experiments in terms of many users. This is also important to cope with data imbalance. The collected data revealed that the majority of user comments were negative, and only when the large dataset was available could we make data balancing between positive and negative user comments even at each user level.

All these proposals are evaluated with experiments under comparative conditions. Evaluative experiments for (1), (2), and (3) show that each proposal, as well as the whole combination of these proposals, have achieved a significant improvement of results compared to different methods and different settings. In particular, we conclude that the PMI-SO-IR method with "NEAR" operation created the best trustworthy sentiment dictionary for this research purpose, which is first confirmed with the large size of dataset, and the final news article sentiment classification was best performed with Support Vector Machines with the linear kernel, whose accuracy was 0.7729, much better than Decision Tree and Naive Bayes.

6.2 Future Tasks to Come

Although the thesis research has achieved the best result, there is much room for improvement. First, this research focused more on noise reduction and how a sentiment dictionary and sentiment estimation of user comments are conducted and employed for final news article classification, but the followings were not included:

- Improvement of word segmentation particularly more tuned to social media text
- Total consideration of negative inversion with negative adverbs
- Consideration of sentence structure for capturing more precise sentiment relations

- Application of more recent machine learning methods such as deep neural networks
- Combining text analysis with analysis of emoticons and other sentiment related non-text features
- Consideration of ironical or sarcastic expressions with which their literal meaning is regarded as positive but their implicational, true meaning as negative.

All these, together with other small refinements, are future tasks to come.

References

Ager, S. (2017). “Types of Writing System”. (Online) Available on http://omniglot.com/writing/types.htm (23 May 2017).

Ahkter, J. K., & Soria, S. (2010). “Sentiment Analysis: Facebook Status Messages.” Unpublished master's thesis, Stanford University, Stanford, CA.

Ahmad, K. (2006). “Multi-lingual sentiment analysis of financial news streams”, *Proceedings of the First International Workshop on Grid Technology for Financial Modeling and Simulation*, Palermo, Italy: 3 - 4 February, 2006.

Aroonmanakun, W. (2002). “Collocation and Thai word segmentation”, *Proceedings of the Fifth SNLP & 5th Oriental COCOSDA Workshop*, Hua Hin, Prachuapkirikhan, Thailand: 9-11 May, 2002.

Balahur, A., & Perea-Ortega, J. M. (2015). “Sentiment analysis system adaptation for multilingual processing: The case of tweets.” *Information Processing & Management*, 51(4), 547-556.

Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., & Belyaeva, J. (2013). “Sentiment analysis in the news.” *arXiv:1309.6202*.

Barragáns-Martínez, A. B., Costa-Montenegro, E., Burguillo, J. C., Rey-López, M., Mikic-Fonte, F. A., & Peleteiro, A. (2010). “A Hybrid Content-Based and Item-Based Collaborative Filtering Approach to Recommend TV Programs Enhanced with Singular Value Decomposition.” *Information Sciences*, *180*(22), 4290-4311.

Bautin, M., Vijayarenu, L., & Skiena, S. (2008, April). “International Sentiment Analysis for News and Blogs”, *Proceedings of the International Conference on Weblogs and Social Media*, Seattle, WA: 30 March - 2 April, 2008.

Bigi, B. (2011). "A Multilingual Text Normalization Approach", *Proceedings of the Fifth Language and Technology Conference*, Poznan, Poland: 25-27 November, 2011.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). "Latent Dirichlet Allocation." *Journal of Machine Learning Research*, *3*, 993-1022.

Breese, J. S., Heckerman, D., & Kadie, C. (1998). "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, Wisconsin: 24-26 July, 1998.

Breiman, L. (2017). *Classification and Regression Trees*. Routledge, New York.

Breitinger, C., Gipp, B., & Langer, S. (2015). "Research-paper Recommender Systems: a Literature Survey." *International Journal on Digital Libraries*, *17*(4), 305-338.

Burke, R. (2002). "Hybrid Recommender Systems: Survey and Experiments." *User Modeling and User-Adapted Interaction*, *12*(4), 331-370.

Cantador, I., Bellogín, A., & Vallet, D. (2010). "Content-based Recommendation in Social Tagging Systems", *Proceedings of the Fourth ACM Conference on Recommender Systems*, Barcelona, Spain: 26-30 September, 2010.

Chang, C. C., & Lin, C. J. (2011). "LIBSVM: a Library for Support Vector Machines." *ACM Transactions on Intelligent Systems and Technology*, 2(3), 1-27.

Chirawichitchai, N. (2014). "Emotion Classification of Thai Text Based Using Term Weighting and Machine Learning Techniques", *Proceedings of the Eleventh International Joint Conference on Computer Science and Software Engineering,* Pattaya, Thailand: 14-16 May, 2014.

Chirawichitchai, N., Sa-nguansat, P., & Meesad, P. (2010). "Developing an Effective Thai Document Categorization Framework Based on Term Relevance Frequency Weighting", *Proceedings of the Eighth International Conference on ICT and Knowledge Engineering*, Bangkok, Thailand: 24-25 November, 2010.

Choksuriwong, J. S. D. A. (2012). "Comparison of Support Vector Machine's Kernel Function for Unsmoke Sheet Rubber Price Forecasting", *Proceedings of the First Annual PSU Phuket International Conference 2012*, Phuket, Thailand: 10-12 2013.

Church, K. W., & Hanks, P. (1990). "Word Association Norms, Mutual Information, and Lexicography." *Computational Linguistics*, *16*(1), 22-29.

Claster, W. B., Cooper, M., & Sallis, P. (2010). "Thailand--Tourism and Conflict: Modeling Sentiment from Twitter Tweets using Naïve Bayes and Unsupervised Artificial Neural Nets", *Proceedings of the Second International Conference on Computational Intelligence, Modelling and Simulation*, Bali, Indonesia: 28-30 September, 2010.

Cortes, C., & Vapnik, V. (1995). "Support-Vector Networks." *Machine Learning*, *20*(3), 273-297.

Das, B., & Chakraborty, S. (2018). "An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation." *arXiv:1806.06407*.

Dumais, S., Platt, J., Heckerman, D., & Sahami, M. (1998). "Inductive Learning Algorithms and Representations for Text Categorization", *Proceedings of the Seventh International Conference on Information and Knowledge Management*, Washington, DC: 2-7 November, 1998.

Ellison, N. B. (2007). "Social Network Sites: Definition, History, and Scholarship." *Journal of Computer Mediated C1ommunication*, 13(1), 210-230.

Esuli, A., & Sebastiani, F. (2005). "Determining the Semantic Orientation of Terms through Gloss Classification", *Proceedings of the Fourteenth ACM International Conference on Information and Knowledge Management*, Bremen, Germany: 31 October - 5 November, 2005.

Fang, X., & Zhan, J. (2015). "Sentiment Analysis Using Product Review Data." *Journal of Big Data*, *2*-5, 14 pages.

Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, Cambridge.

Feng, Y., Huang, R., & Sun, L. (2008). "Two Step Chinese Named Entity Recognition Based on Conditional Random Fields Models", *Proceedings of the Sixth SIGHAN Workshop on Chinese Language Processing*, Hyderabad, India: 11-12 January, 2008.

Gamon, M., & Aue, A. (2005). "Automatic Identification of Sentiment Vocabulary: Exploiting Low Association with Known Sentiment Terms", *Proceedings of the ACL Workshop on Feature Engineering for Machine Learning in Natural Language Processing*, Ann Arbor, MI: 29 June 2005.

García-Crespo, Á., López-Cuadrado, J. L., Colomo-Palacios, R., González-Carrasco, I., & Ruiz-Mezcua, B. (2011). "Sem-Fit: A Semantic Based Expert System to Provide Recommendations in the Tourism Domain." *Expert Systems with Applications*, *38*(10), 13310-13319.

Gebre, B. G., Zampieri, M., Wittenburg, P., & Heskes, T. (2013). "Improving Native Language Identification with Tf-Idf Weighting", *Proceedings of the Eighth NAACL Workshop on Innovative Use of NLP for Building Educational Applications (BEA8)*, Atlanta, GA: 9-14 June 2013.

Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). "Large-Scale Sentiment Analysis for News and Blogs." *ICWSM*, *7*(21), 219-222.

Guzman, E., & Maalej, W. (2014). "How Do Users Like this Feature?: A Fine Grained Sentiment Analysis of App Reviews", *Proceedings of the 22th International Conference on Requirements Engineering Conference (RE)*, Karlskrona, Sweden: 25-29 August, 2014.

Haruechaiyasak, C., & Kongthon, A. (2013). "LexToPlus: A Thai Lexeme Tokenization and Normalization Tool", *Proceedings of the Fourth Workshop on South and Southeast Asian Natural Language Processing*, Nagoya, Japan: 14-18 October, 2013.

Haruechaiyasak, C., Kongthon, A., Palingoon, P., & Trakultaweekoon, K. (2013). "S-Sense: A Sentiment Analysis Framework for Social Media Sensing", *Proceedings of the IJCNLP 2013 Workshop on Natural Language Processing for Social Media (SocialNLP)*, Nagoya, Japan: 1 June, 2013.

He, W., Zha, S., & Li, L. (2013). "Social Media Competitive Analysis and Text Mining: A Case Study in the Pizza Industry." *International Journal of Information Management*, 33(3), 464-472.

Hemalatha, I., Varma, G. S., & Govardhan, A. (2012). "Preprocessing the Informal Text for Efficient Sentiment Analysis." *International Journal of Emerging Trends & Technology in Computer Science, 1*(2), 58-61.

Hirankan, P., Suchato, A., & Punyabukkana, P. (2013). "Detection of Wordplay Generated by Reproduction of Letters in Social Media Texts", *Proceedings of the Tenth International Joint Conference on Computer Science and Software Engineering,* Bangkok, Thailand: 12-14 May, 2010.

Inrak, P., & Sinthupinyo, S. (2010). "Applying Latent Semantic Analysis to Classify Emotions in Thai Text", *Proceedings of the Second International Conference on Computer Engineering and Technology*, Chengdu, China: 16-18 April, 2010.

Jain, T. I., & Nemade, D. (2010). "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis." *International Journal of Computer Applications IJCA*, *7*(5), 5-11.

Jiamthapthaksin, R., Setthawong, P., & Ratanasawetwad, N. (2016). "A System for Popular Thai Slang Extraction from Social Media Content with n-ram Based Tokenization", *Proceedings of the Eighth International Conference on Knowledge and Smart Technology*, Chiang Mai, Thailand: 3-6 February, 2016.

Joachims, T. (1998). "Text Categorization with Support Vector Machines: Learning with Many Relevant Features", *Proceedings of European Conference on Machine Learning*, Chemnitz, Germany: 21-24 April, 1998.

Joachims, T. (1999). “Svmlight: Support vector machine.” *SVM-Light Support Vector Machine http://svmlight. joachims. org/, University of Dortmund*, *19*(4).

Joshi, A., Balamurali, A. R., & Bhattacharyya, P. (2010). A fall-back strategy for sentiment analysis in hindi: a case study. *Proceedings of the Eighth International Conference on Natural Language Processing*, Kharagpur, India: 8-11 December, 2010.

Khan, A. Z., Atique, M., & Thakare, V. M. (2015). “Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis.” *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE)*, 89, 7 pages.

Klinger, R., & Tomanek, K. (2007). *Classical probabilistic models and conditional random fields*. TU, Algorithm Engineering.

Koanantakool, H. T., Karoonboonyanan, T., & Wutiwiwatchai, C. (2009). “Computers and the Thai Language.” *IEEE Annals of the History of Computing*, *31*(1), 46-61.

Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). “Twitter Sentiment Analysis: The Good the Bad and the Omg!.” *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media,* Barcelona, Spain: 17-21 July, 2011.

Kunpattanasopon, N., Tongtep, N., & Hashimoto, K. (2017). “Noise Reduction Effect on Thai Social Texts Sentiment Analysis”, *Proceedings of Artificial Intelligence and Natural Language Processing (iSAI-NLP 2017)*, Hua Hin, Thailand: 27-29 August, 2017.

LEXiTRON. (2016) “LEXiTRON Thai – English Electronic Dictionary” (online) Available on https://www.nectec.or.th/innovation/innovation-software/lexitron.html (23 May 2017).

Lafferty, J., McCallum, A., & Pereira, F. C. (2001). “Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data”, *Proceedings of the 18th International Conference on Machine Learning 2001*, San Francisco, CA: 28 June - 1 July, 2001.

Li, G., & Liu, F. (2010). "A Clustering-based Approach on Sentiment Analysis", *Proceedings of 2010 International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, Hangzhou, China: 15-16 November, 2010.

Linden, G., Smith, B., & York, J. (2003). "Amazon.com Recommendations: Item-to-Item Collaborative Filtering." *IEEE Internet Computing*, *7*(1), 76-80.

Maynard, D., Roberts, I., Greenwood, M. A., Rout, D., & Bontcheva, K. (2017). "A Framework for Real-Time Semantic Social Media Analysis." *Web Semantics: Science, Services and Agents on the World Wide Web*, *44*, 75-88.

Moknarong, N., Suchato, A., & Punyabukkana, P. (2013). "Detecting Romanized Thai Tokens in Social Media Texts", *Proceedings of 2013 International Conference on Computer Science and Engineering Conference (ICSEC)*, Bangkok, Thailand: 4-6 September, 2013.

Mostafa, M. M. (2013). "More than Words: Social Networks' Text Mining for Consumer Brand Sentiments." *Expert Systems with Applications,* 40(10), 4241-4251.

Mr. Bayes, & Price, M. (1763). "An Essay towards Solving a Problem in the Doctrine of Chances." *Philosophical Transactions (1683-1775)*, 370-418.

Mulkalwar, A., & Kelkar, K. (2012). "Sentiment Analysis on Movie Reviews Based on Combined Approach." *International Journal of Science and Research 3(7)*.

Mullen, T., & Collier, N. (2004). "Sentiment Analysis using Support Vector Machines with Diverse Information Sources", *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain: 25-26 July, 2004.

Nakov, P. (2017). "Semantic Sentiment Analysis of Twitter Data." *arXiv:1710.01492*.

National Electronics and Computer Technology Center: NECTEC. (2016) "LexToPlus: A Thai Lexeme Tokenization and Normalization Too." (online) Available on https://www.nectec.or.th/innovation/innovation-software/lextoplus.html (23 May 2017).

Netisopakul, P., & Chattupan, A. (2015). "Thai Stock News Sentiment Classification Using Wordpair Features", *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, Shanghai, China: 30 October – 1 November, 2015.

Ngoc, P. T., & Yoo, M. (2014). "The Lexicon-based Sentiment Analysis for Fan Page Ranking in Facebook", *Proceedings of 2014 International Conference on Information Networking*, Phuket, Thailand: 10-12 February, 2014.

Nielsen, F. Å. (2011). "A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs." *arXiv:1103.2903*.

Nomponkrang, T., & Sanrach, C. (2016). "The Comparison of Algorithms for Thai-Sentence Classification." *International Journal of Information and Education Technology*, 6(10) pp.801-808.

Number of daily active Facebook users worldwide as of 1st quarter (2018). "The Statistics Portal", (online) Available on https://www.statista.com/statistics/346167/facebook-global-dau/ (10 May 2018).

Okazaki, N. (2007). "CRFsuite: a fast implementation of Conditional Random Fields." (online) Available on http://www. chokkan. org/software/crfsuite (24 March 2015).

Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). "Sentiment Analysis in Facebook and its Application to E-learning." *Computers in Human Behavior*, *31*, 527-541.

Paltoglou, G., & Thelwall, M. (2010). A Study of Information Retrieval Weighting Schemes for Sentiment Analysis", *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics,* Uppsala, Sweden: 11-16 July, 2010.

Pak, A., & Paroubek, P. (2010, May). "Twitter as a corpus for sentiment analysis and opinion mining", *LREC 2010 Proceedings*, Malta, Spain: 17-23 May, 2010.

Pazzani, M. J. (1999). "A Framework for Collaborative, Content-based and Demographic Filtering." *Artificial Intelligence Review*, *13*(5-6), 393-408.

Pazzani, M. J., & Billsus, D. (2007). “Content-based Recommendation Systems”, In *The Adaptive Web: Methods and Strategies of Web Personalization*, Springer, Berlin, pp. 325-341.

PyThaiNLP. “Public Source Lexicon Thai” (online) Available on https://github.com/PyThaiNLP/lexicon-thai (24 May 2018).

Quinlan, J. R. (1986). “Induction of Decision Trees.” *Machine learning, 1*(1), 81-106.

Quinlan, J. R. (2014). *C4.5: Programs for Machine Learning*. Elsevier, Amsterdam, North Hollands.

Ray, P., & Chakrabarti, A. (2017). “Twitter Sentiment Analysis for Product Review using Lexicon Method”, *Proceedings of 2017 International Conference on Data Management, Analytics and Innovation,* Pune, India: 24-26 February, 2017.

Rosenthal, S., Farra, N., & Nakov, P. (2017). “SemEval-2017 Task 4: Sentiment Analysis in Twitter”, *Proceedings of the 11th International Workshop on Semantic Evaluation,* Vancouver, Canada: 3-4 August, 2017.

Sarakit, P., Theeramunkong, T., Haruechaiyasak, C., & Okumura, M. (2015). “Classifying Emotion in Thai Youtube Comments”, *Proceedings of 2015 Sixth International Converence on Information and Communication Technology for Embedded Systems (IC-ICTES)*, Hua Hin, Thailand: 22-24 March, 2015.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). “Item-based Collaborative Filtering Recommendation Algorithms”, *Proceedings of the Tenth International Conference on World Wide Web*, Hong Kong: 1-5 May, 2001.

Servia-Rodriguez, S., Fernandez-Vilas, A., Diaz-Redondo, R. P., & Pazos-Arias, J. J. (2013). “Inferring Contexts from Facebook Interactions: A Social Publicity Scenario.” *IEEE Transactions on Multimedia,* 15(6), 1296-1303.

Shi, H. X., & Li, X. J. (2011). “A Sentiment Analysis Model for Hotel Reviews Based on Supervised Learning”, *Proceedings of 2011 International Conference on Machine Learning and Cybernetics,* Guilin, China: 10-13 July, 2011.

Sparck Jones, K. (1972). “A Statistical Interpretation of Term Specificity and its Application in Retrieval.” *Journal of Documentation*, *28*(1), 11-21.

Spärck Jones, K. (2004). “IDF Term Weighting and IR Research Lessons.” *Journal of Documentation*, *60*(5), 521-523.

Sornlertlamvanich, V., Potipiti, T., & Charoenporn, T. (2000). “Automatic Corpus-based Thai Word Extraction with the C4.5 Learning Algorithm”, *Proceedings of the 18th Conference on Computational Linguistics*, Stroudsburg, PA: 31 July - 4 August, 2000.

Steinberger, J., Lenkova, P., Ebrahim, M., Ehrmann, M., Hurriyetoglu, A., Kabadjov, M., ... & Vázquez, S. (2011). “Creating Sentiment Dictionaries via Triangulation”, *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, Portland, Oregon: 24 June, 2011.

Sukhum, K., Nitsuwat, S., & Haruechaiyasak, C. (2011). “Opinion Detection in Thai Political News Columns Based on Subjectivity Analysis.” *Information Technology Journal*, *7*(2), 32-37.

Supnithi, T., Onman, C., Porkaew, P., Ruangrajitpakorn, T., Trakultaweekoon, K., & Kawtrakul, A. (2010). “A Supervised Learning based Chunking in Thai using Categorial Grammar”, *Proceedings of the Eighth Workshop on Asian Language Resources*, Beijing, China: 23-27 August, 2010.

Sutton, C., & McCallum, A. (2012). “An Introduction to Conditional Random Fields.” *Foundations and Trends in Machine Learning*, *4*(4), 267-373.

Taboada, M., Anthony, C., & Voll, K. (2006). “Methods for Creating Semantic Orientation Dictionaries”, *Proceedings of the Fifth International Conference on Language Resources and Evaluation,* Genoa, Italy: 22-28 May 2006.

Takeuchi, K., & Collier, N. (2005). “Bio-Medical Entity Extraction using Support Vector Machines.” *Artificial Intelligence in Medicine*, *33*(2), 125-137.

Thanangthanakij, S., Pacharawongsakda, E., Tongtep, N., Aimmanee, P., & Theeramunkong, T. (2012). “An Empirical Study on Multi-Dimensional Sentiment Analysis from User Service Reviews”, *Proceedings of the Seventh International Conference on Knowledge, Information and Creativity Support Systems,* Melbourne, Australia: 8-10 November, 2012.

Theeramunkong, T., & Usanavasin, S. (2001). “Non-Dictionary-based Thai Word Segmentation using Decision Trees”, *Proceedings of the First International Conference on Human Language Technology Research*, San Diego, CA: 18-21 March, 2001.

Theeramunkong, T., Sornlertlamvanich, V., Tanhermhong, T., & Chinnan, W. (2000). “Character Cluster Based Thai Information Retrieval”, *Proceedings of the Fifth International Workshop on Information Retrieval with Asian Languages*, Chiang Mai, Thailand: 13 November, 2011.

Tongtep, N. (2013). *Extracting Entities and Predicate-oriented Relations in Thai (syllabic Alphabetic Language)*. Doctoral Dissertation, Sirindhorn International Institute of Technology, Thammasat University.

Tongtep, N., & Theeramunkong, T. (2011). “Multi-stage Annotation using Pattern-based and Statistical-based Techniques for Automatic Thai Annotated Corpus Construction”, *Proceedings of the 9th Workshop on Asian Language Resources*, Chiangmai, Thailand: 12-13 November, 2011.

Tongtep, N., & Theeramunkong, T. (2011). “Simultaneous Character-Cluster-based Word Segmentation and Named Entity Recognition in Thai Language”, *Proceedings of the Sixth International Conference on Knowledge, Information and Creativity Support Systems*, Beijing, China: 22-24 October, 2011.

Tongtep, N., & Theeramunkong, T. (2012). “Discovery of Predicate-Oriented Relations among Named Entities Extracted from Thai Texts.” *IEICE Transactions on Information and Systems,* 95(7), 1932-1946.

Troussas, C., Virvou, M., Espinosa, K. J., Llaguno, K., & Caro, J. (2013, July). "Sentiment Analysis of Facebook Statuses using Naive Bayes Classifier for Language Learning," *Proceedings of the Fourth International Conference on Information, Intelligence, Systems and Applications*, Piraeus, Greece: 10-12 July, 2013.

Tsihrintzis, G. A. (2013). "Analysis of Opinions from Questionnaire Surveys of Farming Candidates using Cross Tabulation System", *Proceedings of the Sixth International Conference on Intelligent Interactive Multimedia Systems and Services*, Sesimbra, Portugal: 26-28 June, 2013.

Tuckett, D., Smith, R. E., & Nyman, R. (2014). "Tracking Phantastic Objects: A Computer Algorithmic Investigation of Narrative Evolution in Unstructured Data Sources." *Social Networks*, *38*, 121-133.

Turney, P. D. (2001). "Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL", *Proceedings of European Conference on Machine Learning*, Freiburg, Germany: 5-7 September, 2001.

Turney, P. D. (2002). "Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews", *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Philadelphia, Pennsylvania: 7-12 July, 2002.

Turney, P. D., & Littman, M. L. (2002). "Unsupervised Learning of Semantic Orientation from a Hundred-Billion-Word Corpus." *arXiv preprint cs/0212012*.

Van Zaanen, M., & Kanters, P. (2010). "Automatic Mood Classification Using TF*IDF Based on Lyrics." *Proceedings of the Eleventh International Society for Music Information Retrieval Conference*, Utrecht, Netherlands: 9-13 August, 2010.

Vateekul, P., & Koomsubha, T. (2016). "A Study of Sentiment Analysis using Deep Learning Techniques on Thai Twitter Data", *Proceedings of the 13th International Joint*

Conference on Computer Science and Software Engineering, Khon Kaen, Thailand: 13-15 July, 2016.

Wang, G., Sun, J., Ma, J., Xu, K., & Gu, J. (2014). "Sentiment Classification: The Contribution of Ensemble Learning." *Decision Support Systems*, *57*, 77-93.

Wang, W., Chen, L., Thirunarayan, K., & Sheth, A. P. (2012). "Harnessing Twitter "Big data" for Automatic Emotion Identification", *Proceedings of 2012 ASE/IEEE International Conference on Social Computing*, Amsterdam, Netherlands: 3-5 September, 2012.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis", *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Vancouver, British Columbia, Canada: 6-8 October, 2005.

Xia, T., & Chai, Y. (2011). "An Improvement to TF-IDF: Term Distribution based Term Weight Algorithm." *Journal of Software*, *6*(3), 413-420.

Yassine, M., & Hajj, H. (2010). "A Framework for Emotion Mining from Text in Online Social Networks", *Proceedings of 2010 IEEE International Conference on Data Mining (ICDM),* Sydney, Australia: 13-17 December, 2010.

Yu, D., Mu, Y., & Jin, Y. (2017). "Rating Prediction Using Review Texts with Underlying Sentiments." *Information Processing Letters*, *117*, 10-18.

Yuen, R. W., Chan, T. Y., Lai, T. B., Kwong, O. Y., & T'sou, B. K. (2004). "Morpheme-based Derivation of Bipolar Semantic Orientation of Chinese Words", *Proceedings of the 20th international conference on Computational Linguistics*, Geneva, Switzerland: 23-27 August, 2004.

Zarrinkalam, F., & Kahani, M. (2012). "A Multi-Criteria Hybrid Citation Recommendation System Based on Linked Data", *Proceedings* of 2012 *Second International eConference on Computer and Knowledge Engineering (ICCKE),* Ferdowsi University of Mashhad, Iran: 18-19 October, 2012.

Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., & Liu, B. (2011). *Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis*. HP Laboratories, Technical Report HPL-2011, 89.

Appendices

Appendix A

HTML tags, HTML character entities, Thai punctuations and Thai numbers, Math symbols and Emotions and special Characters

HTML Tags

<a href>		<article>	</article>			
<cite>	</cite>	<dd>	</dd>			<dl>
</dl>	<dt>	</dt>			<figure>	</figure>
<h1>	</h1>	<h2>	</h2>	<h3>	</h3>	<h4>
</h4>	<h5>	</h5>	<h6>	</h6>	<hr>	
	<img src>	<ins>	</ins>			<nav>
</nav>					<pre>	</pre>
<q>	</q>			<style>	</style>	<sup>
</sup>	<table>	</table>	<td>	</td>	<th>	</th>
<time>	</time>	<tr>	</tr>	<tt>	</tt>	<u>
</u>			<var>	</var>	<video>	</video>

HTML Character Entities

No.	HTML	Symbol	No.	HTML	Symbol	No.	HTML	Symbol
1	"	"	28	{	{	55	·	·
2	#	#	29			|	56	¸	¸
3	$	$	30	}	}	57	¹	¹
4	%	%	31			58	º	º
5	&	&	32	¡	¡	59	»	»
6	'	'	33	¢	¢	60	¼	¼
7	(	(	34	£	£	61	½	½
8	)	)	35	¤	¤	62	¾	¾
9	*	*	36	¥	¥	63	¿	¿
10	+	+	37	₹	₹	64	À	À
11	,	,	38	¦	¦	65	Á	Á
12	−	-	39	§	§	66	Â	Â
13	.	.	40	¨	¨	67	Ã	Ã
14	/	/	41	©	©	68	Ä	Ä
15	:	:	42	ª	ª	69	Å	Å
16	;	;	43	«	«	70	Æ	Æ
17	<	<	44	¬	¬	71	Ç	Ç
18	=	=	45			72	È	È
19	>	>	46	®	®	73	É	É
20	?	?	47	¯	¯	74	Ê	Ê
21	@	@	48	°	°	75	Ë	Ë
22	[	[	49	±	±	76	Ì	Ì
23	\	\	50	²	²	77	Í	Í
24	]	]	51	³	³	78	Î	Î
25	^	^	52	´	´	79	Ï	Ï
26	_	_	53	µ	µ	80	Ð	Ð
27	`	`	54	¶	¶	81	Ñ	Ñ

No.	HTML	Symbol	No.	HTML	Symbol	No.	HTML	Symbol
82	Ò	Ò	110	î	î	138	Γ	Γ
83	Ó	Ó	111	ï	ï	139	Δ	Δ
84	Ô	Ô	112	ð	ð	140	Ε	Ε
85	Õ	Õ	113	ñ	ñ	141	Ζ	Ζ
86	Ö	Ö	114	ò	ò	142	Η	Η
87	×	×	115	ó	ó	143	Θ	Θ
88	Ø	Ø	116	ô	ô	144	Ι	Ι
89	Ù	Ù	117	õ	õ	145	Κ	Κ
90	Ú	Ú	118	ö	ö	146	Λ	Λ
91	Û	Û	119	÷	÷	147	Μ	Μ
92	Ü	Ü	120	ø	ø	148	Ν	Ν
93	Ý	Ý	121	ù	ù	149	Ξ	Ξ
94	Þ	Þ	122	ú	ú	150	Ο	Ο
95	ß	ß	123	û	û	151	Π	Π
96	à	à	124	ü	ü	152	Ρ	Ρ
97	á	á	125	ý	ý	153	Σ	Σ
98	â	â	126	þ	þ	154	Τ	Τ
99	ã	ã	127	ÿ	ÿ	155	Υ	Υ
100	ä	ä	128	Œ	Œ	156	Φ	Φ
101	å	å	129	œ	œ	157	Χ	Χ
102	æ	æ	130	Š	Š	158	Ψ	Ψ
103	ç	ç	131	š	š	159	Ω	Ω
104	è	è	132	Ÿ	Ÿ	160	α	α
105	é	é	133	ƒ	*f*	161	β	β
106	ê	ê	134	ˆ	ˆ	162	γ	γ
107	ë	ë	135	˜	˜	163	δ	δ
108	ì	ì	136	Α	Α	164	ε	ε
109	í	í	137	Β	Β	165	ζ	ζ

No.	HTML	Symbol	No.	HTML	Symbol	No.	HTML	Symbol
166	η	η	194	‏		222	↓	↓
167	θ	θ	195	–	–	223	↔	↔
168	ι	ι	196	—	—	224	↵	↵
169	κ	κ	197	‘	‘	225	⇐	⇐
170	λ	λ	198	’	’	226	⇑	⇑
171	μ	μ	199	‚	,	227	⇒	⇒
172	ν	ν	200	“	“	228	⇓	⇓
173	ξ	ξ	201	”	”	229	⇔	⇔
174	ο	ο	202	„	„	230	∀	∀
175	π	π	203	†	†	231	∂	∂
176	ρ	ρ	204	‡	‡	232	∃	∃
177	ς	ς	205	‰	‰	233	∅	∅
178	σ	σ	206	‹	‹	234	∇	∇
179	τ	τ	207	›	›	235	∈	∈
180	υ	υ	208	•	•	236	∉	∉
181	φ	φ	209	…	…	237	∋	∋
182	χ	χ	210	′	′	238	∏	∏
183	ψ	ψ	211	″	″	239	∑	∑
184	ω	ω	212	‾	‾	240	−	−
185	ϑ	ϑ	213	⁄	⁄	241	∗	∗
186	ϒ	ϒ	214	℘	℘	242	√	√
187	ϖ	ϖ	215	ℑ	ℑ	243	∝	∝
188			216	ℜ	ℜ	244	∞	∞
189			217	™	™	245	∠	∠
190			218	ℵ	ℵ	246	∧	⊥
191	‌		219	←	←	247	∨	⊦
192	‍		220	↑	↑	248	∩	∩
193	‎		221	→	→	249	∪	∪

No.	HTML	Symbol
250	∫	∫
251	∴	∴
252	∼	∼
253	≅	≅
254	≈	≈
255	≠	≠
256	≡	≡
257	≤	≤
258	≥	≥
259	⊂	⊂
260	⊃	⊃
261	⊄	⊄
262	⊆	⊆
263	⊇	⊇
264	⊕	⊕
265	⊗	⊗
266	⊥	⊥
267	⋅	⋅
268	⌈	⌈
269	⌉	⌉
270	⌊	⌊
271	⌋	⌋
272	⟨	⟨
273	⟩	⟩
275	♠	♠
276	♣	♣
277	♥	♥
278	♦	♦

Thai Punctuations and Thai Numbers

ๆ	ฯ	ฯลฯ	๏	๛
ฺ	(	)	?	.
,	‘	’	“	”
๐	๑	๒	๓	๔
๕	๖	๗	๘	๙

Math Symbols

+	−	×	*	÷
/	=	≠	≈	%
(	)	[	]	^
{	}	<	>	≥
≥	!	∑	\	$

Emoticons and Special Characters

:)	:-)	:(	:-(	:|
:-|	(^^)	(x_x)	(T_T)	=0=
^_^'	--a	(0‿0)	(^ム^)	(益)
(ˊДˋ)	=A=	^_____^	-3-	-w-
o_Ô	ò_ó	ó_ò	ù_u	q[- -]p

Appendix B

Frequently misspelled words

Misspelled Words List

No.	Incorrect	Correct
1	ก๊	ก็
2	กงกำกง เกวียน	กงเกวียนกำ เกวียน
3	กงศุล	กงสุล
4	กฎ	กฎ
5	กฎเกณฑ์	กฎเกณฑ์
6	กฐิณ	กฐิน
7	ก้นป่อง	ก้นปล่อง
8	กบฎ	กบฏ
9	กบด	กบฏ
10	กบถ	กบฏ
11	กบาร	กระบาล
12	กบาล	กระบาล
13	กรรณแสง	กันแสง
14	กรรมกรณ์	กรรมกร
15	กรรแสง	กันแสง
16	กระเฌอ	กระเชอ
17	กระตือรือล้น	กระตือรือร้น
18	กระทัดรัด	กะทัดรัด
19	กระทันหัน	กะทันหัน
20	กระทิ	กะทิ
21	กระทืบ	กระทืบ
22	กระเทย	กะเทย
23	กระเทาะ	กะเทาะ
24	กระบังลม	กะบังลม
25	กระปิ	กะปิ
26	กระพริบ	กะพริบ
27	กระพรุน	กะพรุน
28	กระเพรา	กะเพรา
29	กระเพราะ	กระเพาะ
30	กระเพา	กะเพรา
31	กระล่อน	กะล่อน
32	กระหรี่	กะหรี่
33	กระหล่ำ	กะหล่ำ
34	กระเหรี่ยง	กะเหรี่ยง
35	กระโหลก	กะโหลก
36	กริยา	กิริยา
37	กุ	กู
38	กรุ	กู
39	กรู	กู
40	กลยุทธ	กลยุทธ์
41	กลยุทธิ์	กลยุทธ์
42	กลยุทธิ	กลยุทธ์
43	กล้วยบวดชี	กล้วยบวชชี
44	กลอฟ	กอล์ฟ
45	กลางครัน	กลางคัน
46	กลาสี	กะลาสี
47	กลิ่นไอ	กลิ่นอาย
48	กเลวราก	กเพวราก
49	กษัตรย์	กษัตริย์
50	กสิน	กสิณ
51	ก้อ	ก็
52	กอป	กอปร
53	ก็อปปี้	ก๊อบปี้
54	กอลฟ์	กอล์ฟ
55	ก็อล์ฟ	กอล์ฟ
56	ก็อลฟ์	กอล์ฟ
57	ก๊อล์ฟ	กอล์ฟ
58	ก๊อลฟ์	กอล์ฟ
59	กะจิดริด	กะจิริด
60	กะเชอ	กระเชอ
61	กะเฌอ	กระเชอ
62	กะทะ	กระทะ
63	กะเท่เร่	กระเท่เร่
64	กะบด	กบฏ
65	กะบาน	กระบาล
66	กะบาร	กระบาล
67	กะบาล	กระบาล
68	กะเพราะ	กระเพาะ
69	กะเพา	กะเพรา
70	กะเพาะ	กระเพาะ
71	กักขระ	กักขพะ
72	กังวน	กังวล
73	กังวาล	กังวาน
74	กัณลักษณ์	กันทรลักษ์
75	กันทรลักษณ์	กันทรลักษ์
76	กัมมะลอ	กำมะลอ
77	กัมมะหยี	กำมะหยี่
78	กากบาด	กากบาท
79	กางครัน	กลางคัน
80	กางคัน	กลางคัน
81	ก๊าฟ	ครับ
82	การุญ	การุณย์
83	การุณ	การุณย์
84	กาลเทศะ	กาลเทศะ
85	กาลสินธ์	กาฬสินธุ์
86	กาละเทศะ	กาลเทศะ
87	กาละมัง	กะละมัง
88	กาละแม	กะละแม
89	กาละแมร์	กะละแม
90	กาลามัง	กะละมัง
91	กาลาแม	กะละแม

No.	Incorrect	Correct
92	กาลาสี	กะลาสี
93	กาแล็กซี่	กาแล็กซี
94	กาฬเวลา	กาลเวลา
95	กาฬสินธ์	กาฬสินธุ์
96	กำเน็จ	กำเหน็จ
97	กำปนาท	กัมปนาท
98	กำหราบ	กำราบ
99	กำเหน็ด	กำเหน็จ
100	กิจจลักษณะ	กิจจะลักษณะ
101	กิจลักษณะ	กิจจะลักษณะ
102	กิติคุณ	กิตติคุณ
103	กิติมศักดิ์	กิตติมศักดิ์
104	กินนรี	กินรี
105	กินนารี	กินรี
106	กิริยา	กริยา
107	กี๋ย	ไก่
108	กุ	กู
109	กุฎ	กุฏิ
110	กุฎิ	กุฏิ
111	กุฎี	กุฏิ
112	กุยไช่	กุยช่าย
113	กุว์	กู
114	เกนทหาร	เกณฑ์ทหาร
115	เกมส์	เกม
116	เกร็ดปลา	เกล็ดปลา
117	เกร็ดเลือด	เกล็ดเลือด
118	เกรียจ	เกลียด
119	เกล็ดความรู้	เกร็ดความรู้
120	เกลือสินเธา	เกลือสินเธาว์
121	เกษมสันติ์	เกษมสันต์
122	เกษมสานต์	เกษมศานต์
123	เกษร	เกสร
124	เกษียณสมุทร	เกษียรสมุทร
125	เกษียน	เกษียณ
126	เกษียร	เกษียณ
127	เก๊าท์	เกาต์
128	เกียตร	เกียรติ
129	เกียรต์	เกียรติ
130	เกียรต	เกียรติ
131	เกียรติ์	เกียรติ
132	เกียรติมศักดิ์	กิตติมศักดิ์
133	แก็ง	แก๊ง
134	แก๊งก์	แก๊ง
135	แก๊งค์	แก๊ง
136	แกงบวช	แกงบวด
137	แก๊ซ	แก๊ส
138	แกรน	แกร็น
139	ไกล้	ใกล้
140	ขนขวาย	ขวนขวาย
141	ขบฎ	ขบถ
142	ขบด	ขบถ
143	ขมักขเม้น	ขะมักเขม้น
144	ขมักเขม้น	ขะมักเขม้น
145	ขมุกขมอม	ขะมุกขะมอม
146	ขยั้นขยอ	คะยั้นคะยอ
147	ขเยก	เขยก
148	ขโยกเขยก	โขยกเขยก
149	ขโยง	โขยง
150	ขร้า	คะ
151	ขร๊า	คะ
152	ขวาญ	ขวาน
153	ขออำไพ	ขออภัย
154	ขะมีขมัน	ขมีขมัน
155	ขะมุกขมัว	ขมุกขมัว
156	ขะเม็ดแขม่	เขม็ดแขม่
157	ขะเหยก	เขยก
158	ขัณฑสกร	ขัณฑสกร
159	ขัดคล่อง	ขัดข้อง
160	ขันฑสกร	ขัณฑสกร
161	ขันฑี	ขันที
162	ขันทสกร	ขัณฑสกร
163	ขัม	ขำ
164	ขาดดุน	ขาดดุล
165	ขาดดุลย์	ขาดดุล
166	ข้าวเหนียวมูล	ข้าวเหนียวมูน
167	ขี้เกลียด	ขี้เกียจ
168	ขี้เกียด	ขี้เกียจ
169	ขึ้นไฉ่	ขึ้นฉ่าย
170	ขึ้นช่าย	ขึ้นฉ่าย
171	ขึ้นไช่	ขึ้นฉ่าย
172	ขู่เข็น	ขู่เข็ญ
173	เขิ๊น	เขิน
174	เขิล	เขิน
175	เขิ๊ล	เขิน
176	โข่ง	โค่ง
177	โขมย	ขโมย
178	ไข่มุกข์	ไข่มุก
179	ไข่มุกด์	ไข่มุก
180	ไข่มุข	ไข่มุก
181	คฑา	คทา
182	คณะบดี	คณบดี
183	คธา	คทา
184	คนธรรม์	คนธรรพ์
185	คนอง	คะนอง

No.	Incorrect	Correct
186	คมสันต์	คมสัน
187	คมำ	คะมำ
188	ครรลัย	ครรไล
189	ครองราช	ครองราชย์
190	คร่ะ	คะ
191	ครัช	ครับ
192	ครัฟ	ครับ
193	ครัย	ใคร
194	คร่า	คะ
195	คร้า	คะ
196	คร๊า	คะ
197	คราคร่ำ	คลาคล่ำ
198	คร๊าบ	ครับ
199	คราย	ใคร
200	คราวะ	คารวะ
201	คร่ำครึก	คร่ำครึ
202	คริป	คลิป
203	คริสต์กาล	คริสตกาล
204	คริสต์จักร	คริสตจักร
205	คริสต ทศวรรษ	คริสต์ ทศวรรษ
206	คริสตมาส	คริสต์มาส
207	คริสต ศตวรรษ	คริสต์ ศตวรรษ
208	คริสตศักราช	คริสต์ศักราช
209	คริสต ศาสนา	คริสต์ศาสนา
210	คริสตศาสนิกชน	คริสต์ศาสนิกชน
211	ครุท	ครุฑ
212	ครุธ	ครุฑ
213	ครุ๊บ	ครับ
214	ครุศึกษา	คุรุศึกษา
215	ครุศึกสา	คุรุศึกษา
216	คฤหาส	คฤหาสน์
217	คฤหาสถ์	คฤหาสน์
218	คลองแคลง	ครองแครง
219	คลอบครุม	ครอบคลุม
220	คลอบคลุม	ครอบคลุม
221	คลอบคุม	ครอบคลุม
222	คลิ้ก	คลิก
223	คลิ๊ก	คลิก
224	คลินิค	คลินิก
225	คลีนิก	คลินิก
226	คลีนิค	คลินิก
227	คลุมเคลือ	คลุมเครือ
228	ค๊อกเทล	ค็อกเทล
229	คอนเสริต	คอนเสิร์ต
230	คอมมิวนิส	คอมมิวนิสต์
231	คอลัม	คอลัมน์
232	คอลัมภ์	คอลัมน์
233	คอลัมม์	คอลัมน์
234	ค่ะ	คะ
235	ค๊ะ	คะ
236	คะนึง	คะนึง
237	คัดสรรค์	คัดสรร
238	คัดสัน	คัดสรร
239	คันลัย	ครรไล
240	คับ	ครับ
241	ค๊บ	ครับ
242	คัฟ	ครับ
243	คัย	ใคร
244	ค๊า	คะ
245	ค่า	คะ
246	ค๊าบ	ครับ
247	คาเฟอีน	กาเฟอีน
248	คาราวะ	คารวะ
249	คำดุษฎี	คำสดุดี
250	คำนวน	คำนวณ
251	คำภีร์	คัมภีร์
252	คิมหัน	คิมหันต์
253	คีบับ	คือแบบ
254	คื่นฉ่าย	ขึ้นฉ่าย
255	คุ้กกี้	คุกกี้
256	คุ๊กกี้	คุกกี้
257	คุรุภัณฑ์	ครุภัณฑ์
258	คุรุศาสตร์	ครุศาสตร์
259	เค็ก	เค้ก
260	เค๊ก	เค้ก
261	เคเบิ้ล	เคเบิล
262	เครื่องลาง	เครื่องราง
263	เครื่องสำอาง	เครื่องสำอาง
264	เคลียจ	เครียด
265	เคลียด	เครียด
266	เคว๋ย	เหี่ย
267	เคหะสถาน	เคหสถาน
268	เคาเตอร์	เคาน์เตอร์
269	เคารวะ	คารวะ
270	เคี่ยน	เฒี่ยน
271	เคี่ยนตี	เฒี่ยนตี
272	แคตตาล็อก	แค็ตตาล็อก
273	แคตาล็อก	แค็ตตาล็อก
274	แค็บหมู	แคบหมู
275	แคปหมู	แคบหมู
276	แค็ปหมู	แคบหมู
277	แคระเกร็น	แคระแกร็น
278	แคระแกน	แคระแกร็น
279	แคลเซี่ยม	แคลเซียม
280	แคะแกน	แคระแกร็น

No.	Incorrect	Correct
281	แคะแกรน	แคระแกร็น
282	โคด	โคตร
283	โคตะระ	โคตร
284	โครงการณ์	โครงการ
285	โครต	โคตร
286	โคลงการ	โครงการ
287	โควต้า	โควตา
288	ใคร่ครวน	ใคร่ครวญ
289	ไค	ใคร
290	ไคร	ใคร
291	ฅน	คน
292	ฆรวาท	ฆราวาส
293	ฆรวาส	ฆราวาส
294	ฆ้อน	ค้อน
295	ฆาตรกร	ฆาตกร
296	ฆาตรกรรม	ฆาตกรรม
297	ฆารวาท	ฆราวาส
298	ฆารวาส	ฆราวาส
299	งบดุลย์	งบดุล
300	ง๊าบ	ครับ
301	งูสวัส	งูสวัด
302	งูสวัสดิ์	งูสวัด
303	จเข้	จระเข้
304	จงกลม	จงกรม
305	จงอย	จะงอย
306	จตุรัส	จัตุรัส
307	จรเข้	จระเข้
308	จรเม็ด	จะละเม็ด
309	จระเม็ด	จะละเม็ด
310	จราจล	จลาจล
311	จลศาสตร์	จลนศาสตร์
312	จละเม็ด	จะละเม็ด
313	จอมเวทย์	จอมเวท

No.	Incorrect	Correct
314	จะเข้	จระเข้
315	จะละหวั่น	จ้าละหวั่น
316	จั๊กจั่น	จักจั่น
317	จักร์	จักร
318	จักรพรรดิ์	จักรพรรดิ
319	จักรยานยน	จักรยานยนต์
320	จักรยานยนตร์	จักรยานยนต์
321	จักรวรรด	จักรวรรดิ
322	จักรวรรดิ์	จักรวรรดิ
323	จักรสาน	จักสาน
324	จักรสาร	จักสาน
325	จักวรรด	จักรวรรดิ
326	จักวรรดิ	จักรวรรดิ
327	จังเบย	จังเลย
328	จังเยย	จังเลย
329	จังรุย	จังเลย
330	จัย	ใจ
331	จัย	ใจ
332	จาย	ใจ
333	จารไน	จาระไน
334	จารบี	จาระบี
335	จาระเม็ด	จะละเม็ด
336	จาลเม็ด	จะละเม็ด
337	จ้าลหวั่น	จ้าละหวั่น
338	จาละเม็ด	จะละเม็ด
339	จำนงค์	จำนง
340	จำวัตร	จำวัด
341	จิง	จริง
342	จิตกร	จิตรกร
343	จิตลดา	จิตรลดา
344	จินตรนาการ	จินตนาการ
345	จินตะนาการ	จินตนาการ

No.	Incorrect	Correct
346	จินตระนาการ	จินตนาการ
347	จิปาฐะ	จิปาถะ
348	จีราฟ	ยีราฟ
349	จึย	ใจ
350	จึย	ใจ
351	จุงเบย	จังเลย
352	เจง	จริง
353	เจตจำนงค์	เจตจำนง
354	เจตนารมย์	เจตนารมณ์
355	เจ้าอาวาท	เจ้าอาวาส
356	เจียรไน	เจียระไน
357	โจทย์จัน	โจษจัน
358	ฉนั้น	ฉะนั้น
359	ฉนี้	ฉะนี้
360	ฉไน	ไฉน
361	ฉเพาะ	เฉพาะ
362	ฉะบับ	ฉบับ
363	ฉะเพาะ	เฉพาะ
364	ฉาน	ฉัน
365	ชนี	ชะนี
366	ชมด	ชะมด
367	ชมภู	ชมพู
368	ชลมาค	ชลมารค
369	ชลมาคร	ชลมารค
370	ชลมาร์ค	ชลมารค
371	ชลอ	ชะลอ
372	ช็อกโกเลต	ช็อกโกแลต
373	ช็อกโกแล็ต	ช็อกโกแลต
374	ช๊อกโกแลต	ช็อกโกแลต
375	ชอคโกแลต	ช็อกโกแลต
376	ช็อคโกแลต	ช็อกโกแลต
377	ช็อคโกแล็ต	ช็อกโกแลต

No.	Incorrect	Correct
378	ช๊อคโกแลต	ช็อกโกแลต
379	ชะมดชม้อย	ชมดชม้อย
380	ชะมะ	ใช่ไหม
381	ชะม๊า	ใช่ไหม
382	ชะอุม	ชอุ่ม
383	ชะอุ่ม	ชอุ่ม
384	ชักคะเย่อ	ชักเย่อ
385	ชัชวาลย์	ชัชวาล
386	ชันสูติ	ชันสูตร
387	ชั้ล	ฉัน
388	ช้าน	ฉัน
389	ชานชะลา	ชานชาลา
390	ช่าย	ใช่
391	ช่ายม้าย	ใช่ไหม
392	ชิมิ	ใช่ไหม
393	ชีพตักษัย	ชีพิตักษัย
394	ชีวะประวัติ	ชีวประวัติ
395	เชรี่ย	เหี่ย
396	เชิ๊ต	เชิ้ต
397	เชิร์ต	เชิ้ต
398	เชีย	เหี่ย
399	ใชเท้า	ไชเท้า
400	ไช้เท้า	ไชเท้า
401	ไช้โป๊	ไชโป๊
402	ไชโป้ว	ไชโป๊
403	ไชโป๊ว	ไชโป๊
404	ไช้โป๊ว	ไชโป๊
405	ซวดทรง	ทรวดทรง
406	ซะละเปา	ซาลาเปา
407	ซาละเปา	ซาลาเปา
408	ซาวด์เสียง	ซาวเสียง
409	ซีดเผือก	ซีดเผือด
410	ซีเม็น	ซีเมนต์
411	ซีเม็นต์	ซีเมนต์
412	ซีเมนท์	ซีเมนต์
413	ซีเม็นท์	ซีเมนต์
414	ซีรีย์	ซีรีส์
415	ซีรี่ย์	ซีรีส์
416	เซ็นต์ชื่อ	เซ็นชื่อ
417	เซ็นติเมตร	เซนติเมตร
418	เซล	เซลล์
419	แซกแซง	แทรกแซง
420	โซม	โทรม
421	ไซร้	ไซ้
422	ฌาณ	ฌาน
423	ฌาปณกิจ	ฌาปนกิจ
424	ญวณ	ญวน
425	ญัติ	ญัตติ
426	ญาต	ญาติ
427	ฎีกา	ฎีกา
428	ฐาณ	ฐาน
429	ฑัณฑ์	ทัณฑ์
430	ฑีฆายุโก	ทีฆายุโก
431	ฑูต	ทูต
432	ฑูต	ทูต
433	ดอกไม้จัน	ดอกไม้จันทน์
434	ดอกไม้จันท์	ดอกไม้จันทน์
435	ดอกไม้จันทร์	ดอกไม้จันทน์
436	ดัชต์	ดัตช์
437	ดัชท์	ดัตช์
438	ดัทช์	ดัตช์
439	ดาดดื่น	ดาษดื่น
440	ดารดาส	ดารดาษ
441	ดาวดึง	ดาวดึงส์
442	ดาวดึงษ์	ดาวดึงส์
443	ดาวโหลด	ดาวน์โหลด
444	ดาษฟ้า	ดาดฟ้า
445	ดาสดา	ดาษดา
446	ดำกฤษณา	ดำฤษณา
447	ดำรงค์	ดำรง
448	ดำริห์	ดำริ
449	ดำละเมื่อม	ดำมะเมื่อม
450	ดำหริ	ดำริ
451	ดีก่า	ดีกว่า
452	ดุลย์	ดุล
453	เดว	เดียว
454	เด๋ว	เดี๋ยว
455	เดะ	เดี๋ยว
456	เดินเหิร	เดินเหิน
457	เดียจฉัน	เดียดฉันท์
458	เดียจฉันท์	เดียดฉันท์
459	แด่ก	แดก
460	แด๊ก	แดก
461	แดรก	แดก
462	แดร่ก	แดก
463	ไดนาไมท์	ไดนาไมต์
464	ไดร์ฟ	ไดรฟ์
465	ตกร่องปล่องชิ้น	ตกล่องปล่องชิ้น
466	ตรรกกะ	ตรรกะ
467	ตรรกะศาสตร์	ตรรกศาสตร์
468	ตระกร้า	ตะกร้า
469	ตระโกน	ตะโกน
470	ตระเวณ	ตระเวน
471	ตระเวร	ตระเวน

No.	Incorrect	Correct
472	ตราสังข์	ตราสัง
473	ตรีโกน	ตรีโกณ
474	ตรึงตา	ตรึงตรา
475	ตลบแตลง	ตลบตะแลง
476	ตลึง	ตะลึง
477	ตะกละตะกราม	ตะกละตะกลาม
478	ตะก้า	ตะกร้า
479	ตะเกียดตะกาย	ตะเกียกตะกาย
480	ตะหงิด	ตงิด
481	ตะเอง	ตัวเอง
482	ตักบาต	ตักบาตร
483	ตักสัย	ตักษัย
484	ตัญหา	ตัณหา
485	ตัดเพ้อ	ตัดพ้อ
486	ตาปู	ตะปู
487	ตาลขโมย	ตานขโมย
488	ตาลปัต	ตาลปัตร
489	ตำหรับ	ตำรับ
490	ติดสัตว์	ติดสัด
491	เตง	ตัวเอง
492	เต๊นท์	เต็นท์
493	โต้ะ	โต๊ะ
494	ใต้ก๋ง	ไต้ก๋ง
495	ใต้ฝุ่น	ไต้ฝุ่น
496	ใต้หวัน	ไต้หวัน
497	ไต้เท้า	ใต้เท้า
498	ไตรตรึงศ์	ตรัยตรึงศ์
499	ไตรยางค์	ไตรยางศ์
500	ไตร่สวน	ไต่สวน
501	ถนนราดยาง	ถนนลาดยาง
502	ถนัดถะหนี่	ถนัดถนี่

No.	Incorrect	Correct
503	ถ่วงดุน	ถ่วงดุล
504	ถ่วงดุลย์	ถ่วงดุล
505	ถั่วพลู	ถั่วพู
506	ถ่ายตัว	ไถ่ตัว
507	ถุ้ย	ถุย
508	ถุ๊ย	ถุย
509	เถาว์	เถา
510	เถาวัล	เถาวัลย์
511	แถลงการ	แถลงการณ์
512	ทนง	ทระนง
513	ทนทายาท	ทนทายาด
514	ทนุถนอม	ทะนุถนอม
515	ทนุทนอม	ทะนุถนอม
516	ทนุบำรุง	ทะนุบำรุง
517	ทมิน	ทมิฬ
518	ทยาน	ทะยาน
519	ทรงกรด	ทรงกลด
520	ทรนง	ทระนง
521	ทรมานทรกรรม	ทรมาทรกรรม
522	ทรมาร	ทรมาน
523	ทรราชย์	ทรราช
524	ทรวดเซ	ซวดเซ
525	ทราบซ่าน	ซาบซ่าน
526	ทราบซึ้ง	ซาบซึ้ง
527	ทราบส้าน	ซาบซ่าน
528	ทลาย	ทะลาย
529	ทลายมะพร้าว	ทะลายมะพร้าว
530	ทศกัณฑ์	ทศกัณฐ์
531	ทศพิตราชธรรม	ทศพิศราชธรรม
532	ทศมาศ	ทศมาส

No.	Incorrect	Correct
533	ท้อฟฟี่	ทอฟฟี่
534	ท๊อฟฟี่	ทอฟฟี่
535	ทอมซิน	ทอนซิล
536	ทอสับ	โทรศัพท์
537	ทะนง	ทระนง
538	ทะมัย	ทำไม
539	ทะมาย	ทำไม
540	ทะโมน	ทโมน
541	ทะยอย	ทยอย
542	ทะแยง	ทแยง
543	ทะแยงมุม	ทแยงมุม
544	ทะลาย	ทลาย
545	ทะเลสาป	ทะเลสาบ
546	ทักขิณาวัฏ	ทักษิณาวรรต
547	ทักษิณาวัฏ	ทักษิณาวรรต
548	ทักษิน	ทักษิณ
549	ทับพี	ทัพพี
550	ทัมมัย	ทำไม
551	ทัมอะรัย	ทำอะไร
552	ทาญาต	ทายาท
553	ทาญาติ	ทายาท
554	ทามมัย	ทำไม
555	ทามมาย	ทำไม
556	ทารุน	ทารุณ
557	ท้าวความ	เท้าความ
558	ทำนุบำรุง	ทะนุบำรุง
559	ทำเนียม	ธรรมเนียม
560	ทำมะรงค์	ธำมรงค์
561	ทำวัด	ทำวัตร
562	ทิฏฐิ	ทิฐิ
563	ทิติ	ทิฐิ

No.	Incorrect	Correct
564	ทุกขกิริยา	ทุกรกิริยา
565	ทุกข์คติ	ทุคติ
566	ทุกขรกิริยา	ทุกรกิริยา
567	ทุจศีล	ทุศีล
568	ทุพลภาพ	ทุพพลภาพ
569	ทุภิกขภัย	ทุพภิกขภัย
570	ทุรกันดาน	ทุรกันดาร
571	ทุรกัลดาน	ทุรกันดาร
572	ทุรกัลดาล	ทุรกันดาร
573	ทูนกระหม่อม	ทูลกระหม่อม
574	ทูนกะหม่อม	ทูลกระหม่อม
575	ทูนเกล้าทูนกระหม่อม	ทูลเกล้าทูลกระหม่อม
576	ทูลกะหม่อม	ทูลกระหม่อม
577	ทูลหัว	ทูนหัว
578	เทคนิก	เทคนิค
579	เทคโนโลยี่	เทคโนโลยี
580	เทพพนม	เทพนม
581	เทเวศ	เทเวศร์
582	เทเวศน์	เทเวศร์
583	เทโวโรหนะ	เทโวโรหณะ
584	เท่ห์	เท่
585	เทอ	เธอ
586	เทอด	เทิด
587	เทอดทูล	เทิดทูน
588	เทอร์	เธอ
589	เทอว์	เธอ
590	เทิญ	เทอญ
591	เทิดทูล	เทิดทูน

No.	Incorrect	Correct
592	เทิม	เทอม
593	เทิร์ม	เทอม
594	แท๊กซี่	แท็กซี่
595	แทยง	ทแยง
596	โทมน	ทโมน
597	โทรทรรศน์	โทรทัศน์
598	โทรทัด	โทรทรรศน์
599	โทรทัศน์	โทรทรรศน์
600	โทรศัพย์	โทรศัพท์
601	โทรสับ	โทรศัพท์
602	โทสับ	โทรศัพท์
603	ธนาณัต	ธนาณัติ
604	ธนานัติ	ธนาณัติ
605	ธนาบัตร	ธนบัตร
606	ธรรมมะ	ธรรมะ
607	ธรรมาส	ธรรมาสน์
608	ธัญญพืช	ธัญพืช
609	ธัญยาหาร	ธัญญาหาร
610	ธำมรง	ธำมรงค์
611	ธำรงค์	ธำรง
612	ธิเบต	ทิเบต
613	ธุดง	ธุดงค์
614	ธุระกิจ	ธุรกิจ
615	เธอว์	เธอ
616	นกพิลาป	นกพิราบ
617	นกพิลาภ	นกพิราบ
618	นกอินทรีย์	นกอินทรี
619	นงเยา	นงเยาว์
620	นนทรีย์	นนทรี
621	นพดล	นภดล
622	นพปดล	นพปฎล
623	นวตกรรม	นวัตกรรม
624	นวลละออ	นวลลออ

No.	Incorrect	Correct
625	น็อต	นอต
626	น๊อต	นอต
627	น๊ะ	นะ
628	นักษัตริย์	นักษัตร
629	นัตถ์	นัตถุ์
630	นั่นล่ะ	นั่นแหละ
631	นั่นหล่ะ	นั่นแหละ
632	นัยตา	นัยน์ตา
633	นัยยะ	นัย
634	นา ๆ	นานา
635	นาฎกรรม	นาฏกรรม
636	นาที	นาที
637	นานับประการ	นานัปการ
638	นานาจิตตัง	นานาจิตตัง
639	น้าม	น้ำ
640	นายหลวง	ในหลวง
641	น่าร๊อค	น่ารัก
642	นาๆ	นานา
643	น้ำแข็งใส	น้ำแข็งไส
644	น้ำจัณท์	น้ำจัณฑ์
645	น้ำจัน	น้ำจัณฑ์
646	น้ำจันท์	น้ำจัณฑ์
647	น้ำมันก๊าซ	น้ำมันก๊าด
648	น้ำมันก๊าส	น้ำมันก๊าด
649	นิจสิน	นิจศีล
650	นิตยาสาร	นิตยสาร
651	นิทัศการ	นิทรรศการ
652	นิเทศ	นิเทศ
653	นิเทศก์	นิเทศ
654	นิเทส	นิเทศ
655	นิมนตร์	นิมนต์
656	นิมิตร	นิมิต

No.	Incorrect	Correct
657	นิรมิตร	นิมิต
658	นิรันด์	นิรันดร์
659	นิวัติ	นิวัต
660	นิเวศน์วิทยา	นิเวศวิทยา
661	นู๋	หนู
662	เนรมิตร	เนรมิต
663	เนืองนิจ	เนืองนิตย์
664	แน่นหนา	หนาแน่น
665	โน๊ต	โน้ต
666	โน้ท	โน้ต
667	โน๊ท	โน้ต
668	ในตา	นัยน์ตา
669	บรเพชร	บอระเพ็ด
670	บรเพ็ด	บอระเพ็ด
671	บรรดาล	บันดาล
672	บรรไดๆ	บันได
673	บรรทัดถาน	บรรทัดฐาน
674	บรรทึก	บันทึก
675	บรรเทิง	บันเทิง
676	บรรลือ	บันลือ
677	บริจาก	บริจาค
678	บริสุทธ	บริสุทธิ์
679	บริสุทธิ	บริสุทธิ์
680	บล๊อก	บล็อก
681	บล็อค	บล็อก
682	บ่วงบาต	บ่วงบาศ
683	บ่วงบาท	บ่วงบาศ
684	บ่วงบาศก์	บ่วงบาศ
685	บ่องตง	บอกตรง
686	บ่องตรง	บอกตรง
687	บอระเพชร	บอระเพ็ด
688	บะหรั่นดี	บรั่นดี
689	บังกาโล	บังกะโล
690	บังคโล	บังกะโล
691	บังสกุล	บังสุกุล
692	บังเอิน	บังเอิญ
693	บัญญัติไตรยางค์	บัญญัติไตรยางศ์
694	บัตรสนเท่	บัตรสนเท่ห์
695	บันทัด	บรรทัด
696	บันทุก	บรรทุก
697	บันลุ	บรรลุ
698	บันเลง	บรรเลง
699	บัลลังค์	บัลลังก์
700	บางลำภู	บางลำพู
701	บาดทยัก	บาดทะยัก
702	บาดพระ	บาตรพระ
703	บาดหลวง	บาทหลวง
704	บาต	บาตร
705	บาตพระ	บาตรพระ
706	บาททะยัก	บาดทะยัก
707	บาทบงส์	บาทบงสุ์
708	บาทยัก	บาดทะยัก
709	บานเกร็ด	บานเกล็ด
710	บาสเก็ตบอล	บาสเกตบอล
711	บำเน็จ	บำเหน็จ
712	บำเหน็ด	บำเหน็จ
713	บิณฑบาตร	บิณฑบาต
714	บิณฑบาท	บิณฑบาต
715	บิดพริ้ว	บิดพลิ้ว
716	บีทา	บีฑา
717	บุคคลากร	บุคลากร
718	บุคคลิก	บุคลิก
719	บุคคลิกภาพ	บุคลิกภาพ
720	บุคคลิค	บุคลิก
721	บุคล	บุคคล
722	บุคบุคคลิคภาพ	บุคลิกภาพ
723	บุคลิค	บุคลิก
724	บุคลิคภาพ	บุคลิกภาพ
725	บุปผาชาติ	บุปผชาติ
726	บุศราคำ	บุษราคัม
727	บุษราคำ	บุษราคัม
728	บูชายัน	บูชายัญ
729	บูชายันต์	บูชายัญ
730	บูรณะปฏิสังขรณ์	บูรณปฏิสังขรณ์
731	เบญจเพศ	เบญจเพส
732	เบญจมาส	เบญจมาศ
733	เบ็นซิน	เบนซิน
734	เบนซิล	เบนซิน
735	เบรค	เบรก
736	เบอ	เบอร์
737	เบ้อเร่อเท่อ	เบ้อเริ่มเทิ่ม
738	แบ็คทีเรีย	แบคทีเรีย
739	แบ็งก์	แบงก์
740	แบ๊งก์	แบงก์
741	แบงค์	แบงก์
742	แบ็งค์	แบงก์
743	แบ๊งค์	แบงก์
744	แบ็ดมินตัน	แบดมินตัน
745	โบนัท	โบนัส
746	โบราน	โบราณ
747	โบว์ผูกผม	โบผูกผม
748	ใบกระเพรา	ใบกะเพรา
749	ปฏิสัมพันธ์	ปฏิสัมพันธ์
750	ปฏิกริยา	ปฏิกิริยา
751	ปฏิธาน	ปณิธาน
752	ปฏิธิน	ปฏิทิน

No.	Incorrect	Correct
753	ปฏิปัก	ปฏิปักษ์
754	ปฏิสังขร	ปฏิสังขรณ์
755	ปฏิสัณฐาน	ปฏิสันถาร
756	ปฏิสันถาน	ปฏิสันถาร
757	ปฐมนิเทศก์	ปฐมนิเทศ
758	ปฐมนิเทศน์	ปฐมนิเทศ
759	ปติทิน	ปฏิทิน
760	ปติมากรรม	ประติมากรรม
761	ปติเสด	ปฏิเสธ
762	ปนิธาน	ปณิธาน
763	ปรณนิบัติ	ปรนนิบัติ
764	ปรมณู	ปรมาณู
765	ปรมานู	ปรมาณู
766	ปรสิท	ปรสิต
767	ประกายพฤกษ์	ประกายพรึก
768	ประกาศณียบัตร	ประกาศนียบัตร
769	ประกาษิต	ประกาศิต
770	ประกาสิต	ประกาศิต
771	ประจัญบาล	ประจัญบาน
772	ประจัญหน้า	ประจันหน้า
773	ประจัญห้อง	ประจันห้อง
774	ประจัน	ประจัญ
775	ประจันบาน	ประจัญบาน
776	ประจันบาล	ประจัญบาน
777	ประจำการณ์	ประจำการ
778	ประชาทันฑ์	ประชาทัณฑ์
779	ประณีประนอม	ประนีประนอม
780	ประดิษฐ์ประดอย	ประดิดประดอย

No.	Incorrect	Correct
781	ประติดประต่อ	ปะติดปะต่อ
782	ประติพัทธ์	ปฏิพัทธ์
783	ประทะ	ปะทะ
784	ประนด	ประณต
785	ประนต	ประณต
786	ประนตน้อม	ประณตน้อม
787	ประนม	ประณม
788	ประนาม	ประณาม
789	ประนิธาน	ปณิธาน
790	ประนีต	ประณีต
791	ประนีประณอม	ประนีประนอม
792	ประนีปราณอม	ประนีประนอม
793	ประนีปรานอม	ประนีประนอม
794	ประมาน	ประมาณ
795	ประเมิณ	ประเมิน
796	ประโยชน์โพธิผล	ประโยชน์โพดผล
797	ประโยชน์โภชน์ผล	ประโยชน์โพดผล
798	ประโยชน์โภชผล	ประโยชน์โพดผล
799	ประสบการ	ประสบการณ์
800	ประสพการณ์	ประสบการณ์
801	ประสพความสำเร็จ	ประสบความสำเร็จ
802	ประสพผลสำเร็จ	ประสบผลสำเร็จ

No.	Incorrect	Correct
803	ประสูต	ประสูติ
804	ประสูตร	ประสูติ
805	ประหัดประหาร	ประหัตประหาร
806	ประหัตถ์ประหาร	ประหัตประหาร
807	ปรัศนีย์	ปรัศนี
808	ปรากฏ	ปรากฏ
809	ปรากด	ปรากฏ
810	ปราณีต	ประณีต
811	ปราณีประนอม	ประนีประนอม
812	ปราณีปราศัย	ปรานีปราศรัย
813	ปราถนา	ปรารถนา
814	ปรานีประนอม	ประนีประนอม
815	ปราศจาค	ปราศจาก
816	ปราศัย	ปราศรัย
817	ปรำปรา	ปรัมปรา
818	ปริปากเงียบ	ปิดปากเงียบ
819	ปริมนฑล	ปริมณฑล
820	ปล้นสดมภ์	ปล้นสะดม
821	ปลาด	ประหลาด
822	ปล่าว	เปล่า
823	ปลาวาล	ปลาวาฬ
824	ปแล่ม	ปะแล่ม
825	ปวารนา	ปวารณา
826	ปองดอง	ปรองดอง
827	ปะปา	ประปา
828	ปะแป้ง	ประแป้ง
829	ปะรำปะรา	ปรัมปรา
830	ปะละ	เปล่า

No.	Incorrect	Correct
831	ป่ะละ	เปล่า
832	ปะหลาด	ประหลาด
833	ปักษิณ	ปักษิน
834	ปั้ม	ปั๊ม
835	ปากฏ	ปรากฏ
836	ปาฏิโมกข์	ปาติโมกข์
837	ปาฏิหารย์	ปาฏิหาริย์
838	ปาฏิหาริย์	ปาฏิหาริย์
839	ปาติหารย์	ปาฏิหาริย์
840	ปายหนาย	ไปไหน
841	ปาร์เก้	ปาร์เกต์
842	ปาริชาติ	ปาริชาต
843	ป่าว	เปล่า
844	ปิคนิค	ปิกนิก
845	ปิติ	ปีติ
846	ปิโตรเลี่ยม	ปิโตรเลียม
847	ปุโลหิต	ปุโรหิต
848	เปง	เป็น
849	เปงยัย	เป็นอะไร
850	เปงใย	เป็นอะไร
851	เปงรัย	เป็นอะไร
852	เปงราย	เป็นอะไร
853	เปงไร	เป็นอะไร
854	เปงลัย	เป็นอะไร
855	เปงลาย	เป็นอะไร
856	เปน	เป็น
857	เปนยัย	เป็นอะไร
858	เปนใย	เป็นอะไร
859	เปนไย	เป็นอะไร
860	เปนรัย	เป็นอะไร
861	เปนราย	เป็นอะไร
862	เปนลัย	เป็นอะไร
863	เปนลาย	เป็นอะไร

No.	Incorrect	Correct
864	เปอเซน	เปอร์เซ็นต์
865	เปอเซนต์	เปอร์เซ็นต์
866	เปอร์เซนต์	เปอร์เซ็นต์
867	เป๋อเล๋อ	เป๋อเหลอ
868	แปล่ม	ปะแล่ม
869	ไป๊	ไป
870	ไปรษณี	ไปรษณีย์
871	ไปรษนี	ไปรษณีย์
872	ไปรษนีย์	ไปรษณีย์
873	ผเชิญ	เผชิญ
874	ผไท	ไผท
875	ผรุทสวาท	ผรุสวาท
876	ผรุสวาส	ผรุสวาท
877	ผลลัพท์	ผลลัพธ์
878	ผลอย	ผล็อย
879	ผลัด	ผัด
880	ผลัดนี้ผลัดสิน	ผัดหนี้ผัดสิน
881	ผลัดผ่อน	ผัดผ่อน
882	ผลัดวันประกันพรุ่ง	ผัดวันประกันพรุ่ง
883	ผลัดเวลา	ผัดเวลา
884	ผลัดหนี้	ผัดหนี้
885	ผลานิสงฆ์	ผลานิสงส์
886	ผลุดลุกผลุดนั่ง	ผุดลุกผุดนั่ง
887	ผเอิญ	เผอิญ
888	ผะเชิญ	เผชิญ
889	ผะดุง	ผดุง
890	ผะไท	ไผท
891	ผะอบ	ผอบ
892	ผะเอิญ	เผอิญ
893	ผัด	ผลัด

No.	Incorrect	Correct
894	ผัดไท	ผัดไทย
895	ผัดเปลี่ยน	ผลัดเปลี่ยน
896	ผัดผ้า	ผลัดผ้า
897	ผัดวันประกันพรุ่ง	ผัดวันประกันพรุ่ง
898	ผัดเวร	ผลัดเวร
899	ผาสุก	ผาสุก
900	ผาสุข	ผาสุก
901	ผิดสำแดง	ผิดสำแลง
902	ผีซ้ำด้ามพลอย	ผีซ้ำด่ำพลอย
903	ผีพุ่งใต้	ผีพุ่งไต้
904	ผูกพันธ์	ผูกพัน
905	ผู้เยา	ผู้เยาว์
906	เผลอเรอ	เผอเรอ
907	เผอไผ	เผลอไผล
908	เผ่าพัน	เผ่าพันธุ์
909	เผ่าพันธ์	เผ่าพันธุ์
910	แผงฤทธิ์	แผลงฤทธิ์
911	แผ่ซ้าน	แผ่ซ่าน
912	แผนการณ์	แผนการ
913	แผ่ส้าน	แผ่ซ่าน
914	ฝรั่งเศษ	ฝรั่งเศส
915	ฝื้น	ฟื่น
916	ฝากครร	ฝากครรภ์
917	ฝาลมี	ฝาละมี
918	ฝีดาด	ฝีดาย
919	ฝึกปลือ	ฝึกปรือ
920	ฝึกปือ	ฝึกปรือ
921	ฝุด	สุด
922	ใฝ	ไฝ
923	พงพันธุ์	พงศ์พันธุ์
924	พงษ์พันธุ์	พงศ์พันธุ์

No.	Incorrect	Correct
925	พจณานุกรม	พจนานุกรม
926	พญาณ	พยาน
927	พญาบาท	พยาบาท
928	พณ	ฯพณฯ
929	พณฯ	ฯพณฯ
930	พณิช	พาณิช
931	พณิชย์	พาณิชย์
932	พนิช	พาณิช
933	พนิชย์	พาณิชย์
934	พเนียด	เพนียด
935	พแนง	พะแนง
936	พยอม	พะยอม
937	พยักเพยิด	พยักพเยิด
938	พยัค	พยัคฆ์
939	พยัคร	พยัคฆ์
940	พยัฆ	พยัคฆ์
941	พยาณ	พยาน
942	พยาบาตร	พยาบาท
943	พยูน	พะยูน
944	พรมจรรย์	พรหมจรรย์
945	พรรณา	พรรณนา
946	พร้อมเพียง	พร้อมเพรียง
947	พระชนพรรษา	พระชนมพรรษา
948	พระวงค์	พระวงศ์
949	พระองศ์	พระองค์
950	พรามณ์	พราหมณ์
951	พราห์มณ์	พราหมณ์
952	พร่ำพรอด	พร่ำพลอด
953	พริ้ว	พลิ้ว
954	พฤศภาคม	พฤษภาคม
955	พฤษจิกายน	พฤศจิกายน
956	พลกำลัง	พละกำลัง
957	พละการ	พลการ
958	พละศึกษา	พลศึกษา
959	พลัดพลาก	พลัดพราก
960	พลางตา	พรางตา
961	พลาสติค	พลาสติก
962	พวักพวน	พะวักพะวน
963	พวังศ์	ภวังค์
964	พว้าพะวง	พะว้าพะวง
965	พหูสูด	พหูสูต
966	พหูสูต	พหูสูต
967	พหูสูตร	พหูสูต
968	พ่อง	พ่อ
969	พะเนียด	เพนียด
970	พะยะค่ะ	พ่ะย่ะค่ะ
971	พะยัก	พยัก
972	พะยักพะเยิด	พยักพเยิด
973	พะยักเพยิด	พยักพเยิด
974	พะยาน	พยาน
975	พะยุง	พยุง
976	พะเยีย	พเยีย
977	พะโยม	โพยม
978	พะวัง	ภวังค์
979	พะเอิญ	เพอิญ
980	พังทะลาย	พังทลาย
981	พันธบัติ	พันธบัตร
982	พันธะกิจ	พันธกิจ
983	พันธะสัญญา	พันธสัญญา
984	พันธารักษ์	ภัณฑารักษ์
985	พันธุ์ทาง	พันทาง
986	พันลวัน	พัลวัน
987	พัศดุ	พัสดุ
988	พัสดี	พัศดี
989	พากพูมิ	ภาคภูมิ
990	พากย์	พากย์
991	พาณิชย	พาณิชย์
992	พานิช	พาณิช
993	พานิชย์	พาณิชย์
994	พานิชย	พาณิชย์
995	พาระกิจ	ภารกิจ
996	พาระกิด	ภารกิจ
997	พาลจะเป็นลม	พานจะเป็นลม
998	พาหุรัต	พาหุรัด
999	พาหุรัตน์	พาหุรัด
1000	พิณภาทย์	พิณพาทย์
1001	พิธีรีตรอง	พิธีรีตอง
1002	พิบูลย์	พิบูล
1003	พิพากย์	พิพากษ์
1004	พิพาทย์	พิพาท
1005	พิพิทธภัณฑ์	พิพิธภัณฑ์
1006	พิราพ	พิราบ
1007	พิเรน	พิเรนทร์
1008	พิโรท	พิโรธ
1009	พิศดาน	พิสดาร
1010	พิศดาร	พิสดาร
1011	พิศมัย	พิสมัย
1012	พิศวาท	พิศวาส
1013	พิสวง	พิศวง
1014	พิสวาท	พิศวาส
1015	พิสวาส	พิศวาส
1016	พิสูตร	พิสูจน์
1017	พึมพัม	พึมพำ
1018	พุชงค์	ภุชงค์
1019	พุดตาล	พุดตาน
1020	พุดตาล	พุดตาน
1021	พุทธชาติ	พุทธชาด

No.	Incorrect	Correct
1022	พุทธตาล	พุดตาน
1023	พูมใจ	ภูมิใจ
1024	พู่ระหงส์	พู่ระหง
1025	เพ็ชร	เพชร
1026	เพชรฆาต	เพชฌฆาต
1027	เพ็ชรฆาต	เพชฌฆาต
1028	เพตรา	เภตรา
1029	เพทภัย	เภทภัย
1030	เพนียด	เพนียด
1031	เพยีย	พเยีย
1032	เพริดพริ้ง	เพริศพริ้ง
1033	เพรียบพร้อม	เพียบพร้อม
1034	เพศภัย	เภทภัย
1035	เพศสัมพันธุ์	เพศสัมพันธ์
1036	เพิมเติม	เพิ่มเติม
1037	เพิ่มพูล	เพิ่มพูน
1038	เพิ่ล	เพื่อน
1039	เพิศพริ้ง	เพริศพริ้ง
1040	เพื่อล	เพื่อน
1041	แพทย์ศาสตร์	แพทยศาสตร์
1042	แพนง	พะแนง
1043	โพชงค์	โพชฌงค์
1044	โพธิ์ดำ	โพดำ
1045	โพธิ์แดง	โพแดง
1046	โพธิ์ทะเล	โพทะเล
1047	โพนทนา	โพนทะนา
1048	โพรดก	โพระดก
1049	โพศพ	โพสพ
1050	โพสภ	โพสพ
1051	ไพฑูรย์	ไพฑูรย์
1052	ฟลิม	ฟิล์ม
1053	ฟังก์ชั่น	ฟังก์ชัน

No.	Incorrect	Correct
1054	ฟันครุฑ	ฟันคุด
1055	ฟันครุท	ฟันคุด
1056	ฟาธอม	ฟาทอม
1057	ฟ่าฟาง	ฝ้าฟาง
1058	ฟามรัก	ความรัก
1059	ฟามลัก	ความรัก
1060	ฟาย	ควาย
1061	ฟิม์ล	ฟิล์ม
1062	ฟิมล์	ฟิล์ม
1063	ฟิลม์	ฟิล์ม
1064	ฟิลิปดา	พิลิปดา
1065	ฟิว	ฟิวส์
1066	ฟุทบอล	ฟุตบอล
1067	ฟูลสแกป	ฟุลสแก๊ป
1068	เฟิร์น	เฟิน
1069	แฟชัน	แฟชั่น
1070	แฟทอม	ฟาทอม
1071	แฟธอม	ฟาทอม
1072	ไฟแชค	ไฟแช็ก
1073	ไฟแช็ค	ไฟแช็ก
1074	ภคิณี	ภคินี
1075	ภววิสัย	ภาววิสัย
1076	ภวัง	ภวังค์
1077	ภาคฑัณฑ์	ภาคทัณฑ์
1078	ภาคภูม	ภาคภูมิ
1079	ภาพพจ	ภาพพจน์
1080	ภาพยนต์	ภาพยนตร์
1081	ภาระกิจ	ภารกิจ
1082	ภาระกิด	ภารกิจ
1083	ภาววิสัย	ภววิสัย
1084	ภาวะการ	ภาวการณ์
1085	ภาวะการณ์	ภาวการณ์
1086	ภาวะวิสัย	ภาววิสัย

No.	Incorrect	Correct
1087	ภู่กัน	พู่กัน
1088	ภูติผี	ภูตผี
1089	ภูติผีปีศาจ	ภูตผีปีศาจ
1090	ภูมใจ	ภูมิใจ
1091	ภูมลำเนา	ภูมิลำเนา
1092	ภู่ระหง	พู่ระหง
1093	ภู่ระหง	พู่ระหง
1094	ภูวนัย	ภูวไนย
1095	โภชงค์	โพชฌงค์
1096	โภรดก	โพระดก
1097	โภศพ	โพสพ
1098	มกุฎ	มกุฎ
1099	มงกุฎ	มงกุฎ
1100	มณเฑียร	มนเทียร
1101	มณฑป	มณฑป
1102	มนฑป	มณฑป
1103	มนเฑียร	มนเทียร
1104	มนเฑียรบาล	มนเทียรบาล
1105	มนทิน	มลทิน
1106	มนุษยวิทยา	มานุษยวิทยา
1107	มนุษย์สัมพันธ์	มนุษยสัมพันธ์
1108	มรกด	มรกต
1109	มรณะกรรม	มรณกรรม
1110	มรณะภาพ	มรณภาพ
1111	มรึง	มึง
1112	มฤตญู	มฤตยู
1113	มว๊าก	มาก
1114	มหรรนพ	มหรรณพ
1115	มหรรศจรรย์	มหัศจรรย์
1116	มหรศพ	มหรสพ
1117	มหะหมัด	มะหะหมัด
1118	มหันนพ	มหรรณพ

No.	Incorrect	Correct
1119	มหาศาล	มหาศาล
1120	มหาสาน	มหาศาล
1121	มหาหิงค์	มหาหิงคุ์
1122	มหาหิงส์	มหาหิงคุ์
1123	มเหศรี	มเหสี
1124	มเหศักดิ์	มเหสักข์
1125	มเหศี	มเหสี
1126	มเหษี	มเหสี
1127	มไหสวรรค์	มไหศวรรย์
1128	มอเตอร์ไซ	มอเตอร์ไซค์
1129	มอเตอร์ไซด์	มอเตอร์ไซค์
1130	มอเตอร์ไซต์	มอเตอร์ไซค์
1131	มอเตอร์ไซท์	มอเตอร์ไซค์
1132	ม่อห้อม	ม่อฮ่อม
1133	ม่อฮ่อม	ม่อฮ่อม
1134	มะรุ	ไม่รู้
1135	มะลังมะเลือง	มลังเมลือง
1136	มะหะมัด	มะหะหมัด
1137	มักกะโรนี	มะกะโรนี
1138	มัคคุเทศ	มัคคุเทศก์
1139	มัคคุเทศน์	มัคคุเทศก์
1140	มังสวิรัต	มังสวิรัติ
1141	มัธยัสต์	มัธยัสถ์
1142	มันฑณศิลป์	มัณฑนศิลป์
1143	มันทนศิลป์	มัณฑนศิลป์
1144	มั้ย	ไหม
1145	มั๊ย	ไหม
1146	มัสตาด	มัสตาร์ด
1147	มัสยา	มัศยา
1148	มัสยา	มัศยา
1149	มัสหมั่น	มัสมั่น
1150	มาตราการ	มาตรการ
1151	มาตราฐาน	มาตรฐาน
1152	ม่าเปงไร	ไม่เป็นไร
1153	ม่ายเปงรัย	ไม่เป็นไร
1154	มาลายู	มลายู
1155	มึง	มึง
1156	มืดมนต์	มืดมน
1157	มืดมล	มืดมน
1158	มุขตลก	มุกตลก
1159	เมพขิง	เทพจริง
1160	เมริง	มึง
1161	เมหสี	มเหสี
1162	เมิง	มึง
1163	เมิง	มึง
1164	เมื่อกี๊	เมื่อกี้
1165	แมงกระพรุน	แมงกะพรุน
1166	แมงดา	แมลงดา
1167	แมงภู่	แมลงภู่
1168	แมงวัน	แมลงวัน
1169	แมงสาบ	แมลงสาบ
1170	แมงสาป	แมลงสาบ
1171	แมลงพู่	แมลงภู่
1172	แมลงมุม	แมงมุม
1173	แมลงสาป	แมลงสาบ
1174	ไม๊	ไหม
1175	ไมค์	ไมล์
1176	ไมร์	ไมล์
1177	ไม่รุ	ไม่รู้
1178	ยรรยงค์	ยรรยง
1179	ยวณ	ญวน
1180	ยศฐาบรรดาศักดิ์	ยศถาบรรดาศักดิ์
1181	ย่อมเยาว์	ย่อมเยา
1182	ยะโส	ยโส
1183	ยัติภัง	ยัติภังค์
1184	ยากเข็น	ยากเข็ญ
1185	ยาเกล็ด	ยาเกร็ด
1186	ยานัด	ยานัตถุ์
1187	ยานัตถ์	ยานัตถุ์
1188	ยาเสพย์ติด	ยาเสพติด
1189	ยุงก้นป่อง	ยุงก้นปล่อง
1190	เยอรมัน	เยอรมนี
1191	เยอรมันนี	เยอรมนี
1192	เยาวัย	เยาว์วัย
1193	เยื่อไย	เยื่อใย
1194	โยธวาฑิต	โยธวาทิต
1195	ใยดี	ไยดี
1196	ใยไพ	ไยไพ
1197	ไยแมงมุม	ใยแมงมุม
1198	รกชัฏ	รกชัฏ
1199	รถเมย์	รถเมล์
1200	รถยนตร์	รถยนต์
1201	รนรงค์	รณรงค์
1202	รมณี	รมณีย์
1203	รมนีย์	รมณีย์
1204	ร่วมหัวจมท้าย	ล่มหัวจมท้าย
1205	รสชาด	รสชาติ
1206	รเห็จ	ระเห็จ
1207	ร้องให้	ร้องไห้
1208	รอดรอนสิทธิ์	ลิดรอนสิทธิ์
1209	ร่อยหลอ	ร่อยหรอ
1210	ระบบนิเวศน์	ระบบนิเวศ
1211	ระเบิ่งเซิ้งแซ่	ระเบงเซ็งแซ่
1212	ระยาง	รยางค์

No.	Incorrect	Correct
1213	ระเห็ด	ระเห็จ
1214	รังเกลียด	รังเกียจ
1215	รังเกียจเดียดฉัน	รังเกียจเดียดฉันท์
1216	รังเกียด	รังเกียจ
1217	รังควาญ	รังควาน
1218	รังศี	รังสี
1219	รังษี	รังสี
1220	รังสิมันต์	รังสิมันตุ์
1221	รัชการ	รัชกาล
1222	รัชฎาภิเษก	รัชดาภิเษก
1223	รัญจวญ	รัญจวน
1224	รัญจวณ	รัญจวน
1225	รัฐวิสาหะกิจ	รัฐวิสาหกิจ
1226	รันจวน	รัญจวน
1227	รัษมี	รัศมี
1228	รัสมี	รัศมี
1229	รากเง่า	รากเหง้า
1230	ราชปแตน	ราชปะแตน
1231	ราชภัฎ	ราชภัฏ
1232	ราชวงค์	ราชวงศ์
1233	ราชสาสน์	ราชสาส์น
1234	ราชันย์	ราชัน
1235	ราดยาง	ลาดยาง
1236	รานรอน	ราญรอน
1237	ราพนาสูร	ราพณาสูร
1238	ราศรี	ราศี
1239	ราสี	ราศี
1240	รำคราญ	รำคาญ
1241	รำคาน	รำคาญ
1242	ริดรอน	ลิดรอน
1243	ริ้บบิ้น	ริบบิ้น
1244	รื่นรมณ์	รื่นรมย์

No.	Incorrect	Correct
1245	รู้	รู้
1246	รู้เท่าไม่ถึงการ	รู้เท่าไม่ถึงการณ์
1247	รูปพันธ์	รูปพรรณ
1248	เรย	เลย
1249	เรห็จ	ระเห็จ
1250	เริ่ดหรู	เลิศหรู
1251	เรี่ยราย	เรี่ยไร
1252	แร็กเก็ต	แร็กเกต
1253	แรงเงา	แรเงา
1254	โรงเตี้ยม	โรงเตี๊ยม
1255	โรงธารกำนัล	โรงธารคำนัล
1256	โรมันคาธอลิค	โรมันคาทอลิก
1257	ไร่เรียง	ไล่เลียง
1258	ไร่เลียง	ไล่เลียง
1259	ฤกษ์ผานาที	ฤกษ์พานาที
1260	ฤทธ์	ฤทธิ์
1261	ฤทธ	ฤทธิ์
1262	ฤษี	ฤๅษี
1263	ฤา	ฤๅ
1264	ฤาษี	ฤๅษี
1265	ฤๅชา	ลือชา
1266	ลดลาวาศอก	ลดราวาศอก
1267	ลดาวัณย์	ลดาวัลย์
1268	ลนหาที่	รนหาที่
1269	ลมปราน	ลมปราณ
1270	ลมหวล	ลมหวน
1271	ลเมียดลไม	ละเมียดละไม
1272	ลโมบ	ละโมบ
1273	ล็อกเก็ต	ล็อกเกต

No.	Incorrect	Correct
1274	ล็อกแล็ก	ลอกแลก
1275	ล่อกแล่ก	ลอกแลก
1276	ลองใน	ลองไน
1277	ล็อตเตอรี่	ลอตเตอรี่
1278	ลอมชอม	รอมชอม
1279	ลเอียดลออ	ละเอียดลออ
1280	ละคอน	ละคร
1281	ละมุนลไม	ละมุนละไม
1282	ละมุลละมัย	ละมุนละไม
1283	ละโมภ	ละโมบ
1284	ละโมภโลภมาก	ละโมบโลภมาก
1285	ละออ	ลออ
1286	ละเอียดละออ	ละเอียดลออ
1287	ล๊ก	รัก
1288	ลักกะปิดลักกะเปิด	ลักปิดลักเปิด
1289	ล๊กซ์	รัก
1290	ลักเพท	ลักเพศ
1291	ลักเพส	ลักเพศ
1292	ล๊ค	รัก
1293	ลัดดาวัลย์	ลดาวัลย์
1294	ลาดตระเวณ	ลาดตระเวน
1295	ลาดหน้า	ราดหน้า
1296	ลายเซ็นต์	ลายเซ็น
1297	ลาวัลย์	ลาวัณย์
1298	ลำคาณ	รำคาญ
1299	ลำใย	ลำไย
1300	ลำใส้	ลำไส้
1301	ลิงค์	ลิงก์
1302	ลิปต์	ลิฟต์
1303	ลิปสติค	ลิปสติก

No.	Incorrect	Correct
1304	ลิฟ	ลิฟต์
1305	ลิฟท์	ลิฟต์
1306	ลุกลี้ลุกรน	ลุกลี้ลุกลน
1307	ลูกเกตุ	ลูกเกด
1308	ลูกนิมิตร	ลูกนิมิต
1309	ลูกบาศ	ลูกบาศก์
1310	เล่กระเท่	เล่ห์กระเท่ห์
1311	เล่นพิเรน	เล่นพิเรนทร์
1312	เล่นพิเรนท์	เล่นพิเรนทร์
1313	เลิกลา	เลิกรา
1314	เลือกสรรค์	เลือกสรร
1315	เลือดกลบปาก	เลือดกบปาก
1316	เลือนลาง	เลือนราง
1317	โลกาภิวัฒน์	โลกาภิวัตน์
1318	โล่ห์	โล่
1319	ไล่เรี่ย	ไล่เลี่ย
1320	ไล่เรียง	ไล่เลียง
1321	ฦๅชา	ลือชา
1322	วงกฎ	วงกต
1323	วงศาคนาญาติ	วงศาคณาญาติ
1324	วงษ์วาน	วงศ์วาน
1325	วนิพพก	วนิพก
1326	วรรณโรค	วัณโรค
1327	วอลเล่ย์บอล	วอลเลย์บอล
1328	วักซีน	วัคซีน
1329	วันทยวุธ	วันทยาวุธ
1330	วันทยาหัตถ์	วันทยหัตถ์
1331	วันโรค	วัณโรค
1332	วางกล้าม	วางก้าม
1333	วาฑิต	วาทิต
1334	วาทะกรรม	วาทกรรม
1335	ว๊าย	ว้าย
1336	วายชน	วายชนม์
1337	วายปราน	วายปราณ
1338	วาระดิถี	วารดิถี
1339	วาล	วาฬ
1340	วิ่งเปรี้ยว	วิ่งเปี้ยว
1341	วิ่งผัด	วิ่งผลัด
1342	วิจาร	วิจารณ์
1343	วิจารณ	วิจารณ์
1344	วิจารณญาน	วิจารณญาณ
1345	วิจารย์	วิจารณ์
1346	วิญญาณ	วิญญาณ
1347	วิญญาน	วิญญาณ
1348	วิตฐาน	วิตถาร
1349	วิตตามิน	วิตามิน
1350	วิตถาน	วิตถาร
1351	วิตถาล	วิตถาร
1352	วิตะมิน	วิตามิน
1353	วินาฑี	วินาที
1354	วินาศะกรรม	วินาศกรรม
1355	วิบัต	วิบัติ
1356	วิปโยก	วิปโยค
1357	วิปลาศ	วิปลาส
1358	วิปัสนา	วิปัสสนา
1359	วิพากย์	วิพากษ์
1360	วิพากย์วิจารณ์	วิพากษ์วิจารณ์
1361	วิพากย์วิจารย์	วิพากษ์วิจารณ์
1362	วิพากษ์วิจารย์	วิพากษ์วิจารณ์
1363	วิศวะกร	วิศวกร
1364	วิศวะกรรม	วิศวกรรม
1365	วิหารคด	วิหารคด
1366	วีดิโอ	วิดีโอ
1367	วีดีทัศน์	วีดิทัศน์
1368	วีดีโอ	วิดีโอ
1369	เวง	เวร
1370	เวทมนต์	เวทมนตร์
1371	เวทย์	เวท
1372	เวทย์มนต์	เวทมนตร์
1373	เวทย์มนตร์	เวทมนตร์
1374	เวน	เวร
1375	เวรคืน	เวนคืน
1376	เวียตนาม	เวียดนาม
1377	ไวยกรณ์	ไวยากรณ์
1378	ศตรู	ศัตรู
1379	ศรีษะ	ศีรษะ
1380	ศรีษะเกษ	ศรีสะเกษ
1381	ศไล	ไศล
1382	ศศิทร	ศศิธร
1383	ศักดิ์	ศักย์
1384	ศัทธา	ศรัทธา
1385	ศัพย์	ศัพท์
1386	ศานุศิษย์	สานุศิษย์
1387	ศาสตราวุธ	ศัสตราวุธ
1388	ศิลป์	ศิลปะ
1389	ศิลป	ศิลปะ
1390	ศิลปะการแสดง	ศิลปะการแสดง
1391	ศิลปะกรรม	ศิลปกรรม
1392	ศิลปะวัฒนธรรม	ศิลปะวัฒนธรรม
1393	ศิลปะวัตถุ	ศิลปวัตถุ
1394	ศิลปะศึกษา	ศิลปศึกษา
1395	ศิวิไล	ศิวิไลซ์

No.	Incorrect	Correct
1396	ศีรษะเกษ	ศรีสะเกษ
1397	ศีศรีสะเกศ	ศรีสะเกษ
1398	ศึกษานิเทศ	ศึกษานิเทศก์
1399	ศึกษานิเทศน์	ศึกษานิเทศก์
1400	ศุข	สุข
1401	ศูนย์เสีย	สูญเสีย
1402	เศรษฐีนี	เศรษฐินี
1403	เศวตร	เศวต
1404	โศกนาฏกรรม	โศกนาฏกรรม
1405	โศกสันต์	โศกศัลย์
1406	โศกะนาฏกรรม	โศกนาฏกรรม
1407	โศรกเศร้า	โศกเศร้า
1408	สกด	สะกด
1409	สกิด	สะกิด
1410	สเก็ต	สเกต
1411	สคราญ	สะคราญ
1412	สงกรานตร์	สงกรานต์
1413	สงกาน	สงกรานต์
1414	สงกานต์	สงกรานต์
1415	สงวนราคา	สนนราคา
1416	สดวก	สะดวก
1417	สติสัมป๊ญชัญญะ	สติสัมปชัญญะ
1418	สเต๊ะ	สะเต๊ะ
1419	สแตมป์	แสตมป์
1420	สถาณการณ์	สถานการณ์
1421	สถานการ	สถานการณ์
1422	สถานะการณ์	สถานการณ์
1423	สถิด	สถิต
1424	สถิตย์	สถิต
1425	สถุน	สถุล
1426	สนุกเกอ	สนุกเกอร์
1427	สนุ้กเกอร์	สนุกเกอร์
1428	สนุ๊กเกอร์	สนุกเกอร์
1429	สเน่ห์	เสน่ห์
1430	สบัด	สะบัด
1431	สเบียง	เสบียง
1432	สปาเก็ตตี้	สปาเกตตี
1433	สเปญ	สเปน
1434	สพาน	สะพาน
1435	สภาวการณ์	สภาพการณ์
1436	สภาวะการณ์	สภาพการณ์
1437	สมดุลย์	สมดุล
1438	สมบูน	สมบูรณ์
1439	สมเพท	สมเพช
1440	สมเพศ	สมเพช
1441	สมเพส	สมเพช
1442	สมมุดติฐาน	สมมติฐาน
1443	สมมุดติถาน	สมมติฐาน
1444	สมมุติ	สมมติ
1445	สมุหบัญชี	สมุห์บัญชี
1446	สรรค์แสร้ง	สรรแสร้ง
1447	สรรค์หา	สรรหา
1448	สรรพยอก	สัพยอก
1449	สรรเพชร	สรรเพชญ
1450	สรัด	สัตว์
1451	สรัส	สัตว์
1452	สร้างสรร	สร้างสรรค์
1453	สลิ่ม	ซ่าหริ่ม
1454	สเลด	เสลด
1455	สวงสวรรค์	สรวงสวรรค์
1456	สวรรค์คต	สวรรคต
1457	สวรรณคต	สวรรคต
1458	สสิธร	ศศิธร
1459	ส้องเสพ	ซ่องเสพ
1460	สอบเชาว์	สอบเชาวน์
1461	สอาด	สะอาด
1462	สอิ้ง	สะอิ้ง
1463	สะกัด	สกัด
1464	สะกาว	สกาว
1465	สะดับ	สดับ
1466	สะบง	สบง
1467	สะบงจีวร	สบงจีวร
1468	สะบาย	สบาย
1469	สะบู่	สบู่
1470	สะเบียง	เสบียง
1471	สะไบ	สไบ
1472	สะพึงกลัว	สะพรึงกลัว
1473	สะเหร่อ	สะเหล่อ
1474	สะเหลด	เสลด
1475	สักกะหลาด	สักหลาด
1476	สักการะบูชา	สักการบูชา
1477	สังเกด	สังเกต
1478	สังเกต	สังเกต
1479	สังเกตุ	สังเกต
1480	สังขทาน	สังฆทาน
1481	สังเขบ	สังเขป
1482	สังคทาน	สังฆทาน
1483	สังคยนา	สังคายนา
1484	สังคยานา	สังคายนา
1485	สังคายนาย	สังคายนา
1486	สังวรณ์	สังวร
1487	สังวาลย์	สังวาล
1488	สังเวท	สังเวช
1489	สังสรร	สังสรรค์

No.	Incorrect	Correct
1490	สัญชาติญาณ	สัญชาตญาณ
1491	สัญญลักษณ์	สัญลักษณ์
1492	สัญญาน	สัญญาณ
1493	สัญโดษ	สันโดษ
1494	สัญลักษ์	สัญลักษณ์
1495	สัณฐาณ	สัณฐาน
1496	สัต	สัตว์
1497	สัตรู	ศัตรู
1498	สันฐาณ	สัณฐาน
1499	สันฐาน	สัณฐาน
1500	สันโดด	สันโดษ
1501	สันทนาการ	นันทนาการ
1502	สันนิฐาณ	สันนิษฐาน
1503	สันนิฐาน	สันนิษฐาน
1504	สันนิษฐาณ	สันนิษฐาน
1505	สับปรับ	สับปลับ
1506	สับปะรส	สับปะรด
1507	สับปะหงก	สัปหงก
1508	สับเพร่า	สะเพร่า
1509	สัปปะรด	สับปะรด
1510	สัปปะเหร่อ	สัปเหร่อ
1511	สัปรด	สับปะรด
1512	สัปลับ	สับปลับ
1513	สัพเหร่อ	สัปเหร่อ
1514	สัมนา	สัมมนา
1515	สัมภาษ	สัมภาษณ์
1516	สัมภาสน์	สัมภาษณ์
1517	สัส	สัตว์
1518	สากกระเบือ	สากกะเบือ
1519	สาด	สัตว์
1520	สาธร	สาทร
1521	สาธารณะชน	สาธารณชน
1522	สาธารณะประโยชน์	สาธารณประโยชน์
1523	สาธารณะสถาน	สาธารณสถาน
1524	สาธารณะสมบัติ	สาธารณสมบัติ
1525	สาธารณะสุข	สาธารณสุข
1526	สานสิญ	สายสิญจน์
1527	สาบแช่ง	สาปแช่ง
1528	สาบสรร	สาปสรร
1529	สาบาญ	สาบาน
1530	สาปศูนย์	สาบสูญ
1531	สาปสูญ	สาบสูญ
1532	สาปสูน	สาบสูญ
1533	สามเศร้า	สามเส้า
1534	สายสิญจ์	สายสิญจน์
1535	สายสิน	สายสิญจน์
1536	สารทุกข์สุขดิบ	สารทุกข์สุกดิบ
1537	สารสำคัญ	สาระสำคัญ
1538	สาระทุกข์สุขดิบ	สารทุกข์สุกดิบ
1539	สาระประโยชน์	สารประโยชน์
1540	สาระพี, สารพี	สารภี
1541	สารัตถสำคัญ	สารัตถะสำคัญ
1542	สารัตถะประโยชน์	สารัตถประโยชน์
1543	สาส	สัตว์
1544	สำภาษณ์	สัมภาษณ์
1545	สำมะนา	สัมมนา
1546	สำเหร็จ	สำเร็จ
1547	สำอางค์	สำอาง
1548	สิงห์โต	สิงโต
1549	สิทธ์	สิทธิ
1550	สิทธ	สิทธิ
1551	สิทธิ์	สิทธิ
1552	สินเทา	สินเธาว์
1553	สิโรราบ	ศิโรราบ
1554	สิวิไล	ศิวิไลซ์
1555	สีสรร	สีสัน
1556	สีสรรค์	สีสัน
1557	สีสวาท	สีสวาด
1558	สีสวาส	สีสวาด
1559	สุกียากี	สุกียากี้
1560	สุกี้ยากี้	สุกียากี้
1561	สุขคติ	สุคติ
1562	สุขติ	สุคติ
1563	สุพรรณบัตร	สุพรรณบัฏ
1564	สุ่ม	ซุ่ม
1565	สุ้ม	ซุ้ม
1566	สุหนัด	สุหนัต
1567	สูจิบัตร	สูติบัตร
1568	สูญญากาศ	สุญญากาศ
1569	สูติบัตร	สูจิบัตร
1570	เสกสรรค์	เสกสรร
1571	เสปน	สเปน
1572	เสพย์	เสพ
1573	เสร่อ	สะเหล่อ
1574	เสริฟ์	เสิร์ฟ
1575	เสริฟ	เสิร์ฟ
1576	เสล่อ	สะเหล่อ
1577	เสียสูญ	เสียศูนย์

No.	Incorrect	Correct
1578	เสื้อกราวน์	เสื้อกาวน์
1579	เสือกใส	เสือกไส
1580	เสื้อกาว	เสื้อกาวน์
1581	เสื้อกาวด์	เสื้อกาวน์
1582	เสื้อเชิ๊ต	เสื้อเชิ้ต
1583	แสกน	สแกน
1584	โสกศัลย์	โศกศัลย์
1585	โสกเศร้า	โศกเศร้า
1586	โสรส	โสพส
1587	ใส้	ไส้
1588	ไสบ	สไบ
1589	ไสล	ไศล
1590	หกคเมน	หกคะเมน
1591	หกคะเมร	หกคะเมน
1592	หงษ์	หงส์
1593	หนาแน่น	แน่นหนา
1594	หน้าปัทม์ นาฬิกา	หน้าปัด นาฬิกา
1595	หม้อห้อม	ม่อฮ่อม
1596	หม้อห้อม	ม่อฮ่อม
1597	หม้อฮ่อม	ม่อฮ่อม
1598	หม้อฮ่อม	ม่อฮ่อม
1599	หมาไน	หมาใน
1600	หมาหมุ้ย	หมามุ่ย
1601	หมี่ซั่ว	มี่สั่ว
1602	หมูหยอง	หมูหย็อง
1603	หยักใย่	หยักไย่
1604	หยักโศก	หยักศก
1605	หยั่มมา	อย่ามา
1606	หยากใย่	หยักไย่
1607	หยากไย่	หยักไย่
1608	หยากไย่	หยักไย่
1609	หยิง	หญิง
1610	หยิบย่ง	หยิบหย่ง
1611	หยิบโย่ง	หยิบหย่ง
1612	หรรม	ขำ
1613	หลงไหล	หลงใหล
1614	หลับไหล	หลับใหล
1615	หลุดรุ่ย	หลุดลุ่ย
1616	หลุมพลาง	หลุมพราง
1617	ห่วงไย	ห่วงใย
1618	หอมหวล	หอมหวน
1619	หอยจ๊อ	ฮ่อยจ๊อ
1620	ห้อยจ้อ	ฮ่อยจ๊อ
1621	ห๊ะ	ฮะ
1622	หินชะนวน	หินชนวน
1623	หิรัญบัตร	หิรัญบัฏ
1624	หุ่นยนตร์	หุ่นยนต์
1625	เหงื่อกาล	เหงื่อกาฬ
1626	เหตุกาน	เหตุการณ์
1627	เหตุการ	เหตุการณ์
1628	เห้น	เห็น
1629	เหนือบ่ากว่าแรง	เหลือบ่ากว่าแรง
1630	เหม็นสาป	เหม็นสาบ
1631	เหรอหรา	เหลอหลา
1632	เหล็กไน	เหล็กใน
1633	เหลวไหล	เหลวไหล
1634	เหิร	เหิน
1635	แหลกราญ	แหลกลาญ
1636	แหลกราน	แหลกลาญ
1637	โหยหวล	โหยหวน
1638	โหรพา	โหระพา
1639	โหระภา	โหระพา
1640	ใหย่	ใหญ่
1641	ใหหลำ	ไหหลำ
1642	ไหร	ไร
1643	ไหลตาย	ใหลตาย
1644	องคชาติ	องคชาต
1645	องคุลิมาร	องคุลิมาล
1646	องคุลีมาร	องคุลิมาล
1647	องคุลีมาล	องคุลิมาล
1648	อณุโลม	อนุโลม
1649	อณุสาวรีย์	อนุสาวรีย์
1650	อดรีนาลิน	อะดรีนาลิน
1651	อธิฏฐาณ	อธิษฐาน
1652	อธิฏฐาน	อธิษฐาน
1653	อธิฐาณ	อธิษฐาน
1654	อธิฐาน	อธิษฐาน
1655	อนาจ	อนาถ
1656	อนาธร	อนาทร
1657	อนุกาชาติ	อนุกาชาด
1658	อนุญาติ	อนุญาต
1659	อนุมัด	อนุมัติ
1660	อนุมัต	อนุมัติ
1661	อนุมาณ	อนุมาน
1662	อนุรักษ์นิยม	อนุรักษนิยม
1663	อนุเสาวรีย์	อนุสาวรีย์
1664	อเนจอนาจ	อเนจอนาถ
1665	อเนถอนาถ	อเนจอนาถ
1666	อโนดาด	อโนดาต
1667	อบอวน	อบอวล
1668	อภิชาติ	อภิชาต
1669	อภินิหารย์	อภินิหาร
1670	อภิรมณ์	อภิรมย์
1671	อภิเสก	อภิเษก
1672	อมต	อมตะ
1673	อมฤต	อำมฤต
1674	อย่าร้าง	หย่าร้าง

No.	Incorrect	Correct
1675	อยุ	อยู่
1676	อยุ่	อยู่
1677	อรทัย	อรไท
1678	อรรถภาพ	อัตภาพ
1679	อริยเมตตรัย	อริยเมตไตรย
1680	อริยเมตไตร	อริยเมตไตรย
1681	อริยสัจจ์	อริยสัจ
1682	อริยาบท	อิริยาบถ
1683	อลังการ์	อลังการ
1684	อลังการณ์	อลังการ
1685	อลุ่มอล่วย	อะลุ่มอล่วย
1686	อลูมิเนียม	อะลูมิเนียม
1687	อวสาณ	อวสาน
1688	อวสานต์	อวสาน
1689	อสงขัย	อสงไขย
1690	อสงไข	อสงไขย
1691	อสุนีบาท	อสุนีบาต
1692	อ๊อกซิเจน	ออกซิเจน
1693	ออกห่าง	ออกหาก
1694	ออฟฟิซ	ออฟฟิศ
1695	อ็อฟฟิซ	ออฟฟิศ
1696	ออฟฟิต	ออฟฟิศ
1697	อ็อฟฟิต	ออฟฟิศ
1698	ออฟฟิส	ออฟฟิศ
1699	อ็อฟฟิส	ออฟฟิศ
1700	อะนุมัต	อนุมัติ
1701	อะมีบ้า	อะมีบา
1702	อะลุ้มอล่วย	อะลุ้มอล่วย
1703	อะลุ่มอะล่วย	อะลุ้มอล่วย
1704	อะลุ้มอะล่วย	อะลุ้มอล่วย
1705	อะหลักอะเหลื่อ	อลักเอลื่อ
1706	อะหลั่ย	อะไหล่

No.	Incorrect	Correct
1707	อะหิวาตกโรค	อหิวาตกโรค
1708	อักขระวิธี	อักขรวิธี
1709	อักศร	อักษร
1710	อักสร	อักษร
1711	อัญชัญ	อัญชัน
1712	อัญญมณี	อัญมณี
1713	อัญมนี	อัญมณี
1714	อัตคัต	อัตคัด
1715	อัตตะนัย	อัตนัย
1716	อัทยาศัย	อัธยาศัย
1717	อัธจันทร์	อัฒจันทร์
1718	อัธยาษัย	อัธยาศัย
1719	อันตพาล	อันธพาล
1720	อันไนย	อัตนัย
1721	อับปรีย์	อัปรีย์
1722	อับปี	อัปรีย์
1723	อัปเดท	อัปเดต
1724	อัปปาง	อับปาง
1725	อัปปางค์	อับปาง
1726	อัปปี	อัปรีย์
1727	อัพเดต	อัปเดต
1728	อัพเดท	อัปเดต
1729	อัพโหลด	อัปโหลด
1730	อัมตะ	อมตะ
1731	อัมพาส	อัมพาต
1732	อัมรินทร์	อมรินทร์
1733	อัมฤต	อำมฤต
1734	อัมหิต	อำมหิต
1735	อัลบัม	อัลบั้ม
1736	อัลฟาทอกซิน	อะฟลาทอกซิน
1737	อัลไร	อะไร

No.	Incorrect	Correct
1738	อัศจันทร์	อัศจรรย์
1739	อัศดง	อัสดง
1740	อากับกิริยา	อากัปกิริยา
1741	อากาส	อากาศ
1742	อาคเณย์	อาคเนย์
1743	อาฆาตมาตร้าย	อาฆาตมาดร้าย
1744	อาจินต์	อาจิณ
1745	อาเจียร	อาเจียน
1746	อาชาไน	อาชาไนย
1747	อาชีวะศึกษา	อาชีวศึกษา
1748	อานิสงค์	อานิสงส์
1749	อานิสงฆ์	อานิสงส์
1750	อาเพท	อาเพศ
1751	อาเพส	อาเพศ
1752	อาเภส	อาเพศ
1753	อายัต	อายัด
1754	อายัติ	อายัด
1755	อารมย์	อารมณ์
1756	อารักษ์ขา	อารักขา
1757	อารัย	อะไร
1758	อารีย์	อารี
1759	อาวร	อาวรณ์
1760	อาวุท	อาวุธ
1761	อาสนสงฆ์	อาสน์สงฆ์
1762	อาสสงฆ์	อาสน์สงฆ์
1763	อาหลั่ย	อะไหล่
1764	อำนาจบาทใหญ่	อำนาจบาตรใหญ่
1765	อำพร	อัมพร
1766	อำมตะ	อมตะ
1767	อำมะหิต	อำมหิต
1768	อินเดียแดง	อินเดียนแดง

No.	Incorrect	Correct
1769	อินทผาลัม	อินทผลัม
1770	อินทร์ธนู	อินทรธนู
1771	อินทรีย์	อินทรี
1772	อินทรีย์วัตถุ	อินทรียวัตถุ
1773	อินทรีวัตถุ	อินทรียวัตถุ
1774	อินธนู	อินทรธนู
1775	อินฟาร์เรด	อินฟราเรด
1776	อินฟาเรด	อินฟราเรด
1777	อิเลคตรอน	อิเล็กตรอน
1778	อิสระภาพ	อิสรภาพ
1779	อิสระเสรี	อิสรเสรี
1780	อิสาณ	อีสาน
1781	อิสาน	อีสาน
1782	อีเมล์	อีเมล
1783	อีเล็คโทรนิคส์	อิเล็กทรอนิกส์
1784	อุกกาบาท	อุกกาบาต
1785	อุกฤติ	อุกฤษฏ์
1786	อุกาบาต	อุกกาบาต
1787	อุดมการ	อุดมการณ์
1788	อุดส่า	อุตส่าห์
1789	อุตสาห์	อุตส่าห์
1790	อุตสาหะพยายาม	อุตสาหพยายาม
1791	อุทาหร	อุทาหรณ์
1792	อุธรณ์	อุทธรณ์
1793	อุธาหรณ์	อุทาหรณ์
1794	อุบาท	อุบาทว์
1795	อุบาทย์	อุบาทว์
1796	อุปการะคุณ	อุปการคุณ
1797	อุปถัมธ์	อุปถัมภ์
1798	อุปถัมน์	อุปถัมภ์
1799	อุปทาน	อุปาทาน

No.	Incorrect	Correct
1800	อุปมาอุปมัย	อุปมาอุปไมย
1801	อุปโลก	อุปโลกน์
1802	อุปัชชา	อุปัชฌาย์
1803	อุปัชฌา	อุปัชฌาย์
1804	อุโมง	อุโมงค์
1805	อุลตร้าไวโอเล็ต	อัลตราไวโอเลต
1806	อุศุภราช	อุสุภราช
1807	เอกขเนก	เอกเขนก
1808	เอกฉัน	เอกฉันท์
1809	เอ็กซเรย์	เอกซเรย์
1810	เอกะฉันท์	เอกฉันท์
1811	เอเชียน	อาเซียน
1812	เอเซีย	เอเชีย
1813	เอนก	อเนก
1814	เอ็นโดรฟีน	เอ็นดอร์ฟิน
1815	เอาใจออกห่าง	เอาใจออกหาก
1816	แอลกอฮอร์	แอลกอฮอล์
1817	โอกาด	โอกาส
1818	โอกาศ	โอกาส
1819	โอป่ะ	โอเครึเปล่า
1820	ไอติม	ไอศกรีม
1821	ไอศครีม	ไอศกรีม
1822	ไอสครีม	ไอศกรีม
1823	ไอสัส	ไอสัตว์
1824	ไอสัสส	ไอสัตว์
1825	ไอเห้	ไอเหี้ย
1826	ไอเฮีย	ไอเหี้ย
1827	ไอเฮี่ย	ไอเหี้ย
1828	ฮิจเราะจ์	ฮิจเราะห์
1829	ฮิบโปโปเตมัส	ฮิปโปโปเตมัส

No.	Incorrect	Correct
1830	ฮีโมโกลบิน	เฮโมโกลบิน
1831	ฮีเลี่ยม	ฮีเลียม
1832	เฮอร์ริเคน	เฮอริเคน
1833	ไฮดร้า	ไฮดรา
1834	ฯพณ	ฯพณฯ

Appendix C

List of stop words retrieve from https://github.com/stopwords-iso/stopwords-th/blob/master/stopwords-th.txt#L115

Stop Word List

No.	Word	No.	Word	No.	Word	No.	Word
1	กล่าว	28	ดัง	55	นี้	82	ลง
2	ก็	29	ด้าน	56	เนื่องจาก	83	เลย
3	กว่า	30	เดียว	57	ใน	84	แล้ว
4	ก่อน	31	เดียวกัน	58	บาง	85	และ
5	กัน	32	โดย	59	แบบ	86	วัน
6	กับ	33	ได้	60	เป็น	87	ว่า
7	การ	34	ต่อ	61	เป็นการ	88	ไว้
8	ขณะ	35	ต้อง	62	เปิด	89	ส่ง
9	ขอ	36	ตั้ง	63	เปิดเผย	90	ส่วน
10	ของ	37	ตั้งแต่	64	ไป	91	สำหรับ
11	ขึ้น	38	ต่าง	65	ผล	92	สุด
12	เขา	39	ตาม	66	พบ	93	หรือ
13	เข้า	40	แต่	67	เพราะ	94	หลัง
14	คง	41	ถ้า	68	เพื่อ	95	หลังจาก
15	ครั้ง	42	ถึง	69	มา	96	หลาย
16	ความ	43	ทั้ง	70	มาก	97	หาก
17	คือ	44	ทั้งนี้	71	มี	98	เห็น
18	เคย	45	ทำ	72	เมื่อ	99	แห่ง
19	จะ	46	ทาง	73	ยัง	100	ให้
20	จัด	47	ทำให้	74	รวม	101	อยาก
21	จาก	48	ที่	75	ร่วม	102	อย่าง
22	จึง	49	ทุก	76	ระหว่าง	103	อยู่
23	เฉพาะ	50	นอกจาก	77	รับ	104	ออก
24	ช่วง	51	นัก	78	รายๆ	105	อะไร
25	เช่น	52	นั้น	79	เรา	106	อาจ
26	ซึ่ง	53	น่า	80	เริ่ม	107	อีก
27	ด้วย	54	นำ	81	แรก	108	เอง

Appendix D

User Comment Sentiment Dictionary

The user comment sentiment dictionary that constructed by SO-PMI-IR algorithm with "NEAR" operation from 8,34,3126 user comments (Approximately110,000,000 words).

SV: Sentiment value

No.	Word	SV	No.	Word	SV	No.	Word	SV
1	555	1.3069	29	กระจอก	-0.0295	57	ก้อ	0.3764
2	ก็	0.3368	30	กระดาษ	-0.0191	58	กอง	0.4135
3	ก้	0.3329	31	กระทรวง	0.1839	59	กองทัพ	0.1951
4	ก็คง	0.0449	32	กระทำ	-0.5780	60	กอด	-0.0095
5	ก็ช่าง	0.5116	33	กระทืบ	0.0714	61	ก่อน	0.3869
6	กฎ	0.6092	34	กระเป๋า	0.7684	62	ก้อน	-0.0062
7	กฎหมาย	0.1200	35	กระแส	0.7919	63	ก่อนที่จะ	0.1537
8	กฎหมาย	0.2324	36	กร่าง	0.3285	64	ก้อย	0.3938
9	กด	0.3017	37	กราบ	0.4401	65	กะ	0.2104
10	กดดัน	0.4768	38	กรุณา	0.3300	66	กะเทย	0.0119
11	ก็ดี	0.8643	39	กล่อง	0.0822	67	กะลา	0.6006
12	ก็ได้	0.5313	40	กล้อง	0.4333	68	กัด	0.5341
13	ก็ต้อง	0.1797	41	กลับ	0.1608	69	กัน	0.3907
14	กตัญญู	1.4790	42	กลับบ้าน	0.6146	70	กันเอง	0.2296
15	ก็ตาม	0.2045	43	กลับมา	0.8005	71	กับ	-0.3949
16	กติกา	0.0087	44	กลัว	0.5117	72	กัว	-0.0665
17	กทม	0.6427	45	กล้า	0.2669	73	กา	0.6903
18	กบฏ	0.2797	46	กลาง	0.6205	74	กาก	-0.1649
19	ก็เพราะ	0.3520	47	กลายเป็น	0.2049	75	กาย	0.6412
20	ก็มี	0.4232	48	กล่าว	0.2193	76	การ	0.2095
21	ก็ไม่ได้	0.4637	49	กล่าวหา	0.4172	77	การกระทำ	-0.2618
22	ก็ยัง	0.6606	50	กลุ่ม	0.4737	78	การโกง	0.4290
23	กร	0.2110	51	ก็แล้วกัน	0.4193	79	การใช้	0.5697
24	กรณี	-0.3007	52	กวน	0.1841	80	การ์ตูน	-0.4994
25	กรรม	-0.5370	53	กว่า	0.6197	81	การทำงาน	0.7320
26	กรรมการ	0.3454	54	กห	0.4188	82	การที่	0.2446
27	กรวย	-0.3822	55	กอ	0.3772	83	การปกครอง	0.3993
28	กระ	0.2445	56	ก่อ	-0.0664	84	การเมือง	0.5304

No.	Word	SV	No.	Word	SV	No.	Word	SV
85	การเลือกตั้ง	0.4505	113	เก่า	0.5239	141	โกง	0.5700
86	การศึกษา	0.2320	114	เกาหลี	0.6389	142	โกงกิน	0.0425
87	ก้าว	0.2715	115	เกาหลีเหนือ	0.7541	143	โกรธ	0.2316
88	กำ	0.0966	116	เกาะ	0.6378	144	โกหก	0.1088
89	กำจัด	0.6917	117	เกิด	-0.2963	145	ใกล้	0.3721
90	กำนัน	0.8862	118	เกิดขึ้น	-0.6222	146	ไก่	0.7730
91	กำไร	0.4595	119	เกิดเหตุ	-0.6273	147	ไกล	0.6423
92	กำลัง	0.2597	120	เกิน	-0.0513	148	ไกล่เกลี่ย	-0.7224
93	กำลังจะ	0.3170	121	เกินไป	-0.0761	149	ไก่อู	0.0174
94	กำลังใจ	0.3057	122	เกี่ยง	-0.4581	150	ขน	0.5878
95	กำหนด	0.3743	123	เกียรติ	0.2616	151	ขนาด	-0.0241
96	กิจ	0.8858	124	เกี่ยว	0.1823	152	ขบวนการ	-0.0670
97	กิน	0.7545	125	เกี่ยวกับ	0.3195	153	ข่มขืน	0.1684
98	กินกัน	0.2301	126	เกี่ยวข้อง	0.0243	154	ขโมย	0.2481
99	กินข้าว	0.5566	127	เกือบ	0.4241	155	ขยะ	0.6804
100	กิเลส	0.2690	128	แก	0.3740	156	ขยะสังคม	-0.1791
101	กี	0.3961	129	แก่	0.1533	157	ขยัน	0.9774
102	กี่	0.2580	130	แก้	0.6272	158	ขวด	0.0297
103	กีฬา	0.7968	131	แก้ไข	0.3461	159	ขวบ	-0.1959
104	กู	0.3739	132	แกง	-0.1411	160	ขวา	0.2187
105	กู้	0.5509	133	แก๊ง	-0.1652	161	ขวาง	-0.0813
106	เก	0.5825	134	แก้ตัว	0.0567	162	ขอ	-0.3747
107	เก่ง	0.9000	135	แกนนำ	0.4112	163	ข้อ	0.3309
108	เก็บ	0.6656	136	แก้ปัญหา	0.9078	164	ข้อความ	0.1530
109	เกม	0.5218	137	แก้ผ้า	-0.0034	165	ของ	0.2175
110	เกรงใจ	-0.0048	138	แก้ม	-0.6806	166	ของจริง	0.8822
111	เกลียด	0.6714	139	แกล้ง	-0.2931	167	ของสงฆ์	0.2517
112	เกษียณ	0.7707	140	โก	0.5803	168	ขอทาน	0.1160

No.	Word	SV	No.	Word	SV	No.	Word	SV
169	ขอโทษ	-0.0307	196	ขาย	0.8865	224	เข้าใจผิด	0.0885
170	ขอบ	0.8419	197	ขายของ	0.1534	225	เข้าไป	0.3508
171	ขอบคุณ	1.1632	198	ขายชาติ	0.2958	226	เข้ามา	0.6501
172	ข้อมูล	0.5727	199	ข้าราชการ	0.7793	227	เขียน	0.5674
173	ขอร้อง	0.2682	200	ขาว	1.5367	228	แข็ง	0.0214
174	ขอแสดงความยินดี	0.9395	201	ข่าว	-0.0806	229	แข่ง	0.8649
			202	ข้าว	0.7529	230	แข็งแรง	1.1115
175	ข้อหา	0.1558	203	ข่าวดี	0.6791	231	ไข่	0.2485
176	ขอให้	-0.1500	204	ข่าวสด	0.5670	232	คค	-0.4762
177	ข้ออ้าง	-0.1301	205	ขำ	0.7071	233	คง	0.1454
178	ขัง	0.2291	206	ขี่	0.2455	234	คงจะ	0.3152
179	ขัด	0.3679	207	ขี้	0.5018	235	คณะ	0.6691
180	ขัดขวาง	0.6453	208	ขี้เกียจ	-0.0521	236	คณะสงฆ์	0.8015
181	ขัดแย้ง	0.5709	209	ขี้โกง	0.3794	237	คดี	0.3170
182	ขั้น	0.1933	210	ขี้ข้า	0.2535	238	คติ	-0.6073
183	ขั้นตอน	0.1376	211	ขึ้น	0.3794	239	คน	0.0951
184	ขับ	-0.0134	212	ขึ้นไป	0.6484	240	ค้น	0.1786
185	ขับรถ	-0.1851	213	ขืน	0.1801	241	คนเก่ง	1.0680
186	ขา	0.7139	214	ขุด	0.6487	242	คนแก่	0.3602
187	ข้า	0.4967	215	ขู่	0.0740	243	คนโกง	0.4303
188	ข้าง	0.5547	216	เข็ด	0.1456	244	คนขับ	-0.5667
189	ข้างใน	0.6185	217	เขต	0.5386	245	คนโง่	-0.1706
190	ข้างหน้า	0.4274	218	เข้มแข็ง	-0.5910	246	คนจน	0.7046
191	ข้างหลัง	-0.5542	219	เขมร	0.4250	247	คนชั่ว	-0.2401
192	ขาด	0.0801	220	เขา	0.2407	248	คนดี	0.8234
193	ขาดทุน	0.3941	221	เข้า	0.3638	249	คนดู	0.7131
194	ขาดสติ	-0.5055	222	เข้าข้าง	0.2018	250	คนเดียว	0.3466
195	ข้าม	0.0772	223	เข้าใจ	-0.0970	251	คนตาย	-0.4338

No.	Word	SV
252	คนใต้	0.6866
253	คนไทย	0.7180
254	คนธรรมดา	0.5044
255	คนใน	0.0407
256	คนบ้า	0.1397
257	คนรวย	0.6082
258	คนรัก	0.2189
259	คนร้าย	-0.7083
260	คนเรา	0.1923
261	คนละ	0.5269
262	คนเลว	-0.2937
263	คนสวย	1.1100
264	คนอื่น	0.3011
265	คบ	0.6274
266	คม	0.4131
267	ครบ	0.5675
268	ครอง	0.3133
269	ครองเมือง	-0.0246
270	ครอบ	-0.6219
271	ครอบครัว	-0.9682
272	ครั้ง	0.1783
273	ครั้งที่	0.1321
274	ครั้งนี้	-0.5763
275	ครับ	0.4204
276	ครับผม	0.6137
277	คราบ	0.4442
278	คร้าบ	0.6500
279	คราวนี้	0.3131
280	ครึ่ง	0.8488
281	ครู	0.3145
282	คลอง	0.3621
283	คลอด	-0.5976
284	คล้าย	0.4946
285	คลิป	0.1238
286	ควบคุม	0.4860
287	ควย	0.1954
288	ควร	0.4314
289	ควรจะ	0.4242
290	ความ	-0.1477
291	ความคิด	0.5816
292	ความคิดเห็น	0.3675
293	ความจริง	0.4398
294	ความเจริญ	0.8401
295	ความเชื่อ	0.1681
296	ความดี	0.8118
297	ความเดือดร้อน	0.3522
298	ความแตกแยก	0.1111
299	ความถูกต้อง	0.7462
300	ความปลอดภัย	0.1679
301	ความเป็นธรรม	0.2980
302	ความผิด	0.0510
303	ความยุติธรรม	0.2334
304	ความรัก	0.3122
305	ความรับผิดชอบ	0.2810
306	ความรุนแรง	0.2379
307	ความรู้	0.5666
308	ความรู้สึก	-0.4356
309	ความเร็ว	0.0598
310	ความวุ่นวาย	0.1583
311	ความสงบ	0.4387
312	ความสามารถ	1.1401
313	ความสุข	0.9184
314	ความเสียใจ	-0.9128
315	ความหมาย	0.4004
316	ความเห็น	0.5931
317	ควาย	0.3203
318	คว่ำ	0.4537
319	คอ	0.0902
320	คอม	0.4020
321	คอย	0.4359
322	ค่อย	0.4345
323	คอรัปชั่น	0.6766
324	คะ	0.3225
325	ค้ะ	0.0197
326	คะแนน	0.2911
327	คัดค้าน	0.2497
328	คัน	0.2272
329	คา	0.2966
330	ค่า	0.5777
331	ค้า	0.7822

No.	Word	SV	No.	Word	SV	No.	Word	SV
332	ค้าขาย	0.4328	360	คุณหมอ	1.1138	388	ฆาตกร	-0.2307
333	คางคก	-0.0494	361	คุม	0.6831	389	ฆ่าตัวตาย	-0.3573
334	ค่าใช้จ่าย	0.3805	362	คุ้ม	0.4980	390	โฆษณา	0.5481
335	ค้าน	0.5770	363	คุ้มครอง	0.3121	391	งก	-0.0066
336	ค่าย	0.5730	364	คุย	0.7312	392	งง	0.3796
337	ค่าแรง	0.1686	365	คู่	0.8859	393	งด	0.2695
338	คำ	-0.0095	366	คู่กรณี	-0.4060	394	งบ	0.3919
339	คำตอบ	0.1574	367	คูณ	0.3303	395	งบประมาณ	0.2833
340	คำถาม	0.3772	368	เคย	0.2737	396	งมงาย	-0.0939
341	คำพูด	0.1933	369	เครียด	-0.0560	397	งวด	1.2127
342	คำสั่ง	-0.0735	370	เครื่อง	0.0720	398	งอ	0.3142
343	คิด	0.4577	371	เครื่องบิน	0.1344	399	ง้อ	0.6984
344	คิดดู	0.2626	372	เครื่องแบบ	0.3109	400	ง้ัน	0.5479
345	คิดได้	0.4594	373	เครื่องมือ	0.1043	401	งัย	0.3577
346	คิดถึง	0.2833	374	เค้า	0.1203	402	งาน	0.7680
347	คิดมาก	0.5296	375	เคารพ	0.5645	403	งาม	1.3916
348	คิว	0.3277	376	แค่	0.1100	404	ง่าย	0.6391
349	คืน	0.4313	377	แค้น	-0.4239	405	งี้	0.7473
350	คือ	0.3530	378	แค่นั้น	0.5281	406	งู	0.5960
351	คุ	0.5439	379	แค่นี้	0.1935	407	เงา	-0.2377
352	คุก	0.3105	380	แคร์	0.9798	408	เงิน	0.6750
353	คุณ	0.5915	381	โค	0.4709	409	เงินก้อน	0.3919
354	คุณธรรม	0.7728	382	โคช	0.6298	410	เงินเดือน	0.5835
355	คุณพระ	0.2662	383	โค้ช	1.0456	411	เงินทอง	0.5916
356	คุณพ่อ	-0.1156	384	โคตร	0.1801	412	เงียบ	0.7817
357	คุณภาพ	0.7816	385	โครงการ	0.7688	413	โง่	0.5859
358	คุณแม่	-0.0969	386	ใคร	0.3429	414	ไง	0.2674
359	คุณยาย	0.5491	387	ฆ่า	-0.1286	415	จง	0.0314

No.	Word	SV	No.	Word	SV	No.	Word	SV
416	จน	0.3180	444	จังเลย	0.1861	472	จีน	1.1061
417	จนกว่า	0.0504	445	จังหวัด	0.7309	473	จึง	0.2145
418	จนได้	0.3385	446	จัญไร	-0.0869	474	จุ	0.7479
419	จนถึง	0.1721	447	จัด	0.9231	475	จุด	0.3730
420	จบ	0.4496	448	จัดการ	0.6051	476	เจ	0.5860
421	จม	0.0973	449	จัดระเบียบ	0.4865	477	เจ้	0.5731
422	จรรยาบรรณ	0.0624	450	จับ	-0.1236	478	เจ๊	0.3736
423	จรวด	-0.2318	451	จับได้	-0.4853	479	เจ๊ง	0.2548
424	จร้า	1.2391	452	จา	0.4124	480	เจ๋ง	0.9977
425	จริง	0.0982	453	จ่า	0.4046	481	เจตนา	0.3728
426	จริงจัง	0.2760	454	จ้า	1.2645	482	เจน	0.4557
427	จริยธรรม	0.1958	455	จาก	0.1626	483	เจ็บ	-0.4844
428	จอ	0.4028	456	จากไป	-0.6479	484	เจ็บปวด	-0.7143
429	จอง	0.4449	457	จ้าง	0.6657	485	เจรจา	-0.0924
430	จ้อง	-0.0624	458	จาน	0.0839	486	เจริญ	0.9725
431	จอด	0.1144	459	จ่าย	0.4716	487	เจอ	0.1543
432	จอม	0.3544	460	จาร	0.3631	488	เจอกัน	0.3355
433	จะ	0.2735	461	จำ	0.3582	489	เจ้า	0.3244
434	จ้ะ	0.7253	462	จำคุก	0.2755	490	เจ้าของ	0.2501
435	จ๊ะ	0.7160	463	จำได้	0.4616	491	เจ้านาย	0.4383
436	จะตาย	0.3382	464	จำนวน	0.3145	492	เจ้าพนักงาน	-0.2700
437	จะทำ	0.3329	465	จำนำ	0.4392	493	เจ้าหน้าที่	-0.1522
438	จะเป็น	0.3314	466	จำไว้	0.4314	494	แจก	0.9443
439	จะไป	0.5243	467	จิ	0.3361	495	แจง	-0.3331
440	จะมา	0.2663	468	จิต	0.1767	496	แจ้ง	0.3010
441	จัก	0.5793	469	จิตใจ	-0.0895	497	แจ้งความ	-0.2527
442	จัง	0.0863	470	จิตสำนึก	0.1701	498	โจร	0.3792
443	จังไร	-0.1257	471	จี้	0.7181	499	ใจ	0.1388

No.	Word	SV	No.	Word	SV	No.	Word	SV
500	ใจดำ	-0.8667	528	ชอบ	0.9000	556	ชาวโลก	0.3383
501	ใจดี	1.3107	529	ชะ	0.3773	557	ชาวสวน	0.5426
502	ใจเย็น	0.2672	530	ชัก	0.3417	558	ชิ	0.5530
503	ใจร้าย	-0.8877	531	ชั่ง	0.2814	559	ชิง	-0.2630
504	ไจ	0.1426	532	ชัด	0.2376	560	ชิน	0.3538
505	ฉบับ	0.4365	533	ชัดเจน	0.7578	561	ชินวัตร	0.1372
506	ฉลาด	1.3747	534	ชั้น	0.5213	562	ชิบหาย	0.5822
507	ฉัน	0.2367	535	ชัย	0.7152	563	ชี	0.1921
508	ฉาย	-0.0855	536	ชั่ว	0.0162	564	ชี้	0.2352
509	ฉิบหาย	0.6114	537	ชั่วโคตร	-0.7966	565	ชี้แจง	0.2158
510	ฉีก	0.2812	538	ชั่วโมง	0.1400	566	ชีวิต	-0.0050
511	เฉพาะ	0.3256	539	ชัวร์	0.4367	567	ชื่นชม	1.4386
512	เฉย	0.6648	540	ชา	0.6601	568	ชื่อ	0.8576
513	แฉ	0.2328	541	ช้า	0.1110	569	ชื้อ	0.6487
514	ชก	0.8091	542	ช่าง	0.1702	570	ชื่อเสียง	0.4431
515	ชง	0.0541	543	ช้าง	0.1240	571	ชุด	1.2773
516	ชงนม	-1.4140	544	ชาติ	0.5118	572	ชุมนุม	0.4640
517	ชดใช้	-0.8037	545	ชาติไทย	0.5171	573	ชู	0.5696
518	ชน	-0.4449	546	ชาติบ้านเมือง	0.6926	574	เช็ค	0.1192
519	ชนะ	1.2919	547	ชาติหน้า	-0.0046	575	เช่น	0.7034
520	ชม	1.1002	548	ชาติหมา	-0.6214	576	เช่นกัน	0.1224
521	ช่วง	0.3328	549	ชาย	0.4425	577	เช่นนี้	0.0924
522	ช่วงนี้	0.3464	550	ชายแดน	0.6109	578	เช้า	0.4396
523	ชวน	0.5101	551	ชาว	0.4840	579	เชิญ	0.9627
524	ช่วย	0.4976	552	ชาวไทย	0.4862	580	เชียงใหม่	0.6542
525	ช่วยกัน	0.4551	553	ชาวนา	0.4472	581	เชียร์	1.1230
526	ช่วยเหลือ	0.7122	554	ชาวบ้าน	0.6102	582	เชียว	0.6489
527	ช่อง	0.6165	555	ชาวพุทธ	0.0776	583	เชื่อ	0.3629

No.	Word	SV	No.	Word	SV	No.	Word	SV
584	เชื้อ	0.1791	612	ซี	0.4688	640	ดับ	-0.0208
585	แช่ง	-0.3804	613	ซี่	0.4187	641	ด่า	0.3358
586	แชมป์	0.8327	614	ซึ่ง	-0.0795	642	ด่า	-0.9000
587	แชร์	0.1416	615	ซึ้ง	0.2986	643	ด่าน	0.3617
588	โชคดี	1.0649	616	ซื้อ	0.8211	644	ด้าน	0.6707
589	โชว์	1.0556	617	ซื้อของ	0.4548	645	ด้าย	0.0609
590	ใช่	0.5375	618	เซ	0.8203	646	ดารา	0.7181
591	ใช้	0.6139	619	เซ็ง	0.1930	647	ดาว	0.5547
592	ใช้ชีวิต	0.2358	620	ญาติ	-1.2110	648	ดาวอังคาร	0.4903
593	ใช้ได้	0.4267	621	ญาติพี่น้อง	-0.8848	649	ด่าว่า	0.2400
594	ใช้เวลา	0.2991	622	ญี่ปุ่น	0.8245	650	ดำ	0.6114
595	ใช้หนี้	0.6319	623	ฐาน	0.1781	651	ดำเนินการ	0.4070
596	ใช่ไหม	0.5871	624	ฐานะ	0.4075	652	ดำเนินคดี	-0.1566
597	ใช้อำนาจ	0.2190	625	ณ	0.1952	653	ดิ	0.6482
598	ไช่	0.7026	626	เณร	-0.1868	654	ดิฉัน	0.2379
599	ซวย	0.1125	627	ดง	-0.9768	655	ดิน	0.3543
600	ซ่อม	0.2855	628	ดราม่า	0.3535	656	ดิ้น	-0.0033
601	ซ้อม	0.2191	629	ดวง	0.0645	657	ดี	0.9000
602	ซอย	0.0622	630	ด่วน	0.0181	658	ดีกว่า	0.9000
603	ซะ	0.4817	631	ด้วย	-0.3681	659	ดีขึ้น	0.6136
604	ซะงั้น	0.0318	632	ด้วยกัน	-0.0021	660	ดีจริง	0.5468
605	ซัก	0.5022	633	ด้วยซ้ำ	0.5660	661	ดีใจ	1.1772
606	ซ่า	0.4575	634	ด้วยว่า	0.2065	662	ดีแต่	0.7775
607	ซ้าย	0.4209	635	ดอ	0.3360	663	ดีมาก	1.0938
608	ซ่าหริ่ม	0.5546	636	ดอก	0.6662	664	ดึง	0.5849
609	ซ้ำ	-0.2036	637	ดอง	0.0565	665	ดื่ม	0.1767
610	ซ้ำเติม	-0.3973	638	ดัง	0.8089	666	ดุ	0.7090
611	ซิ	0.7503	639	ดัน	0.2472	667	ดู	0.4460

No.	Word	SV	No.	Word	SV	No.	Word	SV
668	ดูซิ	-0.2208	696	เดียวกัน	0.2071	724	ตกนรก	-1.0054
669	ดูด	0.2715	697	เดี๋ยวนี้	0.4267	725	ตกลง	0.1910
670	ดูดี	0.8161	698	เดือดร้อน	0.5024	726	ตง	0.3818
671	ดูตัว	0.3015	699	เดือน	0.3055	727	ตน	0.5298
672	ดูถูก	0.8504	700	แดก	0.6905	728	ต้น	0.6360
673	ดูแล	0.5334	701	แดง	0.4006	729	ต้นไม้	0.3798
674	เด่	0.3611	702	แดน	0.1899	730	ต้นเหตุ	0.1155
675	เด้	0.0922	703	โดน	-0.0709	731	ตนเอง	0.4292
676	เดก	0.1490	704	โดย	0.1732	732	ตบ	0.4633
677	เด็ก	0.0433	705	โดยเฉพาะ	0.3274	733	ต้ม	0.2665
678	เด็กดี	0.7480	706	โดยที่	-0.3945	734	ตรง	0.8387
679	เด็กไทย	0.8936	707	โดยเร็ว	-0.6318	735	ตรงกัน	0.3068
680	เด็กน้อย	-0.2520	708	ใด	0.2611	736	ตรงข้าม	-0.0516
681	เด็กนักเรียน	0.6471	709	ได้	0.3199	737	ตรงนี้	0.5760
682	เด้ง	0.4004	710	ได้ขนาด	-0.1467	738	ตรงไหน	0.8612
683	เด็ดขาด	0.4656	711	ได้เงิน	0.5550	739	ตรวจ	0.4244
684	เดน	-0.3770	712	ได้ใจ	0.3465	740	ตรวจสอบ	0.6875
685	เด่น	0.4781	713	ได้ดี	1.1389	741	ตระกูล	0.2886
686	เดรัจฉาน	-0.6294	714	ได้ที่	0.7745	742	ตรี	0.1909
687	เด้อ	0.8358	715	ได้ประโยชน์	0.3684	743	ตรู	0.0107
688	เดา	0.2855	716	ได้มา	0.7242	744	ตลก	0.5570
689	เดิน	0.5029	717	ได้ยิน	0.1097	745	ตลอด	0.6562
690	เดินทาง	0.5833	718	ได้รับ	-0.2755	746	ตลอดชีวิต	-0.2223
691	เดินหน้า	0.5254	719	ได้อีก	0.0947	747	ตลอดไป	1.3459
692	เดิม	0.5757	720	ตก	0.0795	748	ตลอดเวลา	0.1225
693	เดียว	0.2120	721	ตกงาน	0.3766	749	ตลาด	0.5991
694	เดี่ยว	0.5659	722	ตกใจ	0.0527	750	ตอ	0.4037
695	เดี๋ยว	0.4934	723	ตกต่ำ	0.0076	751	ต่อ	0.0461

No.	Word	SV	No.	Word	SV	No.	Word	SV
752	ต้อง	0.0643	780	ตัดสิน	0.4762	808	ตายห่า	0.1947
753	ต้องการ	0.5470	781	ตัดสินใจ	0.7739	809	ตายโหง	-0.5912
754	ต้องหา	0.4110	782	ตัดหัว	-0.0269	810	ตำ	0.6081
755	ต่อต้าน	0.7389	783	ตัน	0.5455	811	ต่ำ	0.0752
756	ตอน	0.4672	784	ตัว	0.2858	812	ตำรวจ	0.2516
757	ตอนที่	0.4558	785	ตัวจริง	0.8131	813	ตำแหน่ง	1.1505
758	ตอนนั้น	0.2287	786	ตัวดี	0.1599	814	ติ	-0.4790
759	ตอนนี้	0.5244	787	ตัวเรา	0.3364	815	ติ่ง	0.3897
760	ต้อนรับ	0.4960	788	ตัวอย่าง	0.7625	816	ติด	0.2155
761	ตอนแรก	0.4446	789	ตัวเอง	0.3758	817	ติดคุก	0.0632
762	ตอบ	0.7355	790	ตา	0.2376	818	ติดต่อ	0.6221
763	ตอบแทน	0.8919	791	ต่าง	0.4886	819	ติดตัว	-0.0330
764	ต่อไป	0.3320	792	ต่างกัน	0.2915	820	ติดตาม	0.5092
765	ต่อย	0.7068	793	ต่างชาติ	0.8439	821	ตี	0.4418
766	ต่อสู้	0.2400	794	ต่างประเทศ	0.6431	822	ตีน	0.4667
767	ต่อหน้า	0.0890	795	ต่างหาก	0.9099	823	ตึก	-0.0968
768	ตอแหล	0.1310	796	ต้าน	0.8374	824	ตื่น	0.4942
769	ต่อให้	0.2628	797	ตาม	-0.0220	825	ตุ	0.2717
770	ตะ	0.4092	798	ตามเคย	-0.3616	826	ตุ๊ด	-0.0746
771	ตะกวด	0.0187	799	ตามใจ	0.7122	827	ตู	0.4084
772	ตัง	0.7891	800	ตามทัน	-0.7914	828	ตู่	1.0377
773	ตั่ง	0.4857	801	ตามที่	0.5321	829	ตู้	0.1935
774	ตั้ง	0.6386	802	ตามนั้น	0.0484	830	ตูด	0.6706
775	ตังค์	0.7560	803	ตามมา	0.3277	831	เต้น	0.7932
776	ตั้งใจ	0.5744	804	ตามสบาย	0.6776	832	เต็ม	0.6135
777	ตั้งแต่	0.3238	805	ตาย	-0.2065	833	เต็มที่	0.5613
778	ตั้งแต่แรก	0.1500	806	ตายดี	-0.8625	834	เตรียม	0.5792
779	ตัด	0.1733	807	ตายแล้ว	0.0017	835	เตรียมตัว	0.4089

No.	Word	SV	No.	Word	SV	No.	Word	SV
836	เตะ	0.5961	864	ถอย	0.6732	892	เถื่อน	0.2863
837	เติม	0.3206	865	ถ่อย	-0.2486	893	แถ	0.0201
838	เตือน	0.3487	866	ถอยหลัง	0.2872	894	แถม	0.6289
839	แต่	0.4200	867	ถา	0.0923	895	แถลง	0.3895
840	แตก	0.4700	868	ถ้า	0.4712	896	แถว	0.6533
841	แต่ก็	0.2942	869	ถาม	0.5300	897	โถ	-0.2460
842	แตกต่าง	0.5108	870	ถ้าไม่	0.3481	898	โถ่	-0.0905
843	แตกแยก	0.4897	871	ถ่าย	0.4900	899	ทน	0.1959
844	แต่กลับ	0.1002	872	ถ่ายรูป	0.6813	900	ทนาย	0.4040
845	แต่ก่อน	0.4296	873	ถีบ	0.1405	901	ทบ	1.1621
846	แต่ง	0.7804	874	ถึง	0.1148	902	ทรง	1.1272
847	แต่งตั้ง	0.2707	875	ถึงกับ	0.0458	903	ทรมาน	-0.7917
848	แต่งตัว	0.6413	876	ถึงขนาด	-0.0727	904	ทรัพย์	0.5652
849	แตงโม	0.4557	877	ถึงจะ	0.3625	905	ทรัพย์สิน	0.1838
850	แต่ถ้า	0.6997	878	ถึงที่สุด	-0.5306	906	ทราบ	0.3021
851	แต่ละ	0.3380	879	ถือ	0.7262	907	ทวง	0.0547
852	แต่ละคน	0.3526	880	ถือว่า	0.7191	908	ท่วม	0.2297
853	แต่ว่า	0.6034	881	ถุง	0.5789	909	ทหาร	0.5104
854	แตะ	0.5810	882	ถุน	-0.1957	910	ท้อ	0.4767
855	โต	0.3576	883	ถุย	0.7690	911	ทอง	0.6366
856	ใต้	0.6030	884	ถูก	0.9000	912	ท้อง	0.0043
857	ไต	0.7548	885	ถูกจับ	-0.2109	913	ทะเล	0.6718
858	ถนน	0.3472	886	ถูกใจ	0.6866	914	ทะเลาะ	0.2898
859	ถนัด	-0.1470	887	ถูกต้อง	1.0164	915	ทัก	0.6710
860	ถล่ม	0.2340	888	ถูกหวย	0.7220	916	ทักษิณ	1.0701
861	ถวาย	0.3809	889	เถอะ	0.2118	917	ทั่ง	0.1999
862	ถอด	0.7303	890	เถิด	-0.0982	918	ทั้ง	0.1319
863	ถอน	0.1372	891	เถียง	0.5005	919	ทั้งคน	-0.3099

No.	Word	SV	No.	Word	SV	No.	Word	SV
920	ทั้งคู่	-0.0080	948	ทำ	0.2628	976	ที่ดิน	-0.0323
921	ทั้งที่	0.0879	949	ทำความ	0.6895	977	ทีเดียว	0.5193
922	ทั้งนั้น	0.3532	950	ทำงาน	0.7212	978	ที่ทำ	-0.1647
923	ทั้งโลก	0.3395	951	ทำใจ	-0.2858	979	ที่แท้จริง	0.5407
924	ทั้งสอง	0.1588	952	ทำชั่ว	-0.1096	980	ที่นั่น	0.5554
925	ทั้งสองฝ่าย	-0.1886	953	ทำดี	1.2866	981	ทีนี้	0.2812
926	ทั้งหมด	0.2012	954	ทำดีที่สุด	0.6486	982	ที่นี่	0.4855
927	ทั้งหลาย	0.4199	955	ทำได้	0.2685	983	ที่ไป	0.2284
928	ทัน	0.2469	956	ทำตัว	0.7367	984	ที่ผ่านมา	0.5640
929	ทั่น	0.4461	957	ทำตาม	0.6045	985	ทีม	1.2456
930	ทันที	0.3360	958	ทำนา	0.4352	986	ที่มา	0.4353
931	ทับ	0.1675	959	ทำบุญ	0.7877	987	ที่มี	0.5271
932	ทั่ว	0.2845	960	ทำผิด	-0.1087	988	ที่รัก	-0.0640
933	ทั่วประเทศ	0.6308	961	ทำมาหากิน	0.5942	989	ทีไร	-0.1378
934	ทั่วไป	0.5534	962	ทำไม	0.0250	990	ที่แล้ว	0.3047
935	ทั่วโลก	0.8892	963	ทำร้าย	-0.1345	991	ที่ว่า	0.6325
936	ทัศนคติ	0.1950	964	ทำลาย	0.5618	992	ทีวี	0.6118
937	ทา	0.5247	965	ทำศพ	0.0671	993	ที่สุด	0.1186
938	ท่า	0.6414	966	ทำหน้าที่	0.7534	994	ทีหลัง	0.2694
939	ท้า	0.5197	967	ทำให้	0.4546	995	ที่ไหน	0.5216
940	ทำ	0.3191	968	ทำอะไร	0.1841	996	ที่อยู่	-0.2648
941	ทาง	0.3969	969	ทิ้ง	0.3075	997	ที่อื่น	0.2235
942	ทางการเมือง	0.4733	970	ที	0.3051	998	ทุ	0.4767
943	ทางโลก	0.2633	971	ที่	0.2562	999	ทุก	0.0344
944	ทาน	0.5547	972	ที่	0.5512	1000	ทุกข์	0.1923
945	ท่าน	0.8040	973	ที่เกิด	-0.2165	1001	ทุกคน	0.1576
946	ทาม	0.3044	974	ที่เกิดขึ้น	-0.8948	1002	ทุกครั้ง	-0.0381
947	ทาส	0.1230	975	ที่จะ	0.3565	1003	ทุกตัว	0.1690

No.	Word	SV	No.	Word	SV	No.	Word	SV
1004	ทุกท่าน	-0.3501	1032	เท่าไหร่	0.3731	1060	นกหวีด	0.1524
1005	ทุกที	-0.0541	1033	เทียน	0.5889	1061	นม	0.4018
1006	ทุกที่	0.4372	1034	เทียบ	0.3891	1062	นโยบาย	0.4113
1007	ทุกปี	0.2438	1035	เที่ยว	0.6032	1063	นรก	-0.1676
1008	ทุกฝ่าย	-0.0181	1036	เทือก	0.1712	1064	นอ	0.4677
1009	ทุกวัน	0.1086	1037	แท้	0.0154	1065	นอก	0.8509
1010	ทุกวันนี้	0.3366	1038	แท็กซี่	0.3600	1066	นอกจาก	0.1697
1011	ทุกอย่าง	0.5456	1039	แทง	0.1736	1067	นอกประเทศ	0.5702
1012	ทุจริต	-0.1733	1040	แทน	0.1526	1068	น้อง	0.0339
1013	ทุน	0.3238	1041	แทนที่จะ	0.0840	1069	นอน	0.4743
1014	ทุบ	0.1542	1042	แทบ	-0.1329	1070	น้อย	0.2135
1015	ทุเรศ	0.2497	1043	แทบจะ	0.0549	1071	น้อยลง	0.1979
1016	เท	0.4755	1044	โท	0.1540	1072	นะ	0.0918
1017	เท่	1.1267	1045	โทด	0.2207	1073	น่ะ	0.0344
1018	เท็จ	-0.0623	1046	โทร	0.5552	1074	นะคะ	-0.1101
1019	เทพ	0.6605	1047	โทษ	-0.0943	1075	นะจ๊ะ	0.7270
1020	เทวดา	0.5946	1048	ไท	0.3099	1076	นะเนี่ย	0.3267
1021	เทศ	0.4355	1049	ไทย	0.8461	1077	นัก	0.7989
1022	เทอญ	-0.1053	1050	ไทยแลนด์	0.6342	1078	นักการเมือง	0.6347
1023	เท่า	0.3535	1051	ธนาคาร	-0.0687	1079	นักกีฬา	0.7565
1024	เท้า	0.5904	1052	ธรรม	0.7760	1080	นักข่าว	0.3227
1025	เท่ากัน	0.6061	1053	ธรรมชาติ	0.8526	1081	นักท่องเที่ยว	0.4341
1026	เท่ากับ	0.1187	1054	ธรรมดา	0.5980	1082	นักโทษ	0.2647
1027	เท่าที่	0.4319	1055	ธิป	1.0588	1083	นักเรียน	0.2348
1028	เท่านั้น	0.4231	1056	ธุรกิจ	0.5255	1084	นักเลง	0.4680
1029	เท่านั้นเอง	0.3772	1057	เธอ	0.8892	1085	นักวิชาการ	0.1208
1030	เท่านี้	0.6124	1058	โธ่	-0.1105	1086	นักศึกษา	0.6049
1031	เท่าไร	0.2841	1059	นก	0.9174	1087	นักหนา	0.4203

No.	Word	SV	No.	Word	SV	No.	Word	SV
1088	นัง	0.2753	1116	นายก	1.2648	1144	นี่แหละ	0.5714
1089	นั่ง	0.3428	1117	น่ายกย่อง	1.1034	1145	นี้แหละ	0.7111
1090	นัด	0.5087	1118	นายทุน	0.2357	1146	นี่เอง	0.3915
1091	นั่น	0.4013	1119	นายพล	0.4427	1147	นึก	0.3772
1092	นั้น	0.2398	1120	น่ารัก	1.6159	1148	นึกถึง	-0.1691
1093	นั่นแหละ	0.0386	1121	น่าสงสาร	-0.8255	1149	นึง	0.3284
1094	นั้นแหละ	0.1450	1122	น่าเห็นใจ	-0.3358	1150	นุ	0.1373
1095	นับ	0.5011	1123	น่าอาย	0.1455	1151	เนชั่น	0.3334
1096	นับถือ	1.2087	1124	นำ	0.6785	1152	เน็ต	0.0981
1097	นา	0.3167	1125	น้ำ	0.6169	1153	เน้น	0.7068
1098	น่า	0.4469	1126	น้ำใจ	1.5207	1154	เนรคุณ	-0.1294
1099	น้า	0.3653	1127	น้ำตา	-0.7079	1155	เน้อ	0.3922
1100	น้ำ	0.4544	1128	น้ำตาไหล	-0.8451	1156	เนอะ	0.3457
1101	น่ากลัว	-0.2287	1129	น้ำท่วม	0.1204	1157	เน่า	0.6185
1102	น่าเกลียด	0.3263	1130	น้ำมัน	1.0105	1158	เนาะ	0.3928
1103	นาง	0.9756	1131	นำมา	0.5083	1159	เนี่ย	0.3459
1104	นางงาม	0.4365	1132	นิ	0.7755	1160	เนี้ย	0.2767
1105	นางฟ้า	0.6221	1133	นิ่ง	0.7690	1161	เนียน	0.1565
1106	น่าจะ	0.5213	1134	นิด	0.3620	1162	เนื้อ	0.1235
1107	น่าจะเป็น	0.5093	1135	นิดเดียว	0.0420	1163	แน่	0.5051
1108	น่าชื่นชม	1.1282	1136	นิดหน่อย	0.4283	1164	แน่จริง	0.3647
1109	น่าดู	0.0877	1137	นิมนต์	0.0787	1165	แน่ใจ	0.2222
1110	นาที	0.3053	1138	นิยม	0.5762	1166	แน่น	0.1865
1111	นาน	0.7798	1139	นิ้ว	0.2784	1167	แน่นอน	0.4932
1112	น่าเบื่อ	0.3226	1140	นิสัย	0.8163	1168	แนว	0.4586
1113	นาม	-0.3474	1141	นี่	0.3825	1169	แนะนำ	0.7906
1114	นามสกุล	-0.1611	1142	นี้	0.3382	1170	โน	0.7075
1115	นาย	0.4202	1143	นี่หว่า	0.1312	1171	โน่น	0.4863

No.	Word	SV	No.	Word	SV	No.	Word	SV
1172	โน้น	0.3228	1200	บ่อย	0.2366	1228	บุ	0.5033
1173	ใน	0.3287	1201	บอล	1.1058	1229	บุก	0.2708
1174	ในใจ	0.3663	1202	บอลโลก	0.5520	1230	บุคคล	-0.0518
1175	ในทาง	0.8853	1203	บัก	-0.0374	1231	บุญ	0.7739
1176	ในบ้าน	0.4390	1204	บัง	0.5168	1232	บุญคุณ	0.3113
1177	ในประเทศ	0.6826	1205	บังคับ	0.6954	1233	บุ๋ม	1.7468
1178	ในเมือง	0.5206	1206	บัตร	0.2836	1234	เบ่ง	-0.1047
1179	ในโลก	0.8230	1207	บัว	1.2464	1235	เบอร์	0.8073
1180	ในหลวง	0.4813	1208	บ้า	0.3691	1236	เบา	0.3317
1181	ในอนาคต	0.5625	1209	บาง	0.4674	1237	เบี้ยว	0.2411
1182	ไน	0.4810	1210	บ้าง	0.3319	1238	เบื่อ	0.9541
1183	บก	-0.0529	1211	บางคน	0.6043	1239	เบื้องหลัง	0.0811
1184	บท	0.5163	1212	บางครั้ง	0.2272	1240	แบ	0.4328
1185	บทเรียน	-0.0911	1213	บางที	0.1075	1241	แบ่ง	0.5896
1186	บน	0.2507	1214	บาดเจ็บ	-1.2702	1242	แบงก์	0.2069
1187	บ่น	0.5423	1215	บาท	0.3709	1243	แบน	0.1510
1188	บรรยาย	-0.7558	1216	บาน	0.1678	1244	แบบ	0.5426
1189	บริการ	0.8235	1217	บ้าน	0.6179	1245	แบบนี้	0.1325
1190	บริจาค	0.3758	1218	บ้านเมือง	0.6841	1246	แบบอย่าง	0.8563
1191	บริษัท	0.4104	1219	บาป	-0.2094	1247	โบราณ	0.6080
1192	บริสุทธิ์	-0.0651	1220	บาปกรรม	-0.5579	1248	ใบ	0.7109
1193	บริหาร	1.0238	1221	บาย	0.3761	1249	ใบขับขี่	0.1653
1194	บวก	0.5216	1222	บารมี	0.6239	1250	ใบสั่ง	0.0617
1195	บวช	0.1884	1223	บ้าอำนาจ	0.2770	1251	ปก	0.3869
1196	บ่อ	0.2972	1224	บิ๊ก	0.4330	1252	ปกครอง	0.7423
1197	บอก	0.4756	1225	บิดเบือน	0.1508	1253	ปกติ	0.6064
1198	บอกตรง	0.4172	1226	บิน	0.4130	1254	ปกป้อง	0.1636
1199	บ่อน	0.6464	1227	บีบ	-0.0772	1255	ปฏิบัติ	0.7474

No.	Word	SV	No.	Word	SV	No.	Word	SV
1256	ปฏิรูป	0.2067	1283	ประมูล	-0.1531	1310	ป่วย	0.1192
1257	ปฏิบัติ	0.7956	1284	ประยุทธ์	0.8982	1311	ปอ	0.4382
1258	ปฏิรูป	0.8263	1285	ประยุทธ	1.1735	1312	ป้องกัน	0.1735
1259	ปฏิวัติ	0.5179	1286	ประโยชน์	0.7661	1313	ปอบ	-0.2625
1260	ปรองดอง	0.1636	1287	ประวัติ	0.4619	1314	ปะ	0.6655
1261	ประ	0.2123	1288	ประวัติศาสตร์	0.6347	1315	ป่ะ	0.5921
1262	ประกวด	0.6187				1316	ปัจจุบัน	0.6682
1263	ประกัน	0.1692	1289	ประสาท	0.0566	1317	ปัญญา	0.4166
1264	ประกาศ	0.6454	1290	ประหยัด	0.9461	1318	ปัญญาอ่อน	0.4058
1265	ประจาน	-0.0264	1291	ประหาร	-0.5077	1319	ปัญหา	0.5413
1266	ประจำ	0.3692	1292	ประหารชีวิต	-0.5724	1320	ปัน	0.3561
1267	ประชด	0.1475	1293	ปรับ	0.8091	1321	ปั่น	0.7078
1268	ประชา	0.6777	1294	ปรับปรุง	0.4939	1322	ปา	0.1998
1269	ประชาชน	0.6608	1295	ปราบ	0.9773	1323	ป่า	0.9015
1270	ประชาธิปไตย	1.0905	1296	ปรี๊ด	-0.2003	1324	ป้า	0.4389
			1297	ปลง	-0.2468	1325	ป๋า	0.4308
1271	ประชานิยม	0.0630	1298	ปลด	0.7166	1326	ปาก	0.8289
1272	ประชามติ	0.4420	1299	ปล้น	0.5169	1327	ปากดี	0.0405
1273	ประชุม	0.2121	1300	ปลวก	0.2999	1328	ปากท้อง	0.4480
1274	ประเด็น	0.4781	1301	ปลอดภัย	0.2072	1329	ปากหมา	0.5090
1275	ประตู	0.4321	1302	ปลอม	0.2723	1330	ป่านนี้	0.2633
1276	ประท้วง	0.6434	1303	ปล่อย	0.3506	1331	ป้าย	0.7355
1277	ประเทศ	0.6653	1304	ปลา	0.3372	1332	ปิด	0.7289
1278	ประเทศชาติ	0.6225	1305	ปลาย	0.6505	1333	ปิดหน้า	-0.7118
1279	ประเทศไทย	0.5169	1306	ปลื้ม	0.8080	1334	ปี	0.5498
1280	ประเภท	0.3954	1307	ปลูก	0.5321	1335	ปืน	0.5635
1281	ประมาณ	0.4398	1308	ปวด	-0.2988	1336	ปู	0.5254
1282	ประมาท	-0.2428	1309	ป่วน	0.3509	1337	ปู่	0.0631

No.	Word	SV	No.	Word	SV	No.	Word	SV
1338	เป้	0.3879	1366	เป่า	0.3946	1394	ผ้า	0.1019
1339	เป็ด	0.1266	1367	เป่านกหวีด	0.2722	1395	ผ่าน	0.1223
1340	เป็น	0.2824	1368	เปิด	0.7106	1396	ผ้าเหลือง	-0.3426
1341	เป็นกลาง	0.4927	1369	เปิดเผย	0.7053	1397	ผิด	-0.9000
1342	เป็นการ	0.2127	1370	แปลก	0.7022	1398	ผิดกฎหมาย	-0.1270
1343	เป็นกำลัง	0.3406	1371	แปลว่า	0.1342	1399	ผิดถูก	0.3689
1344	เป็นข่าว	0.2909	1372	โปรด	0.1408	1400	ผิดพลาด	0.3973
1345	เป็นความ	0.1484	1373	ไป	0.3627	1401	ผิดหวัง	0.5220
1346	เป็นไง	0.6045	1374	ไปกับ	-0.1393	1402	ผี	0.3297
1347	เป็นจริง	0.4273	1375	ไปดู	0.9804	1403	ผึ้ง	0.6725
1348	เป็นได้	0.5648	1376	ไปถึงไหน	0.2436	1404	ผุ	0.0893
1349	เป็นตัว	0.5987	1377	ไปเที่ยว	0.7084	1405	ผู้	-0.2978
1350	เป็นไปได้	0.1621	1378	ไปมา	0.3660	1406	ผูกคอตาย	-0.3075
1351	เป็นไปตาม	-0.1047	1379	ไปสู่สุขคติ	-0.6873	1407	ผู้ชาย	0.4187
1352	เป็นผู้	0.6966	1380	ไปหา	0.3300	1408	ผู้โดยสาร	-0.0403
1353	เป็นเพื่อน	0.3492	1381	ไปไหน	0.3422	1409	ผู้ที่	-0.6066
1354	เป็นเรื่อง	0.1816	1382	ผบ	1.0247	1410	ผู้นำ	0.7343
1355	เป็นโรค	-0.0077	1383	ผม	0.6928	1411	ผู้บริหาร	0.5189
1356	เป็นไร	0.1366	1384	ผล	0.2448	1412	ผู้ปกครอง	-0.4008
1357	เป็นหนี้	0.1004	1385	ผลกรรม	-1.1820	1413	ผู้ว่า	0.7511
1358	เป็นห่วง	0.4887	1386	ผลงาน	0.9946	1414	ผู้เสียชีวิต	-0.7113
1359	เป็นใหญ่	-0.0840	1387	ผลประโยชน์	0.5853	1415	ผู้หญิง	0.3383
1360	เป็นอะไร	-0.1877	1388	ผลิต	0.5061	1416	ผู้ใหญ่	0.2787
1361	เปรต	0.1558	1389	ผ่อน	0.1358	1417	ผู้อื่น	0.2620
1362	เปล่า	0.4817	1390	ผอม	0.3928	1418	เผด็จการ	0.7565
1363	เปลี่ยน	0.8558	1391	ผัก	0.4595	1419	เผา	0.2772
1364	เปลี่ยนไป	0.4973	1392	ผัว	0.1899	1420	เผื่อ	0.6867
1365	เปลือง	0.2642	1393	ผัวเมีย	-0.1102	1421	แผน	0.1072

No.	Word	SV
1422	แผ่น	0.3455
1423	แผ่นดิน	0.4288
1424	แผ่นดินไทย	0.2967
1425	โผล่	-0.0232
1426	ฝน	0.6629
1427	ฝนตก	0.1719
1428	ฝรั่ง	0.9189
1429	ฝัง	0.2088
1430	ฝั่ง	0.3286
1431	ฝัน	0.8500
1432	ฝาก	0.5324
1433	ฝ่าย	0.2579
1434	ฝ่ายตรงข้าม	0.3863
1435	ฝีมือ	1.0930
1436	ฝึก	1.1168
1437	เฝ้า	0.4067
1438	พก	0.2924
1439	พง	0.4284
1440	พ้น	0.2630
1441	พนักงาน	0.2122
1442	พบ	0.2772
1443	พม่า	0.9855
1444	พยาบาล	0.5426
1445	พยายาม	0.4438
1446	พร	0.0824
1447	พรรค	0.6218
1448	พรรคการเมือง	0.1216
1449	พรรคพวก	-0.0515
1450	พร้อม	0.8187
1451	พระ	0.5640
1452	พระเจริญ	-0.2490
1453	พระเจ้า	0.2696
1454	พระพุทธเจ้า	0.3108
1455	พระพุทธศาสนา	0.1447
1456	พระสงฆ์	0.1924
1457	พระองค์	0.5958
1458	พระเอก	0.6316
1459	พรุ่งนี้	0.3973
1460	พฤติกรรม	0.2038
1461	พล	0.3099
1462	พลัง	0.7467
1463	พลาด	0.2285
1464	พวก	0.3925
1465	พวกเขา	-0.0685
1466	พวกคุณ	0.5865
1467	พวกพ้อง	-0.1277
1468	พวกเรา	0.4040
1469	พอ	0.2876
1470	พ่อ	-0.1791
1471	พอกัน	-0.1701
1472	พ่อค้า	0.5140
1473	พอจะ	0.3406
1474	พอใจ	0.8375
1475	พอดี	0.3154
1476	พอได้	0.3477
1477	พอเพียง	0.5892
1478	พ่อแม่	-0.3656
1479	พอแล้ว	-0.0812
1480	พัก	0.4715
1481	พักผ่อน	0.4695
1482	พัง	0.5171
1483	พัฒนา	1.2222
1484	พัทยา	0.5204
1485	พัน	-0.0572
1486	พันธ์	0.3706
1487	พา	0.3746
1488	พาดหัวข่าว	-0.1564
1489	พาย	0.4888
1490	พาล	0.0930
1491	พิการ	0.0572
1492	พิจารณา	0.5846
1493	พิมพ์	0.4668
1494	พิเศษ	0.5772
1495	พิสูจน์	0.4442
1496	พี	0.6049
1497	พี่	0.8154
1498	พี่น้อง	0.1331
1499	พึ่ง	0.3848
1500	พื้นที่	0.3918
1501	พุด	0.4047
1502	พุทธ	0.3739
1503	พูด	0.5456

No.	Word	SV	No.	Word	SV	No.	Word	SV
1504	พูดจา	0.6685	1532	โพส	0.2042	1560	มหาลัย	0.4380
1505	พูดถึง	0.4997	1533	โพสต์	0.2107	1561	มอ	0.6641
1506	พูดว่า	0.3627	1534	ฟรี	0.5950	1562	มอง	0.5592
1507	เพ่	0.3500	1535	ฟ้อง	0.1206	1563	มองไม่เห็น	-0.1190
1508	เพจ	0.0210	1536	ฟัง	0.5401	1564	มองว่า	0.6346
1509	เพชร	0.7200	1537	ฟัน	0.5625	1565	มองเห็น	0.5458
1510	เพราะ	0.3469	1538	ฟ้า	0.6188	1566	ม็อบ	0.1276
1511	เพราะว่า	0.7067	1539	เฟส	0.4749	1567	ม๊อบ	-0.0326
1512	เพราะอะไร	0.2961	1540	แฟน	0.4011	1568	มะ	0.8005
1513	เพลง	0.0601	1541	ไฟ	0.4672	1569	ม่ะ	0.4832
1514	เพลีย	0.0889	1542	ภพ	-0.1610	1570	มะลิ	-0.2054
1515	เพศ	0.4140	1543	ภพภูมิ	-0.3475	1571	มัก	0.3739
1516	เพิ่ง	-0.0598	1544	ภัย	0.3269	1572	มักง่าย	-0.2411
1517	เพิ่ม	0.7245	1545	ภาค	0.8326	1573	มัง	0.6097
1518	เพิ่มขึ้น	0.5955	1546	ภาคใต้	0.7215	1574	มั่ง	0.5725
1519	เพียง	0.2813	1547	ภาพ	-0.0277	1575	มั้ง	0.5321
1520	เพียงแค่	0.3529	1548	ภายใน	0.5321	1576	มัน	-0.0343
1521	เพี้ยน	-0.3256	1549	ภาระ	0.3920	1577	มั่นใจ	1.0428
1522	เพียบ	0.1954	1550	ภาษา	0.6323	1578	มัย	0.3420
1523	เพื่อ	0.3761	1551	ภาษาไทย	0.7165	1579	มัว	0.5477
1524	เพื่อน	0.3032	1552	ภาษี	0.3436	1580	มั่ว	0.3824
1525	เพื่อนบ้าน	0.6593	1553	ภูมิใจ	0.9393	1581	มา	0.1866
1526	เพื่อให้	0.4193	1554	มด	0.2885	1582	ม้า	0.6593
1527	แพ้	0.8292	1555	มติ	0.6285	1583	มาก	0.3774
1528	แพง	0.9729	1556	มนุษย์	-0.0160	1584	มากกว่า	0.3545
1529	แพทย์	0.7341	1557	มโน	0.3505	1585	มากขึ้น	0.5780
1530	แพะ	0.1674	1558	มวย	0.7322	1586	มากนัก	0.5150
1531	โพล	0.1500	1559	มหา	0.4968	1587	มากมาย	0.4554

No.	Word	SV	No.	Word	SV	No.	Word	SV
1588	มาก่อน	0.3938	1616	มีทาง	0.2048	1644	เมือง	0.5713
1589	มาจาก	0.4148	1617	มีน้ำใจ	1.0112	1645	เมืองไทย	0.8209
1590	มาต	0.4742	1618	มีประโยชน์	0.3591	1646	เมืองนอก	0.5366
1591	มาตรฐาน	0.2994	1619	มีปัญหา	0.6184	1647	เมื่อไร	0.2470
1592	มาตรา	-0.0063	1620	มีผล	0.4235	1648	เมื่อวาน	0.1055
1593	มาถึง	0.3556	1621	มีเรื่อง	0.1881	1649	เมื่อไหร่	0.0997
1594	มาน	0.3263	1622	มีลูก	-0.0462	1650	แม่	-0.3119
1595	มานะ	0.1481	1623	มีสติ	0.4075	1651	แม้	0.1639
1596	มาบ	0.4685	1624	มีสิทธิ์	0.1900	1652	แม้กระทั่ง	-0.3161
1597	มาเฟีย	0.2489	1625	มีหน้า	-0.1011	1653	แม่ค้า	0.5754
1598	ม่าย	0.1972	1626	มีหัว	-0.3348	1654	แมง	0.1716
1599	มาร	-0.1860	1627	มีเหตุผล	0.2959	1655	แม่ง	0.2190
1600	มาร์ค	0.1046	1628	มีอำนาจ	0.3440	1656	แม้ง	-0.1728
1601	มารยาท	0.5249	1629	มีโอกาส	0.7006	1657	แม้แต่	-0.0234
1602	มาเลย	0.4648	1630	มึง	0.1881	1658	แม่น	0.5046
1603	มาแล้ว	0.3697	1631	มึน	-0.1056	1659	แมร่ง	0.2312
1604	มาหา	0.1538	1632	มืด	0.0606	1660	แมว	0.0925
1605	มิ	0.4649	1633	มือ	0.5846	1661	แม้ว	0.1393
1606	มิใช่	0.4932	1634	มือถือ	0.0949	1662	โม	0.5981
1607	มิน่า	0.2276	1635	มือใหม่	-0.7259	1663	โม้	0.4249
1608	มี	0.4654	1636	มุง	-0.0611	1664	โมโห	-0.0747
1609	มี่	0.4775	1637	มุม	0.7626	1665	ไม่	0.3659
1610	มีความรู้	0.6861	1638	เม็ด	0.2502	1666	ไม้	0.5101
1611	มีความสุข	0.9460	1639	เม้น	0.3872	1667	ไม่กลัว	0.2711
1612	มีค่า	0.1391	1640	เมา	-0.1677	1668	ไม่กล้า	0.3313
1613	มีเงิน	0.7796	1641	เมีย	0.2828	1669	ไม่กี่	0.4684
1614	มีชีวิต	0.1756	1642	เมื่อ	0.3512	1670	ไม่เกิน	0.1456
1615	มืด	0.2843	1643	เมื่อก่อน	0.6137	1671	ไม่เกี่ยว	0.4906

No.	Word	SV	No.	Word	SV	No.	Word	SV
1672	ไม่เข้าใจ	0.0029	1700	ไม่พอใจ	0.4961	1728	ยกเลิก	0.6217
1673	ไม่ค่อย	0.6945	1701	ไม่มี	0.1854	1729	ยกเว้น	0.3432
1674	ไม่คิด	-0.1938	1702	ไม่มีใคร	0.0500	1730	ยม	0.5783
1675	ไม่คุ้ม	-0.0420	1703	ไม่มีทาง	0.0412	1731	ยศ	0.5368
1676	ไม่เคย	0.3016	1704	ไม่มีวัน	-0.2680	1732	ยอ	0.1906
1677	ไม่งั้น	0.4890	1705	ไม่ยอม	0.4880	1733	ยอด	0.8081
1678	ไม่จำเป็นต้อง	0.7224	1706	ไม่ยอมรับ	0.0955	1734	ยอดเยี่ยม	1.1730
1679	ไม่เชื่อ	0.6839	1707	ไม่ยาก	0.2744	1735	ย้อน	0.0869
1680	ไม่ใช่	0.2808	1708	ไม่รอด	0.3817	1736	ยอม	0.4036
1681	ไม่ดี	0.7389	1709	ไม่รู้	-0.3077	1737	ย่อม	0.4725
1682	ไม่ได้	0.2849	1710	ไม่ร้	-0.4381	1738	ยอมรับ	0.6905
1683	ไม่ต้อง	0.6578	1711	ไม่รู้จัก	-0.0013	1739	ยัง	0.3054
1684	ไม่ต้องการ	0.1567	1712	ไมล์	0.9473	1740	ยังกะ	0.5895
1685	ไม่ทราบ	0.2219	1713	ไม่ว่า	0.2609	1741	ยังงัย	0.1355
1686	ไม่ทัน	-0.1065	1714	ไม่สน	0.1442	1742	ยังงี้	0.3027
1687	ไม่ทำ	0.5464	1715	ไม่สนใจ	0.5510	1743	ยังไง	-0.0488
1688	ไม่น่า	-0.4898	1716	ไม่สมควร	-0.0968	1744	ยังจะ	0.1241
1689	ไม่น่าเชื่อ	0.0097	1717	ไม่เห็น	0.5320	1745	ยังมี	0.4935
1690	ไม่นาน	-0.0915	1718	ไม่เห็นด้วย	0.6455	1746	ยัด	0.1447
1691	ไม่น่าเลย	-0.7246	1719	ไม่เหมาะสม	0.0757	1747	ยัน	0.2165
1692	ไม่แน่	0.0593	1720	ไม่ให้	0.3254	1748	ยา	0.1878
1693	ไม่เป็น	0.5628	1721	ไม่ไหว	0.1256	1749	ย่า	0.4577
1694	ไม่เป็นไร	0.7752	1722	ไม่อยู่	0.0576	1750	ยาก	0.5128
1695	ไม่แปลก	0.8141	1723	ไม่อาย	0.2424	1751	ยาง	1.0157
1696	ไม่ไป	0.5810	1724	ไม่เอา	0.5861	1752	ย่าง	-0.0407
1697	ไม่ผิด	0.4687	1725	ยก	0.9372	1753	ยาบ้า	0.1878
1698	ไม่พ้น	-0.1798	1726	ยกฟ้อง	0.1569	1754	ยาม	0.7969
1699	ไม่พอ	0.2156	1727	ยกย่อง	0.7843	1755	ยาย	0.1970

No.	Word	SV	No.	Word	SV	No.	Word	SV
1756	ย้าย	0.4684	1784	เย็น	0.4529	1812	รวมทั้ง	-0.0264
1757	ยาว	0.9829	1785	เยอรมนี	1.2798	1813	รวย	1.0736
1758	ยาเสพติด	0.2858	1786	เยอะ	0.8231	1814	รส	0.2917
1759	ยิง	0.1322	1787	เยอะแยะ	0.7228	1815	รอ	0.5661
1760	ยิ่ง	0.2996	1788	เยาวชน	0.3280	1816	รอง	0.8065
1761	ยิ่งกว่า	-0.1801	1789	เยี่ยม	1.5961	1817	ร้อง	-0.2175
1762	ยิ่งใหญ่	0.4617	1790	แย่	0.2019	1818	ร้องไห้	-0.7546
1763	ยิน	0.4033	1791	แยก	0.4296	1819	รอด	0.1602
1764	ยินดี	1.6225	1792	แยกแยะ	0.5032	1820	ร้อน	0.4833
1765	ยิ้ม	0.6712	1793	แย่ง	0.5692	1821	รอบ	0.3210
1766	ยี	0.2527	1794	แย่มาก	-0.3005	1822	รอย	0.3071
1767	ยึด	0.9956	1795	แยะ	0.6771	1823	ร้อย	0.4019
1768	ยึดทรัพย์	0.1066	1796	โยง	0.0379	1824	รอลงอาญา	-0.2810
1769	ยึดอำนาจ	0.3336	1797	โยน	0.3959	1825	ระดับ	0.5188
1770	ยืน	0.3151	1798	โยม	0.1246	1826	ระบบ	0.6250
1771	ยื่น	0.0175	1799	รก	-0.3362	1827	ระบอบ	0.4213
1772	ยืนนาน	-1.0803	1800	รด	0.1786	1828	ระเบิด	-0.0467
1773	ยืนยัน	0.6919	1801	รถ	0.3810	1829	ระยำ	-0.6307
1774	ยืม	0.2343	1802	รถติด	0.4712	1830	ระวัง	0.3944
1775	ยุ	0.5779	1803	รถไฟ	0.1764	1831	ระหว่าง	0.3483
1776	ยุค	0.7514	1804	รถยนต์	0.3006	1832	รัก	0.9000
1777	ยุ่ง	0.8077	1805	รบ	0.5244	1833	รักษา	0.6339
1778	ยุด	0.3316	1806	รบกวน	0.4754	1834	รัง	0.2516
1779	ยุติธรรม	0.2053	1807	รม	0.2663	1835	รังแก	0.5562
1780	ยุทธ	0.4442	1808	รวม	0.2600	1836	รัฐ	0.7231
1781	ยุบ	0.2853	1809	ร่วม	0.1308	1837	รัฐธรรมนูญ	0.3674
1782	เย้	1.4970	1810	รวมกัน	0.2650	1838	รัฐบาล	0.6658
1783	เย็ด	-0.0503	1811	ร่วมกัน	0.0423	1839	รัฐประหาร	0.6623

No.	Word	SV	No.	Word	SV	No.	Word	SV
1840	รัฐมนตรี	0.3710	1868	รายได้	0.9032	1896	เรา	0.5538
1841	รับ	0.1181	1869	ร้ายแรง	-0.0005	1897	เริ่ม	0.6003
1842	รับเงิน	-0.0129	1870	รำคาญ	0.1533	1898	เริ่มต้น	0.5979
1843	รับจ้าง	0.0725	1871	ร่ำรวย	0.4469	1899	เรียก	0.4227
1844	รับใช้	0.3128	1872	รี	0.5734	1900	เรียกร้อง	0.2992
1845	รับได้	0.0418	1873	รีด	0.1579	1901	เรียกว่า	0.8903
1846	รับโทษ	-0.7499	1874	รีบ	0.2387	1902	เรียน	0.7570
1847	รับน้อง	0.1268	1875	รึ	0.4147	1903	เรียบร้อย	0.7262
1848	รับผิด	0.0270	1876	รึไง	0.2040	1904	เรือ	0.6059
1849	รับผิดชอบ	-0.0515	1877	รึเปล่า	0.3676	1905	เรือง	0.3978
1850	รับรอง	0.9523	1878	รื้อ	0.3882	1906	เรื่อง	0.2354
1851	รับรู้	-0.3315	1879	รุ่น	0.2999	1907	เรื่องการ	0.7487
1852	รัสเซีย	0.7653	1880	รุนแรง	-0.2633	1908	เรื่องจริง	0.1567
1853	รา	0.4662	1881	รุม	0.0083	1909	เรือคำน้ำ	0.4929
1854	รากหญ้า	0.6322	1882	รู	0.3998	1910	เรื่อย	0.4849
1855	ราคา	1.0409	1883	รู้	0.1609	1911	แรก	0.6650
1856	ราง	0.1163	1884	รู้กัน	0.5775	1912	แรง	0.4795
1857	ร่าง	0.1141	1885	รู้จัก	0.5060	1913	แรงงาน	0.1843
1858	ร่างกาย	0.4508	1886	รู้ดี	0.2584	1914	แร้ว	0.0670
1859	รางวัล	1.2066	1887	รู้ตัว	0.2042	1915	แระ	0.4386
1860	ราช	0.4683	1888	รูป	0.2004	1916	โร	0.4915
1861	ราชการ	0.4339	1889	รูปแบบ	0.0871	1917	โรค	0.2200
1862	ร้าน	0.9119	1890	รู้เรื่อง	0.1423	1918	โรคจิต	0.0867
1863	ราบ	0.3579	1891	รู้สึก	-0.4862	1919	โรง	0.2479
1864	ราย	0.0827	1892	รู้อยู่	0.1327	1920	โรงงาน	0.4150
1865	ร้าย	-0.1980	1893	เร่ง	-0.0133	1921	โรงพยาบาล	0.2229
1866	รายการ	0.7633	1894	เร็ว	-0.1959	1922	โรงพัก	0.1898
1867	รายงานตัว	0.2977	1895	เรอะ	0.2093	1923	โรงเรียน	0.3562

No.	Word	SV	No.	Word	SV	No.	Word	SV
1924	ไร	0.1656	1952	ลอย	0.4137	1980	ลูกค้า	0.3243
1925	ไร่	0.1434	1953	ลอยนวล	-0.1867	1981	ลูกชาย	0.0896
1926	ไร้	0.3506	1954	ละ	0.5271	1982	ลูกน้อง	0.3663
1927	ไร้สาระ	0.7283	1955	ล่ะ	0.3634	1983	ลูกผู้ชาย	0.5206
1928	ลง	0.3009	1956	ละคร	0.6703	1984	ลูกพี่	0.2302
1929	ลงข่าว	0.0922	1957	ลัก	0.7435	1985	ลูกเมีย	-0.3301
1930	ลงทุน	0.4031	1958	ลักษณ์	0.9004	1986	ลูกศิษย์	0.1777
1931	ลงโทษ	-0.5771	1959	ลัง	0.2964	1987	ลูกสาว	-0.0492
1932	ลงนรก	-0.3681	1960	ลับ	0.5571	1988	ลูกหลาน	0.1715
1933	ลงมา	0.3040	1961	ลา	0.1937	1989	เล็ก	0.6059
1934	ลด	0.6298	1962	ล่า	0.5371	1990	เลข	1.1389
1935	ลดโทษ	-0.0312	1963	ลาก	0.0984	1991	เลน	0.2563
1936	ลดลง	0.1764	1964	ล้าง	0.3010	1992	เล่น	0.9091
1937	ล้น	0.0758	1965	ล้าน	0.8660	1993	เล่นการเมือง	0.4900
1938	ลบ	0.6940	1966	ลาย	0.5364	1994	เลย	0.3875
1939	ลบหลู่	0.3058	1967	ลาว	0.8040	1995	เล้ย	0.7048
1940	ลม	0.2761	1968	ลาออก	0.4449	1996	เลว	-0.9000
1941	ล้ม	0.5020	1969	ลำ	0.1442	1997	เลวร้าย	-0.6260
1942	ล่มจม	0.1935	1970	ลำบาก	0.4218	1998	เละ	0.1418
1943	ลวง	0.2092	1971	ลิง	0.3473	1999	เล่า	0.1835
1944	ล่วงหน้า	0.2788	1972	ลิ้น	-0.0277	2000	เลิก	0.5254
1945	ล้วน	0.3216	1973	ลึก	0.2351	2001	เลีย	1.0732
1946	ล่อ	0.2220	1974	ลืม	0.3358	2002	เลี้ยง	0.4097
1947	ล้อ	0.2154	1975	ลุก	0.2259	2003	เลี้ยงลูก	-0.0348
1948	ลอก	-0.0981	1976	ลุง	0.6585	2004	เลือก	0.8636
1949	ลอง	0.5783	1977	ลุ้น	1.0759	2005	เลือกตั้ง	0.8213
1950	ลองดู	0.5913	1978	ลุย	0.7613	2006	เลือด	0.1755
1951	ลอด	0.4835	1979	ลูก	0.0514	2007	เลื่อน	0.0449

No.	Word	SV	No.	Word	SV	No.	Word	SV
2008	แล	0.6012	2036	ว่ะ	0.2946	2064	วิธี	0.4436
2009	แลก	0.0738	2037	ว๊ะ	-0.0435	2065	วิธีการ	0.6122
2010	แล้ง	0.3001	2038	วัฒนธรรม	0.4290	2066	วิน	0.7925
2011	แลนด์	0.3756	2039	วัด	0.8297	2067	วินัย	0.8895
2012	แล้ว	0.3046	2040	วัน	0.1956	2068	วุ่นวาย	0.7618
2013	แล้วก็	0.1915	2041	วันที่	0.3858	2069	เวที	0.4862
2014	แล้วกัน	0.3736	2042	วันนั้น	0.0411	2070	เว้น	-0.0666
2015	แล้วจะ	0.1484	2043	วันนี้	0.4626	2071	เว้ย	0.5361
2016	แล้วแต่	0.7094	2044	วันละ	0.1531	2072	เวร	-0.1304
2017	แล้วไป	0.4234	2045	วัย	0.1996	2073	เวรกรรม	-0.6107
2018	แล้วยัง	0.2224	2046	วัว	0.0701	2074	เวลา	0.4527
2019	แล้วว่า	0.4021	2047	วา	0.6181	2075	เวอร์	0.9324
2020	และ	0.1119	2048	ว่า	0.4011	2076	เวียดนาม	0.6578
2021	โล	0.6090	2049	ว้า	0.3751	2077	โว้ย	0.4532
2022	โลก	0.6945	2050	ว่าการ	0.1002	2078	โวยวาย	-0.0253
2023	โลง	0.2677	2051	วาง	0.3757	2079	ไว	0.1232
2024	โลด	0.2296	2052	ว่าง	0.5662	2080	ไว้	0.2426
2025	โล้น	0.0195	2053	วางแผน	0.3708	2081	ไว้ใจ	-0.1593
2026	ไล่	0.5253	2054	ว่าไง	0.5946	2082	ไว้อาลัย	-0.9084
2027	ไลน์	0.3673	2055	ว่าแต่	0.6585	2083	ศพ	-0.2486
2028	ไล่ออก	0.1316	2056	ว่าที่	0.6585	2084	ศรัทธา	0.4940
2029	วก	0.3776	2057	วาม	0.0018	2085	ศักดิ์ศรี	0.9121
2030	วง	0.0543	2058	วาย	0.3608	2086	ศักย์	-0.0127
2031	วงการ	0.1779	2059	วิเคราะห์	0.4030	2087	ศาล	0.3458
2032	วน	0.4082	2060	วิ่ง	0.5283	2088	ศาสนา	0.2396
2033	วว	0.0531	2061	วิจารณ์	0.5050	2089	ศาสนาพุทธ	0.2453
2034	วอ	0.2724	2062	วิชา	0.3827	2090	ศิษย์	0.3618
2035	วะ	0.2942	2063	วิญญาณ	-0.5637	2091	ศีล	0.1820

No.	Word	SV	No.	Word	SV	No.	Word	SV
2092	ศึก	0.4223	2120	สนามบิน	-0.0592	2148	ส่วนตัว	0.6071
2093	ศึกษา	0.4072	2121	สนุก	0.5298	2149	ส่วนมาก	0.7065
2094	เศรษฐกิจ	0.8498	2122	สบาย	-0.0377	2150	ส่วนรวม	0.8403
2095	เศร้า	-0.8233	2123	สบายใจ	0.7159	2151	ส่วนใหญ่	0.6767
2096	เศร้าใจ	-1.0850	2124	สภา	-0.1537	2152	สวย	1.2426
2097	เศษ	0.3762	2125	สภาพ	-0.1583	2153	สวยงาม	1.2849
2098	สกปรก	0.5911	2126	สม	0.9019	2154	สวรรค์	0.0562
2099	ส่ง	0.1565	2127	สมควร	0.3567	2155	สวะ	-0.0066
2100	สงคราม	0.0702	2128	สมใจ	0.5971	2156	สวัสดี	0.5826
2101	สงฆ์	0.0315	2129	สมเด็จ	0.1410	2157	สว่าง	0.4865
2102	สงบ	0.7055	2130	สมน้ำหน้า	0.3394	2158	สส	0.7692
2103	สงบสุข	0.6185	2131	สมบัติ	0.5044	2159	สหรัฐ	0.4668
2104	สงสัย	0.4509	2132	สมอง	0.5606	2160	ส่อ	0.0564
2105	สงสาร	-0.9000	2133	สมัคร	0.4494	2161	สอง	0.3429
2106	ส่งเสริม	0.8217	2134	สมัย	0.6483	2162	สอน	0.8677
2107	สด	0.5484	2135	สมัยก่อน	0.3696	2163	สอบ	0.6344
2108	สติ	0.2031	2136	สมัยนี้	0.4009	2164	สอบซ่อม	-0.2281
2109	สถาน	-0.1582	2137	สมาคม	0.6018	2165	สอบสวน	-0.3721
2110	สถานการณ์	0.6251	2138	สยาม	0.4210	2166	สะใจ	0.5101
2111	สถานที่	0.3490	2139	สรรหา	-0.3414	2167	สะสม	0.2442
2112	สถาบัน	0.4153	2140	สร้าง	0.7415	2168	สะอาด	0.8063
2113	สน	0.8433	2141	สร้างภาพ	0.9294	2169	สัก	0.5674
2114	สั้น	-0.3758	2142	สรุป	0.1840	2170	สักที	0.3494
2115	สนใจ	1.1756	2143	สลาย	-0.7085	2171	สักนิด	0.0434
2116	สั้นตีน	0.2301	2144	สว	0.2003	2172	สักวัน	0.2278
2117	สนอง	-0.7218	2145	สวด	0.6153	2173	สั่ง	0.3418
2118	สนับสนุน	0.9576	2146	สวน	0.3759	2174	สังเกต	0.2050
2119	สนาม	0.8859	2147	ส่วน	0.1014	2175	สังคม	0.3877

No.	Word	SV	No.	Word	SV	No.	Word	SV
2176	สั่งสอน	0.1746	2204	สาว	1.0624	2232	สุดยอด	1.4345
2177	สัญญา	0.0569	2205	สาสม	-0.9203	2233	สุเทพ	0.4139
2178	สัด	-0.1411	2206	สาเหตุ	-0.1573	2234	สุนัข	0.4870
2179	สัตย์	-0.4889	2207	สำคัญ	0.6805	2235	สู่	-0.1818
2180	สัตว์	-0.0331	2208	สำนึก	-0.0005	2236	สู้	0.5051
2181	สัตว์นรก	-0.5804	2209	สำรวจ	0.2943	2237	สูง	0.7453
2182	สัน	0.0945	2210	สำเร็จ	0.7614	2238	สูงขึ้น	0.3572
2183	สั้น	0.2244	2211	สำหรับ	0.0860	2239	สูงสุด	0.2257
2184	สันดาน	0.1085	2212	สิ	0.6405	2240	สูญเสีย	-1.5741
2185	สัมภาษณ์	0.3286	2213	สิ่ง	0.4982	2241	สูบ	0.2394
2186	สา	0.2662	2214	สิทธิ	0.4696	2242	เสก	0.2217
2187	สากล	0.5921	2215	สิ้น	0.3499	2243	เส้น	0.4279
2188	สาขา	0.5028	2216	สินค้า	0.7609	2244	เสนอ	0.7217
2189	สาธารณะ	0.3147	2217	สิ้นดี	-0.3636	2245	เสนอหน้า	0.0216
2190	สาธุ	0.6401	2218	สินะ	0.4159	2246	เสพ	0.0443
2191	สาน	-0.0819	2219	สิบ	0.5217	2247	เสมอ	0.7609
2192	สาบ	-0.4886	2220	สี	0.6385	2248	เสร็จ	0.6104
2193	สาปแช่ง	-0.7260	2221	สึก	-0.0499	2249	เสร็จแล้ว	0.2861
2194	สาม	0.4409	2222	สืบ	0.2633	2250	เสริม	0.7677
2195	สามัคคี	0.6753	2223	สื่อ	0.3332	2251	เสรี	0.4552
2196	สามารถ	0.6360	2224	สู้	0.5761	2252	เสรีภาพ	0.4303
2197	สามี	0.2430	2225	สุข	-0.4879	2253	เสา	0.2738
2198	สาย	0.2998	2226	สุขภาพ	1.3752	2254	เสีย	-0.3358
2199	สายตา	0.3474	2227	สุคติ	-0.9447	2255	เสียง	0.7546
2200	สาร	-0.3260	2228	สุจริต	0.5331	2256	เสี่ยง	0.1713
2201	สารพัด	0.0336	2229	สุด	-0.0299	2257	เสียเงิน	0.3087
2202	สารเลว	-0.5440	2230	สุดซึ้ง	-1.3414	2258	เสียจัย	-1.1226
2203	สาระ	0.1732	2231	สุดท้าย	0.0055	2259	เสียใจ	-0.9000

No.	Word	SV	No.	Word	SV	No.	Word	SV
2260	เสียชีวิต	-1.0721	2287	หน่วยงาน	0.2512	2315	หมิ่น	0.2765
2261	เสียชื่อ	0.2382	2288	หนอ	0.0923	2316	หมื่น	0.6595
2262	เสียดาย	0.3140	2289	หน่อย	0.5745	2317	หมู	0.5142
2263	เสียที	0.3423	2290	หนัก	0.1488	2318	หมู่	0.3756
2264	เสียภาษี	-0.0219	2291	หนักแผ่นดิน	0.3465	2319	หมู่บ้าน	0.2697
2265	เสียเวลา	0.4069	2292	หนัง	0.2815	2320	หย่า	0.3934
2266	เสียสละ	0.3731	2293	หนังสือ	0.5029	2321	หยุด	0.2104
2267	เสียหาย	0.1416	2294	หนา	0.4002	2322	หรอ	0.2262
2268	เสือ	0.4927	2295	หน้า	0.3050	2323	หรอก	0.2050
2269	เสื้อ	0.5926	2296	หน้าด้าน	0.1389	2324	หรา	0.4622
2270	เสือก	0.6227	2297	หน้าตา	0.4914	2325	หรือ	0.3141
2271	เสื่อม	0.0490	2298	หน้าที่	0.2657	2326	หรือไง	0.2615
2272	แสดง	-1.0322	2299	หนาว	0.3147	2327	หรือเปล่า	0.2595
2273	แสดงความคิดเห็น	0.2094	2300	หนี	0.1339	2328	หรือไม่	0.5530
			2301	หนี้	0.5404	2329	หรือไม่ก็	0.1601
2274	แสดงว่า	0.5140	2302	หนึ่ง	0.2578	2330	หรือยัง	0.3532
2275	แสดงออก	0.7437	2303	หนุน	0.6985	2331	หรือว่า	0.4345
2276	แสน	0.2329	2304	หนุ่ม	0.5993	2332	หรู	0.8477
2277	แสนล้าน	-0.0297	2305	หนู	0.7050	2333	หลง	0.4497
2278	ใส	1.1095	2306	ห่ม	-0.1867	2334	หลบ	0.2806
2279	ใส่	0.6961	2307	หมด	0.4032	2335	หลวง	0.6218
2280	ใส่ใจ	0.6023	2308	หมดแล้ว	0.3183	2336	หลวงพ่อ	0.6930
2281	ใส่ร้าย	0.5271	2309	หมดอายุ	0.0437	2337	หล่อ	1.3379
2282	ไส้	0.2930	2310	หมอ	0.7974	2338	หลอก	0.6302
2283	หญ้า	0.2099	2311	หมา	0.3618	2339	หล่อน	0.4352
2284	หญิง	0.8363	2312	หมาก	0.2837	2340	หละ	0.5335
2285	หน	0.3378	2313	หมาย	0.4549	2341	หล่ะ	0.3881
2286	หน่วย	0.2116	2314	หมายถึง	0.6384	2342	หลัก	0.4167

No.	Word	SV	No.	Word	SV	No.	Word	SV
2343	หลักฐาน	0.5660	2371	ห่า	0.1942	2399	เหตุผล	0.2807
2344	หลัง	0.4100	2372	ห้า	0.5051	2400	เห็น	-0.0717
2345	หลังจาก	0.2575	2373	หาก	0.5551	2401	เห็นแก่	0.3258
2346	หลับ	-0.4045	2374	หากิน	0.4635	2402	เห็นแก่ตัว	0.1105
2347	หลาน	0.3154	2375	หาความ	0.2716	2403	เห็นใจ	-0.6288
2348	หลาย	0.6470	2376	หาง	0.3668	2404	เห็นด้วย	1.0454
2349	หลายครั้ง	0.0791	2377	ห้าง	0.2490	2405	เห็นภาพ	-0.8439
2350	หลายอย่าง	0.7252	2378	หาเงิน	0.4901	2406	เหนียว	0.8265
2351	หลุด	0.3286	2379	หาทาง	0.5151	2407	เหนือ	0.6262
2352	ห่วง	0.5853	2380	ห้าม	0.5358	2408	เหนื่อย	0.1526
2353	หวย	0.8196	2381	หาย	0.1979	2409	เห็บ	-0.3949
2354	ห่วย	0.1131	2382	หายใจ	-0.3006	2410	เหม็น	0.1056
2355	ห่วยแตก	0.1731	2383	หายไป	0.4300	2411	เหมา	0.0672
2356	หวัง	0.5561	2384	หาร	0.6112	2412	เหมาะ	0.7818
2357	หวังว่า	0.2358	2385	หาเรื่อง	0.3989	2413	เหมาะสม	1.4707
2358	หว่า	0.1463	2386	หาว่า	0.2827	2414	เหมือน	0.4568
2359	ห้อง	0.2179	2387	หำ	0.1404	2415	เหมือนกัน	0.1301
2360	หอย	0.3248	2388	หิว	-0.2020	2416	เหมือนกับ	0.0095
2361	หัก	0.3534	2389	หี	-0.4914	2417	เหมือนเดิม	0.8325
2362	หัด	0.4754	2390	หึ	0.2354	2418	เหยียบ	0.1545
2363	หันมา	0.7217	2391	หึหึ	0.1547	2419	เหยื่อ	-0.6174
2364	หั้ย	0.0349	2392	หุ่น	0.6792	2420	เหรอ	0.3260
2365	หัว	0.5512	2393	หุบปาก	0.0795	2421	เหรียญ	0.5444
2366	หัวใจ	-0.1959	2394	หู	0.3718	2422	เหล่	0.0462
2367	หัวเราะ	0.2190	2395	เห	0.1436	2423	เหล่า	0.0579
2368	หัวหน้า	0.3345	2396	เห็ด	0.2486	2424	เหล้า	0.0127
2369	หัวอก	-1.3696	2397	เหตุ	-0.3575	2425	เหล่านั้น	0.3297
2370	หา	0.5596	2398	เหตุการณ์	-0.9552	2426	เหล่านี้	0.3565

No.	Word	SV	No.	Word	SV	No.	Word	SV
2427	เหลิม	-0.2814	2455	ให้เป็น	0.6315	2483	อเมริกา	1.0313
2428	เหลี่ยม	0.3443	2456	ใหม่	0.5705	2484	อย่า	0.3486
2429	เหลือ	0.5131	2457	ให้อภัย	0.1146	2485	อยาก	0.3165
2430	เหลือเกิน	-0.1904	2458	ให้ออก	0.5078	2486	อยากจะ	0.0242
2431	เหลือง	0.6178	2459	ให้โอกาส	0.3775	2487	อยากได้	0.9005
2432	เห้อ	-0.3108	2460	ไห	0.5192	2488	อยากเป็น	0.1170
2433	เหอะ	0.3903	2461	ไห้	0.1296	2489	อยากรู้	0.4501
2434	เห่า	0.5989	2462	ไหน	0.1154	2490	อย่าง	-0.1159
2435	เห่าหอน	0.2277	2463	ไหม	0.5363	2491	อย่างงี้	0.6639
2436	เหี่ย	0.7202	2464	ไหล	-0.3617	2492	อย่างเดียว	0.3218
2437	เหี้ย	-0.9000	2465	ไหว	0.3800	2493	อย่างน้อย	0.6528
2438	แห	0.4297	2466	ไหว้	0.1493	2494	อย่างนั้น	0.4151
2439	แห่	0.5717	2467	อก	0.4660	2495	อย่างนี้	0.4180
2440	แหก	0.2188	2468	อคติ	0.6229	2496	อย่างยิ่ง	-0.0828
2441	แหกตา	-0.0086	2469	องค์	0.7714	2497	อย่างไร	0.0744
2442	แห่ง	0.5875	2470	องค์กร	0.1115	2498	อย่างอื่น	0.8436
2443	แหม	0.5208	2471	อด	0.1699	2499	อย่าได้	0.3724
2444	แหม่	0.5098	2472	อดตาย	0.2126	2500	อย่ามา	0.6161
2445	แหล	0.3839	2473	อดทน	0.3674	2501	อย่าลืม	0.7871
2446	แหละ	0.5623	2474	อดีต	0.4480	2502	อยู	0.5064
2447	โห	0.3035	2475	อธิบาย	0.0156	2503	อยู่	0.4187
2448	โหด	-0.2796	2476	อนาคต	0.5129	2504	อยู่ดี	0.2991
2449	โหดร้าย	-1.2901	2477	อนุโมทนา	0.9749	2505	อยู่ใน	0.4341
2450	โหน	-0.0694	2478	อบ	0.3345	2506	อร่อย	0.8862
2451	ให้	0.2507	2479	อบรม	0.4213	2507	อ้วก	0.0181
2452	ให้กำลังใจ	0.4137	2480	อภัย	0.1730	2508	อวด	0.5481
2453	ใหญ่	0.6004	2481	อภิสิทธิ์	0.1118	2509	อ้วน	0.4172
2454	ให้ท่า	0.8045	2482	อม	0.3966	2510	อวย	0.2852

No.	Word	SV	No.	Word	SV	No.	Word	SV
2511	ออ	0.2842	2539	อ่าน	-0.0371	2567	อุ	-0.0766
2512	อ้อ	0.4877	2540	อาย	0.5779	2568	อุดมการณ์	0.6843
2513	ออก	0.5045	2541	อ้าย	0.2285	2569	อุบัติเหตุ	-0.5591
2514	ออกจาก	0.3024	2542	อายุ	0.2008	2570	อุบาทว์	-0.5691
2515	ออกไป	0.6673	2543	อารมณ์	0.0256	2571	อุ้ม	0.1435
2516	ออกมา	0.3680	2544	อ้าว	0.1124	2572	เอ	0.6106
2517	อ่อน	-0.0707	2545	อาวุธ	0.6353	2573	เอกชน	0.0068
2518	อะ	0.5553	2546	อาศัย	-0.0370	2574	เอกสาร	-0.0886
2519	อ่ะ	0.4708	2547	อาหาร	0.8697	2575	เอง	0.3829
2520	อะรัย	0.0363	2548	อำ	0.4225	2576	เอ็ง	0.2755
2521	อะไร	0.1030	2549	อำนาจ	0.8724	2577	เอย	0.3326
2522	อะไรบ้าง	0.1904	2550	อิจฉา	0.8521	2578	เอ่ย	-0.0086
2523	อังกฤษ	0.8753	2551	อิ่ม	0.5388	2579	เอ้ย	-0.0060
2524	อัด	0.1466	2552	อิสระ	0.0235	2580	เอ๊ย	-0.0570
2525	อัน	0.2248	2553	อิอิ	1.3504	2581	เอ๋ย	0.0886
2526	อันตราย	0.1901	2554	อี	0.3229	2582	เออ	0.7462
2527	อันนี้	0.5959	2555	อีก	0.2289	2583	เอ่อ	0.4339
2528	อา	0.2751	2556	อีกคน	0.2173	2584	เอ๋อ	0.3876
2529	อ่า	0.3432	2557	อีกครั้ง	0.4275	2585	เอะอะ	0.3402
2530	อาการ	0.0095	2558	อีกต่อไป	0.1301	2586	เอา	0.5380
2531	อากาศ	0.1906	2559	อีกที	0.2941	2587	เอ้า	0.3083
2532	อ้าง	0.6208	2560	อีกไม่นาน	0.0831	2588	เอาความ	0.3686
2533	อ้างว่า	0.0915	2561	อีกแล้ว	-0.0458	2589	เอาคืน	0.1344
2534	อาจ	0.2749	2562	อีกหน่อย	0.3263	2590	เอาจริง	0.4242
2535	อาจจะ	0.3894	2563	อีด	-0.0136	2591	เอาใจ	0.6929
2536	อาจารย์	0.6362	2564	อีสาน	0.8897	2592	เอาใจช่วย	0.8211
2537	อาชีพ	0.6380	2565	อื่น	0.7476	2593	เอาดี	0.1656
2538	อาทิตย์	0.5547	2566	อืม	0.3978	2594	เอาแต่	0.2601

No.	Word	SV
2595	เอาเปรียบ	0.4128
2596	เอาผิด	0.0053
2597	เอามา	0.7080
2598	เอาเรื่อง	-0.0176
2599	เอาเลย	0.0824
2600	เอาไว้	0.5436
2601	เอาหน้า	0.3917
2602	เอาเอง	0.3423
2603	เอียง	0.0858
2604	แอด	0.0356
2605	แอบ	0.6133
2606	แอร์	0.3908
2607	โอ	0.7485
2608	โอ้	0.3009
2609	โอกาส	0.3861
2610	โอเค	0.8292
2611	โอ้ย	-0.1890
2612	ไอ	0.0840
2613	ไอ้	-0.0016
2614	ไอ้ชาติ	-0.5676
2615	ไอ้บ้า	-0.4081
2616	ไอ้เวร	-0.4075
2617	ฮะ	0.4089
2618	ฮา	0.8011
2619	ฮ่า	0.6022
2620	ฮีโร่	0.7448
2621	เฮง	0.7353
2622	เฮงซวย	-0.0789

No.	Word	SV
2623	เฮ้ย	0.0978
2624	เฮ้อ	-0.1224
2625	เฮีย	0.3962
2626	เฮี้ย	0.1744

Appendix E

News Article Sentiment Dictionary

The news article sentiment dictionary that constructed by SO-PMI-IR algorithm with "NEAR" operation from 172,189 news articles (Approximately 85,000,000 words).

Polar: Polarity value

P: Positive

N: Negative

NE: Neutral

No.	Word	Polar	No.	Word	Polar	No.	Word	Polar
1	กกต	N	28	กระแสข่าว	N	55	การกระทำ	N
2	กค	N	29	กรุงเทพ	N	56	การกำ	N
3	กฎหมาย	N	30	กรุงเทพ มหานคร	N	57	การแก้ปัญหา	N
4	กฎหมาย	N				58	การแข่งขัน	NE
5	กดดัน	N	31	กล้องวงจรปิด	N	59	การควบ	N
6	กติกา	N	32	กลับบ้าน	N	60	การค้า	N
7	กปปส	N	33	กล้า	N	61	การเคลื่อนไหว	N
8	กพ	N	34	กล่าวถึง	N	62	การจับกุม	N
9	กม	N	35	กล่าวว่า	N	63	การใช้	N
10	กย	N	36	กล่าวหา	N	64	การณ์	N
11	กรกฎาคม	N	37	กลาโหม	N	65	การเดินทาง	N
12	กรณี	N	38	กว่า	P	66	การต่อสู้	N
13	กรณีที่	N	39	กว้าง	N	67	การท่องเที่ยว	N
14	กรม	N	40	กอง	N	68	การนำ	N
15	กรรม	N	41	กองกำลัง	N	69	การบริหาร	N
16	กรรมการ	N	42	กองทัพ	N	70	การปกครอง	N
17	กรอบ	N	43	กองทัพบก	N	71	การปฏิบัติ	N
18	กระตุ้น	N	44	ก่อนที่	N	72	การปฏิรูป	N
19	กระทบ	N	45	ก่อนที่จะ	N	73	การประชุม	N
20	กระทรวง	N	46	ก่อนหน้า	N	74	การป้องกัน	N
21	กระทรวง มหาดไทย	N	47	ก่อนหน้านี้	N	75	การ เปลี่ยนแปลง	N
			48	ก่อสร้าง	N			
22	กระทั่ง	N	49	ก่อเหตุ	NE	76	การผลิต	N
23	กระทำ	N	50	ก่อให้เกิด	N	77	การพัฒนา	N
24	กระทำผิด	N	51	กังวล	N	78	การพิจารณา	N
25	กระบวนการ	N	52	กัน	P	79	การพูด	N
26	กระเป๋า	N	53	กันยายน	NE	80	การมี	N
27	กระสุน	N	54	กา	N	81	การรักษา	N

No.	Word	Polar
82	การลงทุน	N
83	การศึกษา	N
84	การสร้าง	N
85	การแสดง	N
86	การหารือ	N
87	ก้าว	N
88	กำลังใจ	N
89	กำหนด	N
90	กำหนดให้	N
91	กิจ	N
92	กิจการ	NE
93	กิโลเมตร	N
94	กุม	N
95	กุมภาพันธ์	N
96	กุล	N
97	กู	N
98	กู้ภัย	N
99	เก	N
100	เก็บ	N
101	เกม	P
102	เกรง	N
103	เกษตร	N
104	เก่า	P
105	เกาะ	N
106	เกิดเหตุ	N
107	เกิน	N
108	เกียรติ	N
109	เกี่ยวกับ	N
110	เกี่ยวข้อง	N
111	เกือบ	N
112	แก่	N
113	แก้	NE
114	แกนนำ	N
115	แก้ปัญหา	P
116	แกรม	N
117	แก้ว	N
118	โก	N
119	ใกล้	NE
120	ใกล้เคียง	N
121	ไกล	N
122	ขณะ	N
123	ขณะเดียวกัน	N
124	ขน	N
125	ขนาดใหญ่	N
126	ขบวน	N
127	ขบวนการ	N
128	ข่มขู่	N
129	ขยาย	N
130	ขยายผล	N
131	ขวา	N
132	ข้อ	N
133	ข้อกล่าวหา	N
134	ขอขอบคุณ	N
135	ข้อความ	N
136	ขอความร่วมมือ	N
137	ของกลาง	N
138	ของกอง	N
139	ข้อเท็จจริง	N
140	ขอโทษ	N
141	ข้อมูล	N
142	ข้อเสนอ	NE
143	ข้อหา	N
144	ขัดแย้ง	N
145	ขั้น	N
146	ขั้นตอน	N
147	ขับ	N
148	ขับเคลื่อน	N
149	ขับรถ	N
150	ขา	NE
151	ข้าง	N
152	ข้างหน้า	N
153	ขาด	N
154	ข้าม	N
155	ขาย	P
156	ข้าราชการ	N
157	ขาว	N
158	ข่าว	N
159	ข้าว	N
160	ข่าวสด	N
161	ขี่	N
162	ขึ้น	P
163	ขึ้นไป	N
164	ขึ้นรถ	N

No.	Word	Polar
165	ขึ้นอยู่กับ	N
166	เขต	N
167	เข้าข่าย	N
168	เข้าที่	N
169	เข้าไป	N
170	เข้าพบ	NE
171	เข้าร่วม	N
172	เขียน	NE
173	แข่งขัน	N
174	แข็งแรง	N
175	แขน	N
176	แขวง	N
177	ค	N
178	คณะ	N
179	คณะกรรมการ	N
180	คณะทำงาน	N
181	คณะรักษาความสงบแห่งชาติ	N
182	คดี	N
183	คนขับ	N
184	คนดัง	N
185	คนดี	N
186	คนร้าย	N
187	คนละ	N
188	คนี้	N
189	คร	N
190	ครบ	N
191	ครม	N
192	ครอบครอง	N
193	ครั้ง	N
194	ครั้งที่	N
195	ครั้งนี้	N
196	ครับ	P
197	ครึ่ง	N
198	คลอง	N
199	คลิป	N
200	คลื่น	N
201	ควบคุม	N
202	ควบคุมตัว	N
203	ควร	P
204	ควรจะ	N
205	คว้า	N
206	ความเข้าใจ	N
207	ความคิด	N
208	ความคิดเห็น	N
209	ความคืบหน้า	N
210	ความเคลื่อนไหว	N
211	ความจริง	N
212	ความช่วยเหลือ	N
213	ความเชื่อมั่น	N
214	ความเดือดร้อน	N
215	ความต้องการ	N
216	ความปลอดภัย	N
217	ความเป็นธรรม	N
218	ความผิด	N
219	ความพยายาม	N
220	ความพร้อม	N
221	ความมั่นคง	N
222	ความร่วมมือ	N
223	ความรัก	NE
224	ความรับผิดชอบ	N
225	ความรุนแรง	N
226	ความรู้	N
227	ความรู้สึก	N
228	ความเร็ว	N
229	ความสงบ	N
230	ความสงบเรียบร้อย	N
231	ความสัมพันธ์	N
232	ความสามารถ	N
233	ความสำคัญ	N
234	ความสุข	NE
235	ความเสียหาย	N
236	คส	N
237	คสช	N
238	คอ	N
239	ค่อย	N
240	ค่ะ	P
241	คะแนน	N
242	คัดค้าน	N
243	คา	N

No.	Word	Polar	No.	Word	Polar	No.	Word	Polar
244	ค่า	N	272	แค่	P	300	จันทร์	N
245	ค้า	N	273	โครงการ	N	301	จับ	N
246	ค่าใช้จ่าย	NE	274	โครงสร้าง	N	302	จับกุม	N
247	คาด	N	275	ฆ่า	N	303	จา	N
248	ค่าย	N	276	โฆษก	N	304	จากนั้น	N
249	ค่ำ	N	277	ง	N	305	จากนั้นก็	N
250	คำตอบ	N	278	งด	N	306	จากนี้	N
251	คำถาม	N	279	งบ	N	307	จ่ายเงิน	N
252	คำพูด	N	280	งบประมาณ	N	308	จำ	N
253	คำสั่ง	N	281	งาน	P	309	จำกัด	N
254	คิด	P	282	ง่าย	P	310	จำคุก	N
255	คืน	N	283	เงินสด	N	311	จำนวนมาก	N
256	คุณภาพ	N	284	เงื่อนไข	N	312	จำนวนหนึ่ง	N
257	คุม	N	285	จ	N	313	จำเป็นต้อง	N
258	คุ้มครอง	N	286	จนกระทั่ง	N	314	จำเลย	N
259	คุมตัว	N	287	จนกว่า	N	315	จำหน่าย	N
260	คุย	P	288	จนได้	N	316	จิต	N
261	คู่	NE	289	จนเป็น	N	317	จิตใจ	N
262	เค	N	290	จนมี	N	318	จี	N
263	เคย	N	291	จม	N	319	จีน	NE
264	เครียด	N	292	จยย	N	320	จุด	N
265	เครือข่าย	N	293	จราจร	N	321	เจตนา	N
266	เครื่อง	N	294	จอด	N	322	เจรจา	N
267	เครื่องบิน	N	295	จักร	N	323	เจริญ	N
268	เครื่องมือ	N	296	จังหวะ	N	324	เจ้า	N
269	เคลื่อน	N	297	จังหวัด	N	325	เจ้าของ	N
270	เคลื่อนไหว	N	298	จัดกิจกรรม	N	326	เจ้าตัว	N
271	เคารพ	NE	299	จัดทำ	N	327	เจ้าหน้าที่	N

No.	Word	Polar
328	เจ้าหน้าที่ตำรวจ	N
329	เจ้าอาวาส	N
330	แจก	N
331	แจ้ง	N
332	แจ้งความ	N
333	แจ้งว่า	N
334	โจมตี	NE
335	ใจ	P
336	ฉายา	N
337	เฉพาะ	N
338	ช	N
339	ชน	N
340	ชนะ	N
341	ชนิด	N
342	ชลบุรี	N
343	ช่วงนี้	NE
344	ช่วงเวลา	N
345	ชวน	N
346	ช่วยเหลือ	N
347	ช่องทาง	N
348	ชอบ	P
349	ชัด	N
350	ชั้น	N
351	ชัย	N
352	ชั่วคราว	N
353	ชั่วโมง	N
354	ชา	N
355	ช้า	N
356	ชาติ	P
357	ชาย	N
358	ชาวต่างชาติ	N
359	ชาวบ้าน	N
360	ชิ	N
361	ชี้	N
362	ชี้แจง	N
363	ชื่นชม	N
364	ชื่อดัง	NE
365	ชุด	N
366	ชุดสืบสวน	N
367	ชุมชน	N
368	ชุมนุม	N
369	ชู	N
370	เช่นเดียวกับ	N
371	เช่นนี้	N
372	เชิง	N
373	เชิญ	N
374	เชื่อมโยง	N
375	แชร์	N
376	โชคดี	N
377	โชว์	N
378	ใช้	P
379	ใช้เวลา	N
380	ใช้อำนาจ	N
381	ไชย	N
382	ซา	N
383	ซ้าย	N
384	ซ้ำ	N
385	ซิ	N
386	ซี	N
387	ซื้อ	P
388	โซ	N
389	โซเชียล	N
390	ญาติ	N
391	ญี่ปุ่น	N
392	ฐาน	N
393	ณ	N
394	ณรงค์	N
395	ด	N
396	ดร	N
397	ด่วน	N
398	ด้วยกัน	N
399	ด้วยว่า	N
400	ดัง	P
401	ดังกล่าว	N
402	ดังนี้	N
403	ดับ	N
404	ดา	N
405	ด่าน	N
406	ด้านหน้า	N
407	ด้านหลัง	N
408	ดารา	N
409	ดำ	N
410	ดำเนิน	N

No.	Word	Polar	No.	Word	Polar	No.	Word	Polar
411	ดำเนินการ	N	439	ได้อีก	N	467	ต่อว่า	N
412	ดำเนินคดี	N	440	ต	N	468	ต่อสู้	N
413	ดำเนินงาน	N	441	ตก	NE	469	ตั้งใจ	N
414	ดำรงตำแหน่ง	N	442	ตกใจ	N	470	ตั้งอยู่	N
415	ดิน	N	443	ตกเป็น	N	471	ตัดสินใจ	NE
416	ดี	P	444	ตกลง	N	472	ตัน	N
417	ดีกว่า	P	445	ตค	N	473	ตัวแทน	N
418	ดีขึ้น	NE	446	ตท	N	474	ตัวผู้	N
419	ดีใจ	N	447	ต้น	N	475	ตัวเลข	NE
420	ดึง	N	448	ตร	N	476	ตัวเอง	P
421	ดู	P	449	ตรง	N	477	ตา	N
422	เด	N	450	ตรงกัน	N	478	ต่างชาติ	N
423	เดช	N	451	ตรงกับ	N	479	ต่างประเทศ	N
424	เดลินิวส์	N	452	ตรวจ	N	480	ตาม	N
425	เดินหน้า	N	453	ตรวจค้น	N	481	ตามกฎหมาย	N
426	แดง	NE	454	ตรวจสอบ	N	482	ตามที่	N
427	โดน	N	455	ตรา	N	483	ตามปกติ	N
428	โดยจะ	N	456	ตลอด	N	484	ตามมา	N
429	โดยเฉพาะ	NE	457	ตลอดจน	N	485	ตาย	N
430	โดยเร็ว	N	458	ตลาด	NE	486	ต่ำ	N
431	ได้แก่	N	459	ตห	N	487	ตำบล	N
432	ได้ดี	N	460	ตอ	N	488	ตำรวจ	N
433	ได้ที่	N	461	ต้องหา	N	489	ตำแหน่ง	NE
434	ได้มา	N	462	ต่อต้าน	N	490	ติ	N
435	ได้มีการ	N	463	ต้อนรับ	N	491	ติดต่อ	N
436	ได้ยิน	N	464	ต่อเนื่อง	N	492	ติดตาม	N
437	ได้รับแจ้ง	N	465	ต่อไป	N	493	ตี	N
438	ได้รับบาดเจ็บ	N	466	ต่อมา	N	494	ตุลาคม	N

No.	Word	Polar	No.	Word	Polar	No.	Word	Polar
495	ตู้	N	523	ทน	N	551	ทั่วโลก	N
496	เตรียม	N	524	ทนายความ	N	552	ทา	N
497	แตก	N	525	ทบ	N	553	ทาง	N
498	แตกต่าง	N	526	ทบทวน	N	554	ทางการเมือง	NE
499	แต่กลับ	N	527	ทยอย	N	555	ทางด้าน	N
500	แต่ง	N	528	ทรัพย์	N	556	ท่าน	P
501	แต่งตั้ง	N	529	ทรัพย์สิน	N	557	ท่ามกลาง	N
502	แต่ละ	N	530	ทราบ	N	558	ท้าย	N
503	แต่อย่างใด	NE	531	ทราบชื่อ	N	559	ทำการ	N
504	โตโยต้า	N	532	ทราบว่า	N	560	ทำความเข้าใจ	N
505	ถ	N	533	ทหาร	N	561	ทำตาม	N
506	ถล่ม	N	534	ทอง	N	562	ทำบุญ	N
507	ถวาย	N	535	ท้องถิ่น	N	563	ทำไม	N
508	ถาม	NE	536	ท่องเที่ยว	N	564	ทำร้าย	N
509	ถามถึง	N	537	ทะเบียน	N	565	ทำหน้าที่	NE
510	ถ่ายรูป	N	538	ทะเล	N	566	ทำอะไร	N
511	ถาวร	N	539	ทะเลาะ	NE	567	ทิพย์	N
512	ถึงขั้น	N	540	ทั้งที่	N	568	ทิศทาง	N
513	ถือ	N	541	ทั้งนี้	N	569	ที่เกิดขึ้น	N
514	ถือเป็น	N	542	ทั้งสอง	N	570	ที่เกิดเหตุ	N
515	ถูก	P	543	ทั้งสิ้น	N	571	ที่เกี่ยวข้องกับ	N
516	ถูกจับ	N	544	ทั้งหมด	N	572	ที่จะ	P
517	ถูกต้อง	N	545	ทันที	N	573	ที่ดิน	NE
518	แถม	N	546	ทับ	NE	574	ทีเดียว	N
519	แถลง	N	547	ทัพ	N	575	ที่ตั้ง	N
520	แถลงการณ์	N	548	ทั่ว	N	576	ที่ถูก	N
521	แถลงข่าว	N	549	ทั่วประเทศ	N	577	ที่ทาง	N
522	ท	N	550	ทั่วไป	P	578	ที่แท้จริง	N

No.	Word	Polar
579	ที่นั่ง	N
580	ที่ไป	N
581	ที่ผิด	N
582	ทีมงาน	N
583	ที่วัด	N
584	ทีวี	NE
585	ที่สุดใน	NE
586	ที่เหลือ	N
587	ที่อยู่	N
588	ทุกวันนี้	N
589	ทุจริต	N
590	เท	N
591	เทคโนโลยี	N
592	เท่า	N
593	เท่ากับ	N
594	เท่าที่	N
595	เท่านั้น	N
596	เทียบ	NE
597	เที่ยว	N
598	แทบ	N
599	โทร	N
600	โทรศัพท์	N
601	โทษ	N
602	ไทย	P
603	ธ	NE
604	ธนาคาร	N
605	ธรรม	N
606	ธรรมชาติ	N
607	ธิ	N
608	ธุรกิจ	N
609	น	N
610	นก	N
611	นท	N
612	นนท์	N
613	นปช	N
614	นพ	N
615	นโยบาย	N
616	นวันที่	N
617	นส	N
618	นอก	N
619	นอกจากนี้ยัง	N
620	นอง	N
621	น้อง	NE
622	นอน	N
623	น้อย	NE
624	นะ	P
625	นัก	P
626	นักร้อง	N
627	นักเรียน	N
628	นักวิชาการ	N
629	นักศึกษา	N
630	นักแสดง	N
631	นั่ง	N
632	นัด	N
633	นั่น	N
634	นันท์	N
635	นา	N
636	นางสาว	N
637	น่าจะเป็น	N
638	นาที	N
639	นาม	N
640	นาย	N
641	นายสุเทพ เทือกสุบรรณ	N
642	น้ำ	N
643	นำเข้า	N
644	นำตัว	N
645	นำไปสู่	N
646	น้ำมัน	N
647	น้ำหนัก	N
648	นิ	N
649	นิยม	N
650	นี	N
651	นุ	N
652	เน	N
653	เน็ต	N
654	เน้น	N
655	เนียบ	N
656	แน่	N
657	แน่นอน	N
658	แนว	N
659	แนวคิด	N
660	แนวทาง	N
661	แนะนำ	NE

No.	Word	Polar
662	โน	N
663	ในขณะที่	N
664	ในขณะนี้	N
665	ในช่วง	N
666	ในทาง	N
667	ในที่สุด	N
668	ในบ้าน	N
669	ในประเทศ	N
670	ในปัจจุบัน	N
671	ในพื้นที่	N
672	ในเมือง	N
673	ในโลก	N
674	ในอนาคต	N
675	บ	N
676	บก	N
677	บท	N
678	บทบาท	N
679	บรรจุ	N
680	บรรยากาศ	N
681	บริษัท	N
682	บ่อย	P
683	บัญชี	N
684	บัตร	NE
685	บันทึก	N
686	บา	N
687	บาดเจ็บ	N
688	บาดแผล	N
689	บ้านพัก	N
690	บ่าย	N
691	บิ๊ก	N
692	บิน	N
693	บี	N
694	บุ	N
695	บุก	N
696	บุ๊ก	N
697	บุคคล	N
698	บุญ	N
699	บุตร	N
700	เบาะแส	N
701	เบื้องต้น	N
702	แบ่ง	N
703	แบ่งเป็น	N
704	แบบ	P
705	ใบ	N
706	ใบหน้า	N
707	ป	N
708	ปกครอง	N
709	ปกติ	P
710	ปกป้อง	N
711	ปฏิบัติ	N
712	ปฏิบัติการ	N
713	ปฏิบัติตาม	N
714	ปฏิเสธ	N
715	ปทุมธานี	N
716	ปป	N
717	ปปช	N
718	ปรองดอง	N
719	ประ	N
720	ประกอบ	N
721	ประกอบด้วย	N
722	ประการ	N
723	ประกาศ	N
724	ประจำ	N
725	ประจำปี	N
726	ประชา	N
727	ประชาธิปัตย์	N
728	ประชาสัมพันธ์	N
729	ประชุม	N
730	ประเด็น	NE
731	ประตู	N
732	ประเทศ	P
733	ประเทศจีน	N
734	ประเทศชาติ	N
735	ประเภท	N
736	ประมาณ	N
737	ประเมิน	N
738	ประโยชน์	N
739	ประวัติ	N
740	ประวัติศาสตร์	N
741	ประวิตร วงษ์สุวรรณ	N
742	ประสบการณ์	N
743	ประสบความสำเร็จ	N

No.	Word	Polar	No.	Word	Polar	No.	Word	Polar
744	ประสาน	N	772	เป็นประโยชน์	N	800	ผบช	N
745	ประสานงาน	N	773	เป็นปัญหา	N	801	ผล	N
746	ปรับ	N	774	เป็นไป	N	802	ผลกระทบ	N
747	ปรับปรุง	N	775	เป็นไปได้	N	803	ผลงาน	NE
748	ปรากฏ	N	776	เป็นไปตาม	N	804	ผลประโยชน์	N
749	ปรากฏว่า	N	777	เป็นผู้	N	805	ผลักดัน	N
750	ปราบปราม	N	778	เป็นระยะ	N	806	ผลิต	N
751	ปริมาณ	N	779	เป็นหลัก	N	807	ผอ	N
752	ปรึกษา	N	780	เป็นห่วง	N	808	ผ่าน	P
753	ปล่อย	N	781	เป็นเหตุให้	N	809	ผิด	N
754	ปลา	NE	782	เป็นอย่างดี	N	810	ผิดกฎหมาย	N
755	ป่วย	N	783	เปรียบเทียบ	N	811	ผู้คน	N
756	ปอ	N	784	เปลี่ยน	P	812	ผู้จัดการ	N
757	ป้องกัน	N	785	เปลี่ยนแปลง	N	813	ผู้ช่วย	N
758	ปัจจัย	N	786	เปอร์เซ็นต์	N	814	ผู้ชุมนุม	N
759	ปัญหา	P	787	เป้าหมาย	N	815	ผู้ใช้	N
760	ป่า	N	788	เปิดตัว	N	816	ผู้โดยสาร	N
761	ปาก	N	789	เปิดเผย	N	817	ผู้ต้องหา	N
762	ปิด	N	790	เปิดเผยว่า	N	818	ผู้ตาย	N
763	ปืน	N	791	ไปจน	NE	819	ผู้นำ	N
764	เป็น	P	792	ไปดู	N	820	ผู้บริหาร	N
765	เป็นข่าว	N	793	ไปที่	N	821	ผู้บังคับบัญชา	N
766	เป็นความ	NE	794	ไปมา	N	822	ผู้บัญชาการ	N
767	เป็นเจ้าของ	N	795	ไปยัง	N	823	ผู้ประกอบการ	N
768	เป็นช่วง	N	796	ไปหา	N	824	ผู้ป่วย	N
769	เป็นต้น	N	797	ผ	N	825	ผู้สื่อข่าว	N
770	เป็นต้นไป	N	798	ผบ	N	826	ผู้เสียชีวิต	N
771	เป็นตัว	N	799	ผบก	N	827	ผู้เสียหาย	N

No.	Word	Polar
828	ผู้ใหญ่	NE
829	ผู้อำนวยการ	N
830	ผู้อื่น	N
831	เผยแพร่	N
832	แผน	N
833	ฝน	N
834	ฝนตก	N
835	ฝั่ง	N
836	ฝ่าย	N
837	ฝ่ายสืบสวน	N
838	เฝ้า	N
839	พ	N
840	พค	N
841	พง	N
842	พงศ์	N
843	พนักงาน	N
844	พนักงานสอบสวน	N
845	พบ	N
846	พย	NE
847	พยาน	N
848	พยานหลักฐาน	N
849	พร	N
850	พรบ	N
851	พรรคการเมือง	N
852	พรรคประชาธิปัตย์	N
853	พรรคเพื่อไทย	N
854	พร้อม	N
855	พร้อมกับ	N
856	พร้อมด้วย	N
857	พร้อมทั้ง	N
858	พระ	N
859	พระองค์	N
860	พระเอก	N
861	พฤติกรรม	NE
862	พฤษภาคม	N
863	พล	N
864	พลเมือง	N
865	พลัง	N
866	พลาด	N
867	พวกเรา	N
868	พศ	N
869	พ่อแม่	N
870	พอสมควร	N
871	พัง	N
872	พัฒนา	NE
873	พันธุ์	N
874	พิ	N
875	พิจารณา	N
876	พิทักษ์	N
877	พิธี	N
878	พิเศษ	N
879	พิสูจน์	N
880	พี่	NE
881	พื้น	N
882	พื้นฐาน	N
883	พุ่ม	N
884	พูดคุย	N
885	เพจ	N
886	เพชร	N
887	เพราะฉะนั้น	N
888	เพิ่ม	N
889	เพิ่มขึ้น	N
890	เพิ่มเติม	N
891	เพียงแค่	N
892	เพียงแต่	N
893	เพียงพอ	N
894	เพื่อนบ้าน	N
895	เพื่อไม่ให้	N
896	แพ้	N
897	แพง	N
898	แพทย์	N
899	แพร่	N
900	โพสต์	N
901	ฟ	N
902	ฟ้อง	N
903	ฟ้า	N
904	เฟซ	N
905	แฟนคลับ	N
906	ไฟ	NE
907	ภ	N
908	ภรรยา	N
909	ภัย	N

No.	Word	Polar	No.	Word	Polar	No.	Word	Polar
910	ภา	N	938	มอบ	N	966	มีส่วน	N
911	ภาคใต้	N	939	มอบหมาย	N	967	มีส่วนร่วม	N
912	ภาครัฐ	N	940	มอบให้	N	968	มีหน้า	N
913	ภาคเหนือ	N	941	มัก	N	969	มีอำนาจ	N
914	ภาพ	N	942	มักจะ	NE	970	มีโอกาส	NE
915	ภาพข่าว	N	943	มัน	P	971	มุ่ง	N
916	ภาพประกอบ	N	944	มาก	P	972	มุ่งหน้า	N
917	ภาพรวม	N	945	มากมาย	N	973	มุม	N
918	ภายใต้	N	946	มาก่อน	N	974	มูลค่า	N
919	ภายใน	N	947	มาต	N	975	มูลนิธิ	N
920	ภายหลัง	N	948	มาตร	N	976	เม	N
921	ภารกิจ	N	949	มาตรการ	N	977	เมตร	N
922	ภาวะ	N	950	มาตรา	N	978	เมีย	NE
923	ภาษี	N	951	มายัง	N	979	เมือง	N
924	ภูมิภาค	N	952	มาเลเซีย	N	980	เมื่อวันที่	N
925	ม	N	953	มาแล้ว	N	981	เมื่อเวลา	N
926	มกราคม	N	954	มิ	N	982	แม่	N
927	มค	N	955	มีความผิด	N	983	แม้จะ	N
928	มงคล	N	956	มีความสุข	N	984	แม้แต่	N
929	มติ	N	957	มีเจ้า	N	985	แม้ว่า	N
930	มนุษย์	N	958	มีด	N	986	โม	N
931	มม	N	959	มีเดีย	N	987	ไม้	N
932	ม่วง	N	960	มีปัญหา	P	988	ไม่กี่	N
933	มวลชน	NE	961	มีผล	NE	989	ไม่เกิน	N
934	มหา	N	962	มีเพี	N	990	ไม่ค่อย	P
935	มหาชน	N	963	มีมติ	N	991	ไม่ได้รับ	N
936	มหาวิทยาลัย	N	964	มีเรื่อง	N	992	ไม่ต้อง	P
937	มองว่า	P	965	มีลูก	N	993	ไม่ต้องการ	N

No.	Word	Polar
994	ไม่ต่ำกว่า	N
995	ไม่ทราบ	N
996	ไม่ทัน	N
997	ไม่ทำ	N
998	ไม่น่า	N
999	ไม่พบ	N
1000	ไม่พอใจ	N
1001	ไม่มีใคร	N
1002	ไม่รับ	N
1003	ไม่ว่า	N
1004	ไม่ว่าจะเป็น	N
1005	ไม่เห็น	N
1006	ไม่เห็นด้วย	N
1007	ไม่เหมาะสม	N
1008	ไม่ให้	NE
1009	ไม่อยู่	N
1010	ย	N
1011	ยกเลิก	N
1012	ยง	N
1013	ยนต์	N
1014	ยม	N
1015	ยศ	N
1016	ยอด	N
1017	ยอม	N
1018	ย่อม	N
1019	ยอมรับ	NE
1020	ยะ	N
1021	ยังคง	N
1022	ยังไง	N
1023	ยั่งยืน	N
1024	ยัน	N
1025	ยา	N
1026	ยาง	N
1027	ย่าน	N
1028	ยาว	N
1029	ยาเสพติด	N
1030	ย้ำ	N
1031	ยิง	N
1032	ยิ่งลักษณ์ ชินวัตร	N
1033	ยินดี	N
1034	ยี่ห้อ	N
1035	ยึด	N
1036	ยืน	N
1037	ยื่น	N
1038	ยืนยัน	N
1039	ยุค	N
1040	ยุติ	N
1041	ยุติธรรม	N
1042	ยุทธศาสตร์	N
1043	ยู	N
1044	เย็น	N
1045	เยอะ	P
1046	เยี่ยม	N
1047	แยก	N
1048	โย	N
1049	ร	N
1050	รณรงค์	N
1051	รถ	N
1052	รถกระบะ	N
1053	รถจักรยานยนต์	N
1054	รถยนต์	N
1055	รบ	N
1056	รพ	N
1057	รม	N
1058	รมว	N
1059	รร	N
1060	รวบรวม	N
1061	รวม	N
1062	ร่วมกัน	N
1063	ร่วมกับ	N
1064	ร่วมงาน	N
1065	ร่วมมือ	N
1066	รส	N
1067	รอง	N
1068	ร้อง	N
1069	รองนายกรัฐมนตรี	N
1070	รองประธาน	N
1071	ร่องรอย	N
1072	ร้องเรียน	N
1073	ร้อน	N
1074	รอย	N

No.	Word	Polar
1075	ร้อยเวร	N
1076	ระงับ	N
1077	ระดม	N
1078	ระดับสูง	N
1079	ระบบ	NE
1080	ระบอบประชาธิปไตย	N
1081	ระบุ	N
1082	ระบุว่า	N
1083	ระเบิด	N
1084	ระเบียบ	N
1085	ระมัดระวัง	N
1086	ระยะ	N
1087	ระยะเวลา	N
1088	ระวัง	N
1089	รัก	P
1090	รักษ์	N
1091	รักษาการ	N
1092	รักษาตัว	N
1093	รัฐ	N
1094	รัฐธรรมนูญ	P
1095	รัฐประหาร	N
1096	รัฐมนตรี	N
1097	รัฐสภา	N
1098	รัตน์	N
1099	รัตน	N
1100	รับแจ้ง	N
1101	รับทราบ	N
1102	รับประทาน	N
1103	รับผิดชอบ	N
1104	รับรอง	N
1105	รับรู้	N
1106	รับสารภาพ	N
1107	รา	N
1108	ราคา	NE
1109	ร่างกาย	N
1110	ร่างรัฐธรรมนูญ	NE
1111	ราช	N
1112	ราชการ	N
1113	ราย	N
1114	รายการ	N
1115	รายงาน	N
1116	รายงานข่าว	N
1117	รายงานตัว	N
1118	รายชื่อ	N
1119	รายได้	NE
1120	รายละเอียด	N
1121	ราว	N
1122	ริ	N
1123	ริม	N
1124	ริมถนน	N
1125	รี	N
1126	รีบ	N
1127	รุด	N
1128	รุนแรง	N
1129	รู้จัก	NE
1130	รูป	N
1131	รูปแบบ	N
1132	เร่ง	N
1133	เร่งด่วน	N
1134	เรา	P
1135	เริ่มต้น	N
1136	เรียก	N
1137	เรียกร้อง	N
1138	เรียกว่า	NE
1139	เรียนรู้	N
1140	เรียบร้อย	N
1141	เรียบร้อยแล้ว	N
1142	เรือ	N
1143	เรื่อย	N
1144	แรงงาน	N
1145	โร	N
1146	โรค	N
1147	โรงพยาบาล	N
1148	โรงแรม	N
1149	โรด	N
1150	ไร่	N
1151	ไร้	N
1152	ฤ	N
1153	ล	N
1154	ลงทุน	N
1155	ลงโทษ	N
1156	ลงนาม	N

No.	Word	Polar	No.	Word	Polar	No.	Word	Polar
1157	ลงพื้นที่	N	1185	ลูกสาว	N	1213	วันเสาร์	N
1158	ลงมา	N	1186	เล	N	1214	วัย	N
1159	ลงมือ	N	1187	เล็กน้อย	N	1215	วัยรุ่น	NE
1160	ลดลง	N	1188	เลข	NE	1216	วา	N
1161	ลม	N	1189	เลขที่	N	1217	วาง	N
1162	ล้ม	N	1190	เลขาธิการ	N	1218	ว่าด้วย	N
1163	ล่วงหน้า	N	1191	เล่น	P	1219	ว่าที่	N
1164	ล้วน	N	1192	เลย	P	1220	วาน	N
1165	ล้อ	N	1193	เลิก	N	1221	วาระ	N
1166	ลอด	N	1194	เลี้ยง	N	1222	วิ	N
1167	ลอบ	N	1195	เลือก	P	1223	วิกฤต	N
1168	ละคร	P	1196	เลือด	NE	1224	วิเคราะห์	N
1169	ลักษณ์	N	1197	เลื่อน	N	1225	วิจารณ์	N
1170	ลักษณะ	N	1198	แล้วแต่	N	1226	วิน	N
1171	ลับ	N	1199	แล้วเสร็จ	N	1227	วินัย	N
1172	ลา	N	1200	ไล่	N	1228	วินิจฉัย	N
1173	ล่า	N	1201	ว	N	1229	วิพากษ์วิจารณ์	N
1174	ลาน	N	1202	วงศ์	N	1230	วุฒิ	N
1175	ล้าน	N	1203	วงษ์	N	1231	เวที	N
1176	ล้านบาท	N	1204	วจ	N	1232	เว็บไซต์	N
1177	ลาย	N	1205	วร	N	1233	ไว้ได้	N
1178	ล่าสุด	NE	1206	วรรณ	N	1234	ศ	N
1179	ลาออก	N	1207	วัฒน์	N	1235	ศพ	N
1180	ลำ	N	1208	วัฒนธรรม	N	1236	ศรี	N
1181	ลิ	N	1209	วัด	N	1237	ศักดิ์	N
1182	ลี	N	1210	วันจันทร์	N	1238	ศาล	N
1183	ลึก	N	1211	วันเดียวกัน	N	1239	ศาลรัฐธรรมนูญ	N
1184	ลูกค้า	NE	1212	วันที่	N			

No.	Word	Polar
1240	ศิริ	N
1241	ศิรษะ	N
1242	ศึก	N
1243	ศึกษา	NE
1244	ศูนย์	N
1245	เศรษฐกิจ	N
1246	ส	N
1247	สค	N
1248	ส่ง	N
1249	สงคราม	N
1250	สงบ	N
1251	ส่งไป	N
1252	ส่งผล	N
1253	ส่งผลกระทบ	N
1254	ส่งผลให้	N
1255	สงสัย	N
1256	ส่งเสริม	N
1257	สด	N
1258	สต	N
1259	สตา	N
1260	สถาน	N
1261	สถานี	N
1262	สถาบัน	N
1263	สน	N
1264	สนช	N
1265	สนาม	N
1266	สนามบิน	N
1267	สป	N
1268	สปช	P
1269	สภ	N
1270	สภานิติบัญญัติแห่งชาติ	NE
1271	สภาปฏิรูป	N
1272	สม	N
1273	สมบูรณ์	N
1274	สมัคร	N
1275	สมาคม	N
1276	สมาชิกสภา	N
1277	สรรหา	N
1278	สรุป	N
1279	สว	N
1280	สวน	N
1281	ส่วนตัว	NE
1282	ส่วนหนึ่ง	N
1283	สวม	N
1284	สว่าง	NE
1285	สหรัฐ	N
1286	สอดคล้อง	N
1287	สอบ	N
1288	สอบถาม	N
1289	สอบสวน	N
1290	สะพาน	N
1291	สะสม	NE
1292	สัก	N
1293	สังกัด	N
1294	สั่งการ	N
1295	สัญญา	N
1296	สัญลักษณ์	N
1297	สั้น	N
1298	สัปดาห์	N
1299	สัมภาษณ์	N
1300	สา	N
1301	สากล	N
1302	สาขา	N
1303	สาธารณะ	N
1304	สามี	N
1305	สาเหตุ	N
1306	สำคัญ	NE
1307	สำนวน	N
1308	สำนัก	N
1309	สำนักข่าวต่างประเทศ	N
1310	สำนักงาน	N
1311	สำนักงานตำรวจแห่งชาติ	N
1312	สำนักนายกรัฐมนตรี	N
1313	สำรวจ	N
1314	สำเร็จ	N
1315	สิ	N
1316	สิงหาคม	N
1317	สิทธิ์	N
1318	สิน	N
1319	สี	N

No.	Word	Polar
1320	สีดำ	N
1321	สีแดง	N
1322	สืมา	N
1323	สืบเนื่องจาก	N
1324	สืบสวน	N
1325	สื่อ	NE
1326	สื่อมวลชน	NE
1327	สุ	N
1328	สุข	N
1329	สุขภาพ	NE
1330	สุว	N
1331	สู่	N
1332	สู้	NE
1333	สูง	NE
1334	สูงสุด	N
1335	เส้นทาง	N
1336	เสมอ	N
1337	เสร็จ	N
1338	เสร็จสิ้น	N
1339	เสริม	N
1340	เสี่ยง	N
1341	เสียใจ	N
1342	เสียชีวิต	N
1343	เสียหาย	N
1344	เสื้อ	NE
1345	แสง	N
1346	แสดง	N
1347	แสดงให้เห็น	N
1348	แสดงความคิดเห็น	N
1349	แสน	N
1350	ใส	N
1351	ใส่	N
1352	หญิงสาว	N
1353	หน	N
1354	หน่วย	N
1355	หน่วยงาน	N
1356	หน่อย	N
1357	หนัก	N
1358	หนังสือ	N
1359	หนังสือพิมพ์	N
1360	หน้าที่	N
1361	หนี	N
1362	หมดแล้ว	N
1363	หมอ	N
1364	หมาย	N
1365	หมายจับ	N
1366	หมายถึง	N
1367	หมายเลข	N
1368	หมื่น	N
1369	หมู่	N
1370	หรอก	N
1371	หรือเปล่า	N
1372	หลบหนี	N
1373	หลวง	N
1374	หลัก	NE
1375	หลักการ	N
1376	หลักฐาน	N
1377	หลังจาก	N
1378	หลังจากที่	N
1379	หลังจากนั้น	N
1380	หลังจากนี้	N
1381	หลากหลาย	N
1382	หลายครั้ง	N
1383	หลายราย	N
1384	หลีกเลี่ยง	N
1385	หลุด	N
1386	ห่วง	N
1387	หวังว่า	NE
1388	ห้อง	N
1389	หัก	N
1390	หันมา	N
1391	หัว	N
1392	หัวใจ	NE
1393	หัวหน้า	N
1394	หัวหน้าพรรค	N
1395	ห่าง	N
1396	ห่างจาก	N
1397	หาทาง	N
1398	ห้าม	N
1399	หาย	N
1400	หายไป	N
1401	หารือ	N
1402	เหตุ	N

No.	Word	Polar
1403	เหตุการณ์	N
1404	เห็นชอบ	N
1405	เห็นด้วย	N
1406	เหนือ	N
1407	เหมาะสม	N
1408	เหมือน	P
1409	เหมือนกับ	N
1410	เหมือนเดิม	P
1411	เหยื่อ	N
1412	เหล็ก	N
1413	เหล่า	N
1414	แห่	N
1415	แห่งชาติ	N
1416	แห่งหนึ่ง	N
1417	แหล่ง	N
1418	ให้การ	N
1419	ให้กำลังใจ	N
1420	ให้ความสำคัญ	N
1421	ใหม่	P
1422	ให้ออก	N
1423	ไหม	NE
1424	ไหล	N
1425	อก	N
1426	องค์	N
1427	อดีต	N
1428	อธิบดี	N
1429	อธิบาย	N
1430	อน	N
1431	อนาคต	N
1432	อนุญาต	N
1433	อนุญาตให้	N
1434	อนุมัติ	N
1435	อบ	N
1436	อย	N
1437	อย่าง	P
1438	อย่างใกล้ชิด	N
1439	อย่างชัดเจน	N
1440	อย่างเดียว	N
1441	อย่างต่อเนื่อง	N
1442	อย่างเต็มที่	N
1443	อย่างนี้	N
1444	อย่างแน่นอน	N
1445	อย่างเป็นทางการ	N
1446	อย่างมาก	N
1447	อย่างยิ่ง	N
1448	อย่างรวดเร็ว	N
1449	อย่างรุนแรง	N
1450	อย่างไรก็ตาม	N
1451	อย่างละเอียด	N
1452	อย่างหนัก	N
1453	อยู่	P
1454	อยู่ระหว่าง	N
1455	ออ	N
1456	ออกจาก	NE
1457	ออกแบบ	N
1458	ออกหมายจับ	N
1459	อ่อน	N
1460	ออนไลน์	N
1461	อังกฤษ	N
1462	อัตรา	N
1463	อัน	N
1464	อันดับ	N
1465	อันตราย	N
1466	อันเป็น	N
1467	อัยการ	N
1468	อาการ	N
1469	อากาศ	N
1470	อาคาร	N
1471	อ้าง	N
1472	อ้างว่า	N
1473	อาจจะมี	N
1474	อาจารย์	N
1475	อาชีพ	N
1476	อาเซียน	N
1477	อาทิ	N
1478	อ่าน	N
1479	อารมณ์	N
1480	อาวุธ	N
1481	อาวุธปืน	N
1482	อาศัย	N
1483	อาศัยอยู่	N
1484	อาหาร	NE
1485	อำนาจ	NE

No.	Word	Polar
1486	อำเภอ	N
1487	อิน	NE
1488	อิสระ	N
1489	อี	N
1490	อีกคน	N
1491	อีกครั้ง	N
1492	อุดม	N
1493	อุบัติเหตุ	N
1494	อุปกรณ์	N
1495	อุปสรรค	N
1496	เอ	N
1497	เอก	N
1498	เอกชน	N
1499	เอกสาร	N
1500	เอง	P
1501	เอเชีย	N
1502	เอ็ม	N
1503	โอ	N
1504	โอกาส	P
1505	โอชา	N
1506	ฮอนด้า	N

Appendix F

Feature vector of sentiment classification

This table shows the top 1,000 of frequency sentiment word

Freq.: Frequency

No.	Words	Freq.	No.	Words	Freq.	No.	Words	Freq.
1	คน	131815	26	ประเทศ	38760	51	หรือไม่	28409
2	ไม่	108979	27	ที่จะ	37044	52	รายงาน	28370
3	ปี	89664	28	นำ	35457	53	คสช	28335
4	วันที่	69142	29	พลอ	33957	54	ใหม่	28028
5	เรื่อง	66393	30	ตรวจสอบ	33635	55	บอก	27853
6	นาย	65865	31	ได้รับ	33435	56	อายุ	27499
7	น	60793	32	ดี	32716	57	อดีต	27090
8	ไม่ได้	57667	33	มัน	31922	58	รถ	26795
9	ถูก	55609	34	กล่าวว่า	31894	59	คิด	26763
10	ประชาชน	55096	35	บริเวณ	31871	60	ผู้	26506
11	เวลา	53884	36	พล	31576	61	วันนี้	26401
12	ใช้	51111	37	พร้อม	31489	62	บ้าน	26009
13	ไทย	49049	38	พศ	31267	63	เดินทาง	25674
14	ผม	48785	39	บาท	31051	64	พื้นที่	25374
15	สามารถ	47586	40	ที่ผ่านมา	30866	65	ต่อไป	25080
16	ทำให้	47078	41	เข้ามา	30866	66	จำนวน	24714
17	ทำ	46402	42	จน	30831	67	อยู่ใน	24696
18	ตน	46239	43	ผ่าน	30549	68	สังคม	24631
19	ดังกล่าว	43673	44	งาน	30218	69	ใคร	24490
20	เกิด	43172	45	สำหรับ	30131	70	คดี	24127
21	ไม่มี	42322	46	ออกมา	29987	71	สร้าง	24023
22	รัฐบาล	41128	47	เมื่อวันที่	29607	72	ปัญหา	23427
23	เจ้าหน้าที่	40146	48	หนึ่ง	29418	73	จ	23236
24	กลุ่ม	40106	49	กำลัง	28853	74	ดู	23130
25	ที่มี	39558	50	ตัว	28468	75	ตำรวจ	22893

No.	Words	Freq.	No.	Words	Freq.	No.	Words	Freq.
76	ไม่ใช่	22737	101	ข่าว	19359	126	เสนอ	16686
77	ประมาณ	22543	102	บน	19259	127	การเลือกตั้ง	16629
78	จริง	22537	103	ผู้สื่อข่าว	19140	128	ประธาน	16581
79	ทหาร	22441	104	ช่วย	18915	129	จุด	16380
80	นายกรัฐมนตรี	22300	105	เข้าไป	18841	130	ผบ	16370
81	กรณี	21951	106	ควร	18762	131	ในช่วง	16348
82	พ	21686	107	อ	18351	132	เพียง	16192
83	รัฐธรรมนูญ	21668	108	รอ	18167	133	พูด	16091
84	ท่าน	21491	109	ตอ	18138	134	ขอให้	16075
85	ภายใน	21221	110	ถาม	18041	135	รอง	16072
86	ฝ่าย	21197	111	กลับ	17996	136	ขณะนี้	16000
87	สิ่ง	20979	112	อำนาจ	17935	137	น้ำ	15993
88	อย่างไร	20825	113	ประเทศไทย	17800	138	คุณ	15916
89	ภาพ	20790	114	ใด	17749	139	ต้องการ	15896
90	เดือน	20225	115	พิจารณา	17719	140	จะต้อง	15863
91	ทั้งหมด	20221	116	ราคา	17666	141	สูง	15781
92	ประยุทธ์	20203	117	ทำงาน	17611	142	เพื่อให้	15726
93	ดำเนินการ	20175	118	ตท	17544	143	จะเป็น	15713
94	เงิน	19966	119	อื่น	17498	144	ใหญ่	15699
95	ขณะที่	19937	120	ข้อมูล	17419	145	หา	15636
96	ว่าจะ	19930	121	เมือง	17284	146	ทีม	15595
97	โครงการ	19885	122	สำคัญ	17201	147	ระบบ	15547
98	หน้า	19830	123	ทุกคน	16859	148	เป็นผู้	15340
99	กฎหมาย	19809	124	นายก	16799	149	บริษัท	15318
100	ธ	19402	125	ยิ่ง	16794	150	ประกาศ	15278

No.	Words	Freq.	No.	Words	Freq.	No.	Words	Freq.
151	ยืนยัน	15254	176	ประชาธิปไตย	13696	201	เมื่อเวลา	13068
152	นาง	15086	177	เกี่ยวกับ	13644	202	ตำแหน่ง	12978
153	ตัวเอง	15018	178	รัฐ	13602	203	อาจจะ	12946
154	เอา	15017	179	ขาย	13547	204	ก็ต้อง	12900
155	พอ	15013	180	อย่างไรก็ตาม	13531	205	รวมทั้ง	12888
156	ดูแล	14935	181	ประเด็น	13512	206	ทางการเมือง	12826
157	จะทำ	14897	182	ลูก	13484	207	เรื่องนี้	12806
158	ชุด	14840	183	แก่	13463	208	เลือกตั้ง	12804
159	เด็ก	14750	184	เศรษฐกิจ	13442	209	เหตุ	12799
160	ล้านบาท	14711	185	พยายาม	13439	210	การเมือง	12764
161	ให้กับ	14680	186	เหมือน	13428	211	ถนน	12760
162	ที่มา	14669	187	ติดตาม	13414	212	ร่วมกัน	12742
163	ณ	14489	188	ระบุว่า	13392	213	ในพื้นที่	12683
164	คณะกรรมการ	14410	189	เพิ่ม	13340	214	ซื้อ	12670
165	เกิดขึ้น	14304	190	สอบสวน	13326	215	สส	12666
166	เชื่อ	14284	191	เป็นเรื่อง	13299	216	อยู่ที่	12637
167	แค่	14246	192	ยังมี	13272	217	ครั้งนี้	12527
168	โดยเฉพาะ	14058	193	ไหน	13238	218	ดังนั้น	12472
169	บ้าง	13983	194	สาว	13230	219	ระดับ	12462
170	ฉบับ	13846	195	ศาล	13218	220	บุคคล	12442
171	เลือก	13844	196	มาจาก	13212	221	สถานการณ์	12440
172	สนช	13805	197	ที่สุด	13198	222	ข้าว	12438
173	เหตุการณ์	13752	198	น้อง	13194	223	แจ้ง	12423
174	ครับ	13726	199	จังหวัด	13183	224	ล่าสุด	12414
175	เท่านั้น	13717	200	รวมถึง	13164	225	จันทร์	12391

No.	Words	Freq.	No.	Words	Freq.	No.	Words	Freq.
226	ต	12346	250	ชาติ	11496	275	พวก	10739
227	ชาว	12324	251	มากกว่า	11395	276	ชื่อ	10725
228	ร่าง รัฐธรรมนูญ	12295	252	ยิง	11308	277	นส	10653
			253	วัด	11275	278	เบื้องต้น	10643
229	กำหนด	12258	254	ตร	11252	279	ชอบ	10635
230	เสียชีวิต	12233	255	นาน	11246	280	กลับมา	10618
231	ก็ได้	12230	256	เดิน	11196	281	ขึ้นมา	10608
232	ก็มี	12175	257	รู้	11155	282	อุณหภูมิ	10604
233	ชีวิต	12152	258	สื่อ	11153	283	คนร้าย	10591
234	นะ	12085	259	รอบ	11133	284	เรื่องการ	10588
235	ร่าง	12076	260	การใช้	11126	285	รู้สึก	10553
236	ภาค	11976	261	ผู้ต้องหา	11103	286	ครอบครัว	10543
237	ตอนนี้	11957	262	มาตรา	11042	287	ชัย	10524
238	ยก	11906	263	หัวหน้า	11039	288	ช่วยเหลือ	10468
239	ชาวบ้าน	11878	264	โลก	11002	289	ค	10448
240	จากนั้น	11870	265	ร้อยละ	10966	290	อย่า	10421
241	ย	11868	266	เกี่ยวข้อง	10959	291	เจ้าหน้าที่ ตำรวจ	10417
242	ฐานะ	11855	267	ร	10902			
243	สภ	11813	268	ที่ทำ	10863	292	มิ	10416
244	ผู้ที่	11714	269	ชม	10833	293	จับกุม	10400
245	รักษา	11692	270	รัก	10833	294	นี่	10396
246	เธอ	11690	271	ตนเอง	10819	295	คำ	10393
247	เล่น	11612	272	ใส่	10805	296	คาดว่า	10374
248	หน้าที่	11597	273	มากขึ้น	10796	297	ศพ	10374
249	แสดง	11553	274	นโยบาย	10786	298	ทราบ	10371

No.	Words	Freq.	No.	Words	Freq.	No.	Words	Freq.
299	คนไทย	10356	324	กค	9824	349	แม่	9416
300	คณะ	10355	325	กกต	9781	350	ประชุม	9403
301	หน่วยงาน	10342	326	ความเห็น	9769	351	เดิม	9394
302	ปฏิรูป	10310	327	กม	9763	352	ปัจจุบัน	9323
303	อาการ	10301	328	ไม่เคย	9753	353	ชุมนุม	9322
304	เพื่อน	10300	329	ชาย	9745	354	อัน	9318
305	หลัก	10292	330	ต่อมา	9737	355	อีกครั้ง	9295
306	สนับสนุน	10264	331	พัฒนา	9731	356	ยอมรับ	9267
307	ไปยัง	10246	332	ศรี	9716	357	องค์กร	9227
308	เปิดเผยว่า	10196	333	การทำงาน	9697	358	ส	9227
309	นัด	10173	334	สอง	9697	359	เรียก	9222
310	ข้อความ	10166	335	ลด	9659	360	ร้าน	9197
311	ละ	10163	336	องศาเซลเซียส	9655	361	พา	9175
312	แต่ก็	10096	337	ชัดเจน	9633	362	เข้าใจ	9135
313	ถือว่า	10085	338	จีน	9628	363	ยังคง	9129
314	ที่ถูก	10063	339	นั่ง	9623	364	น่าจะ	9109
315	พระ	10030	340	จัดการ	9575	365	ส่วนใหญ่	9105
316	สาย	10016	341	สี	9570	366	เฟซ	9075
317	แก้ไข	9990	342	นอกจากนี้	9566	367	ลักษณะ	9068
318	ให้การ	9964	343	คัน	9564	368	ชั้น	9067
319	การประชุม	9941	344	ปรับ	9540	369	ดำเนินคดี	9038
320	โพสต์	9895	345	ออกไป	9517	370	โอชา	9005
321	ถือเป็น	9894	346	ขนาด	9471	371	คู่	8995
322	ตรวจ	9860	347	ตต	9426	372	คำสั่ง	8993
323	เสียง	9854	348	เกิดเหตุ	9422	373	คุย	8989

No.	Words	Freq.	No.	Words	Freq.	No.	Words	Freq.
374	หมู่	8988	399	ต่างประเทศ	8634	421	ผิด	8264
375	ที่เกิดขึ้น	8966	400	ม	8584	422	น้อย	8257
376	ปปช	8965	401	ออกจาก	8569	423	ยา	8249
377	ก็ยัง	8964	402	คลิป	8559	424	ออนไลน์	8240
378	แนวทาง	8909	403	ลักษณ์	8537	425	สค	8239
379	ข้าราชการ	8906	404	เปลี่ยน	8486	426	ป	8223
380	กลายเป็น	8876	405	คืน	8482	427	ที่เกิดเหตุ	8210
381	ส่วนตัว	8867	406	เขต	8473	428	กรณีที่	8207
382	ราชการ	8861	407	ระบุ	8471	429	ตามที่	8203
383	เตรียม	8854	408	ในประเทศ	8467	430	แบบนี้	8180
384	ล้าน	8832	409	ไม่รู้	8451	431	สูงสุด	8167
385	กทม	8826	410	เร่ง	8419	432	หนังสือ	8160
386	สั่ง	8816	411	ชี้แจง	8412	433	เพิ่มเติม	8141
387	การปฏิรูป	8797	412	แทน	8409	434	ปล่อย	8140
388	จำนวนมาก	8794	413	เขียน	8368	435	จำเลย	8137
389	ทอง	8769	414	บุ๊ก	8346	436	นักศึกษา	8135
390	ฟัง	8757	415	คณะรักษาความสงบแห่งชาติ	8340	437	พิเศษ	8102
391	สภาพ	8732				438	มอง	8085
392	ไม่ต้อง	8730				439	ติด	8078
393	กระบวนการ	8718	416	เก็บ	8332	440	สุดท้าย	8075
394	สิทธิ	8703	417	เมตร	8326	441	ตลาด	8056
395	ไม่ให้	8694	418	พนักงานสอบสวน	8318	442	สัมภาษณ์	8051
396	พรรค	8659				443	ป้องกัน	8038
397	การที่	8643	419	ช่อง	8315	444	ฐาน	8012
398	หมด	8642	420	สว	8314	445	พี่	7990

No.	Words	Freq.	No.	Words	Freq.	No.	Words	Freq.
446	ประจำ	7980	471	ค่ะ	7653	496	สาเหตุ	7333
447	แต่งตั้ง	7978	472	ตอน	7648	497	ล	7310
448	ชน	7975	473	รถยนต์	7637	498	จำกัด	7294
449	เจอ	7956	474	ระเบิด	7618	499	โอกาส	7292
450	ที่อยู่	7944	475	กร	7596	500	ครม	7290
451	ทำการ	7903	476	แต่ละ	7590	501	นักท่องเที่ยว	7275
452	งบประมาณ	7885	477	แห่งชาติ	7568	502	พ่อ	7261
453	มองว่า	7870	478	เรียกร้อง	7560	503	สปช	7253
454	เหล่านี้	7853	479	ตุลาคม	7551	504	เส้นทาง	7248
455	สน	7849	480	จบ	7543	505	แฟน	7242
456	กันยายน	7836	481	สมาชิก	7529	506	ทุกอย่าง	7238
457	กลาง	7817	482	สหรัฐ	7517	507	กิน	7223
458	รายการ	7816	483	ธุรกิจ	7494	508	พรบ	7220
459	ปกติ	7813	484	สู่	7494	509	ง่าย	7199
460	ให้เป็น	7810	485	ข้อ	7458	510	ศึกษา	7195
461	แยก	7778	486	ทำหน้าที่	7445	511	ห้อง	7181
462	ยื่น	7768	487	ฝั่ง	7440	512	แดง	7162
463	ทุจริต	7762	488	กล่าวถึง	7427	513	ความขัดแย้ง	7158
464	เสีย	7753	489	ผู้ตาย	7423	514	ภายหลัง	7157
465	แพทย์	7746	490	รุ่น	7417	515	ทำไม	7151
466	ทรัพย์สิน	7717	491	ว่าการ	7402	516	ความผิด	7142
467	ปิด	7698	492	นาที	7390	517	ห้าม	7138
468	ศูนย์	7697	493	แม้	7367	518	รายได้	7122
469	โรงเรียน	7690	494	หนุ่ม	7359	519	กรรมการ	7120
470	เสื้อ	7667	495	นพ	7348	520	ก่อเหตุ	7118

No.	Words	Freq.	No.	Words	Freq.	No.	Words	Freq.
521	หญิง	7081	546	ร่วมกับ	6668	571	ตัดสินใจ	6415
522	เข้าร่วม	7067	547	เหลือ	6664	572	เช่นกัน	6399
523	รี	7039	548	เวที	6660	573	จ่าย	6389
524	นา	6954	549	เลขที่	6635	574	เรียน	6378
525	อาหาร	6945	550	หลักฐาน	6632	575	คลื่น	6338
526	เผย	6935	551	ก่อนหน้านี้	6632	576	ดร	6332
527	ควบคุม	6914	552	นักการเมือง	6631	577	รายละเอียด	6327
528	กปปส	6889	553	ประสาน	6616	578	อ่าน	6325
529	กรุงเทพ	6881	554	ผกก	6603	579	พร้อมกับ	6305
530	โรงพยาบาล	6858	555	กย	6576	580	สิงหาคม	6286
531	ทันที	6836	556	สภา	6561	581	ทั้งสอง	6284
532	สินค้า	6835	557	อร์	6537	582	การศึกษา	6280
533	ราช	6816	558	รางวัล	6528	583	มีปัญหา	6274
534	รพ	6799	559	เร็ว	6523	584	กระแส	6270
535	พง	6786	560	ไม่ทราบ	6521	585	ผู้เสียชีวิต	6262
536	ตลอด	6780	561	ได้มีการ	6518	586	ข้อเท็จจริง	6253
537	ไม่ว่า	6779	562	เกม	6510	587	วัย	6252
538	หนัก	6767	563	ความรัก	6509	588	ผบก	6243
539	ยาก	6759	564	ชั่วคราว	6493	589	แก้ปัญหา	6192
540	พร	6758	565	ใจ	6484	590	ทรง	6173
541	ตก	6738	566	ทั่วไป	6480	591	เพิ่มขึ้น	6167
542	แกนนำ	6718	567	พูดคุย	6470	592	มาแล้ว	6156
543	มาตรการ	6715	568	ขั้นตอน	6434	593	มค	6135
544	โดน	6699	569	กิจกรรม	6431	594	ยาง	6130
545	แสน	6690	570	ประโยชน์	6429	595	หารือ	6127

No.	Words	Freq.	No.	Words	Freq.	No.	Words	Freq.
596	สอบ	6123	621	คนอื่น	5962	646	เก่า	5717
597	อันดับ	6123	622	อิน	5951	647	ริ	5707
598	อีกทั้ง	6112	623	บริการ	5949	648	ประกอบ	5702
599	ที่ประชุม	6112	624	ไหม	5922	649	เยอะ	5686
600	ทราบว่า	6110	625	รัฐประหาร	5901	650	ทบ	5655
601	ข้อหา	6096	626	วิธี	5896	651	ความคืบหน้า	5644
602	เครื่อง	6090	627	รูป	5875	652	คส	5640
603	อาคาร	6088	628	ระวัง	5855	653	ลงไป	5640
604	มกราคม	6071	629	ได้แก่	5839	654	การนำ	5638
605	สอบถาม	6061	630	เลขาธิการ	5836	655	ทำได้	5637
606	พย	6061	631	การกระทำ	5832	656	ได้นำ	5634
607	สื่อมวลชน	6056	632	ใกล้	5828	657	ตอบ	5628
608	ต่อต้าน	6046	633	ตค	5820	658	เชิญ	5613
609	ความเสียหาย	6043	634	ขอบคุณ	5810	659	ใช้เวลา	5611
610	โฆษก	6039	635	แก้ไขปัญหา	5809	660	จะไป	5600
611	ลดลง	6038	636	วาง	5799	661	สุขภาพ	5596
612	อาทิ	6031	637	แรง	5795	662	พวกเขา	5594
613	บุญ	6016	638	นำมา	5791	663	อก	5592
614	ติดต่อ	5999	639	การพิจารณา	5775	664	บางคน	5585
615	จับ	5997	640	แน่นอน	5767	665	ได้มา	5585
616	ประ	5992	641	รัฐมนตรี	5753	666	หัว	5571
617	หยุด	5970	642	หลบหนี	5738	667	ขาว	5571
618	เข้าสู่	5967	643	ชุมชน	5730	668	ส่งเสริม	5562
619	เม	5966	644	คนใน	5725	669	สัปดาห์	5556
620	ช	5963	645	ข้าง	5722	670	รมว	5554

No.	Words	Freq.	No.	Words	Freq.	No.	Words	Freq.
671	ท้องถิ่น	5554	696	กพ	5441	721	เกิน	5318
672	ชนะ	5553	697	เล็ก	5435	722	เพิ่ง	5316
673	หมู่บ้าน	5549	698	มาถึง	5434	723	ต้น	5303
674	หน่วย	5548	699	เช้า	5427	724	แก้	5301
675	เกือบ	5542	700	รีบ	5423	725	พฤษภาคม	5291
676	นั่น	5536	701	ด้วยกัน	5419	726	ปลูก	5278
677	สำนักงาน	5523	702	สุ	5417	727	ยืน	5271
678	รา	5519	703	ไม่เป็น	5416	728	จอด	5264
679	ใต้	5516	704	ช่วยกัน	5403	729	อาวุธปืน	5263
680	พค	5515	705	ญาติ	5397	730	ทิ้ง	5247
681	ยึด	5505	706	วิ่ง	5379	731	บริหาร	5241
682	ษ์	5504	707	แรงงาน	5378	732	ตามมา	5241
683	อากาศ	5497	708	เจ้าของ	5374	733	อาจารย์	5235
684	ภรรยา	5496	709	ธันวาคม	5369	734	ป่า	5231
685	คร	5494	710	ขณะเดียวกัน	5366	735	เอ	5228
686	รับผิดชอบ	5494	711	ที่ใช้	5359	736	กระทั่ง	5225
687	มือ	5491	712	ตา	5355	737	นำไป	5221
688	บรรยากาศ	5483	713	บ้านเมือง	5351	738	เป็นความ	5215
689	ทะเบียน	5477	714	วิ	5351	739	ต่ำสุด	5207
690	รร	5472	715	ปอ	5344	740	ความเร็ว	5206
691	แต่อย่างใด	5469	716	ความมั่นคง	5342	741	พรรคเพื่อไทย	5197
692	จนถึง	5454	717	กิจการ	5338	742	การปฏิบัติ	5196
693	ว	5451	718	พร้อมด้วย	5337	743	ประกอบด้วย	5195
694	ใบ	5442	719	ประตู	5330	744	พรรคการเมือง	5192
695	ความปลอดภัย	5441	720	ภายใต้	5325	745	บก	5189

No.	Words	Freq.	No.	Words	Freq.	No.	Words	Freq.
746	มหาวิทยาลัย	5188	771	เครือข่าย	5008	796	เดช	4799
747	สมัย	5188	772	วิธีการ	5004	797	เชียงใหม่	4795
748	เรือ	5187	773	กองทัพ	5004	798	คณะรัฐมนตรี	4794
749	ผู้นำ	5173	774	ท	5002	799	พัก	4792
750	แก้ว	5172	775	ยาว	4985	800	การให้	4790
751	เดินหน้า	5157	776	ดีขึ้น	4978	801	การสร้าง	4787
752	มีโอกาส	5154	777	ฆ่า	4962	802	เรียกว่า	4782
753	บิน	5149	778	มนุษย์	4958	803	ประเภท	4768
754	ค่า	5144	779	พันธุ์	4946	804	สถาบัน	4767
755	ขาด	5136	780	บางส่วน	4941	805	ในทาง	4766
756	รูปแบบ	5126	781	คะแนน	4941	806	เอาไว้	4762
757	เกษตรกร	5113	782	ตาย	4936	807	ความรุนแรง	4760
758	แล้วก็	5103	783	ขับรถ	4935	808	ฉัน	4752
759	ธรรม	5103	784	เอกสาร	4935	809	ปฏิบัติหน้าที่	4751
760	ญี่ปุ่น	5097	785	การพัฒนา	4931	810	นักเรียน	4734
761	แถลง	5095	786	ตามกฎหมาย	4916	811	ไม่เกิน	4733
762	เย็น	5090	787	ครั้งที่	4914	812	ศักดิ์	4728
763	ไม่มีใคร	5080	788	การบริหาร	4898	813	ที่ดิน	4727
764	ดำรงตำแหน่ง	5080	789	เจรจา	4867	814	ทำอะไร	4727
765	ครู	5053	790	ชั่วโมง	4858	815	ควบคุมตัว	4725
766	เหมือนกัน	5048	791	อาวุธ	4851	816	ไม้	4723
767	ต่อว่า	5042	792	ชี้	4839	817	พนักงาน	4718
768	อย่างต่อเนื่อง	5034	793	กรกฎาคม	4835	818	นปช	4717
769	สถานที่	5017	794	โต	4817	819	ไม่ดี	4711
770	ลม	5011	795	ยกเลิก	4809	820	โดยจะ	4708

No.	Words	Freq.	No.	Words	Freq.	No.	Words	Freq.
821	ร่างกาย	4703	846	สุข	4573	871	ละคร	4443
822	เหตุผล	4702	847	ยาเสพติด	4571	872	เล่า	4435
823	แล้วว่า	4701	848	ง	4567	873	คอย	4416
824	อาชีพ	4700	849	ยุค	4559	874	ชาวนา	4410
825	ผู้โดยสาร	4665	850	กุมภาพันธ์	4556	875	วงศ์	4409
826	พิธี	4657	851	กรอบ	4552	876	สามี	4405
827	ด้วยว่า	4653	852	สั่งการ	4550	877	กอง	4401
828	เสียหาย	4649	853	ผู้หญิง	4548	878	เพลง	4401
829	ฟ้อง	4640	854	แม้จะ	4542	879	บังคับ	4398
830	ราว	4626	855	ชนิด	4541	880	พฤติกรรม	4387
831	เครื่องบิน	4626	856	ทะเล	4535	881	ผลิต	4384
832	ดา	4625	857	ทุกฝ่าย	4532	882	ข้อเสนอ	4370
833	สรรหา	4617	858	ภาคใต้	4510	883	พัน	4364
834	ทางการ	4609	859	ร์	4506	884	ความคิด	4363
835	อา	4605	860	เผยแพร่	4505	885	ทำลาย	4341
836	หลังจากที่	4604	861	เลี้ยง	4495	886	ตัด	4340
837	เจ	4603	862	รถไฟ	4492	887	ดีกว่า	4327
838	ผู้อำนวยการ	4600	863	นอน	4484	888	จำหน่าย	4327
839	ปฏิเสธ	4595	864	ทำให้เกิด	4478	889	ปืน	4322
840	จะมา	4594	865	รุนแรง	4464	890	ล่า	4310
841	นอก	4594	866	อยู่ระหว่าง	4460	891	ปราบปราม	4307
842	ปฏิบัติ	4589	867	ฝาก	4452	892	น้ำมัน	4303
843	แต่ถ้า	4589	868	ฝน	4452	893	ไปที่	4301
844	ลงมา	4584	869	ความจริง	4452	894	เน้น	4297
845	ทั้งที่	4583	870	กา	4452	895	ลูกค้า	4296

No.	Words	Freq.	No.	Words	Freq.	No.	Words	Freq.
896	มูลค่า	4288	921	อธิบดี	4162	945	กระจาย	4041
897	นำตัว	4284	922	บัญชี	4160	946	ต่อสู้	4040
898	ก็คง	4279	923	ความสัมพันธ์	4155	947	ผู้ใหญ่	4036
899	เพชร	4277	924	สม	4154	948	ที่ปรึกษา	4035
900	ทั่วประเทศ	4275	925	ผลกระทบ	4148	949	ที	4031
901	รัตน์	4269	926	อุบัติเหตุ	4137	950	ลงโทษ	4024
902	การดำเนินการ	4264	927	อ้างว่า	4136	951	รู้จัก	4023
903	ผู้บริหาร	4260	928	ชื่อดัง	4132	952	เป็นหลัก	4022
904	บ	4254	929	สาม	4128	953	เมษายน	4021
905	มอบ	4254	930	กระทรวง	4121	954	ส่วนหนึ่ง	4019
906	เช่นนี้	4244	931	เรื่อย	4120	955	รม	4013
907	ตัวแทน	4232	932	กลัว	4115	956	มาร่วม	4013
908	การแข่งขัน	4225	933	การเงิน	4105	957	ปรากฏ	4011
909	คล้าย	4212	934	กระทำ	4098	958	รับฟัง	4006
910	ระยะ	4209	935	จี	4098	959	รายชื่อ	3983
911	มากมาย	4196	936	ความสุข	4086	960	ในอนาคต	3982
912	มีนาคม	4191	937	วงการ	4080	961	ส่งผลให้	3981
913	จัดทำ	4190	938	ขึ้นไป	4078	962	เตือน	3975
914	ระยะเวลา	4180	939	เต็ม	4070	963	วง	3971
915	พลเอก	4179	940	ประวัติศาสตร์	4070	964	ลงทุน	3962
916	ด	4175	941	ย้าย	4067	965	อาเซียน	3958
917	เป็นต้น	4172	942	มีอำนาจ	4054	966	ที่เกิด	3956
918	ยอด	4171	943	ผลงาน	4052	967	ตรงนี้	3954
919	เหมาะสม	4169	944	ยิ่งลักษณ์ ชินวัตร	4047	968	สรุป	3953
920	ก็ไม่ได้	4166				969	บรรดา	3949

No.	Words	Freq.
970	หนู	3949
971	สาขา	3944
972	ตัดสิน	3935
973	มูลนิธิ	3925
974	ที่ว่า	3924
975	ปริมาณ	3920
976	ทั้งสิ้น	3920
977	ศาล รัฐธรรมนูญ	3918
978	หนี	3916
979	สวน	3915
980	ย่อม	3913
981	สำนวน	3910
982	เชิง	3904
983	มีผล	3897
984	ซอย	3895
985	หน	3890
986	เคลื่อนไหว	3887
987	สมาคม	3872
988	นับ	3871
989	ถอดถอน	3869
990	ยังไง	3869
991	มาก่อน	3868
992	เว็บไซต์	3859
993	ดำเนิน	3859
994	รอง นายกรัฐมนตรี	3858
995	เจริญ	3857
996	อีกด้วย	3854
997	ไม่ว่าจะเป็น	3849
998	ไม่ได้รับ	3849
999	หนัง	3847
1000	ไม่ค่อย	3846

VITAE

Name Nut Kunpattanasopon

Student ID 5830223003

Educational Attainment

Degree	Name of Institution	Year of Graduation
Bachelor of Science (Information Technology)	Prince of Songkla University	2011

Work – Position and Address

2012 - 2014	IT Support I.C.C. International Public Co., Ltd.
2014 - 2015	Web programmer, ThaiRoute.com

List of Publication and Proceedings

Kunpattanasopon, N., Tongtep, N., & Hashimoto, K. (2017). "Noise Reduction Effect on Thai Social Texts Sentiment Analysis", *Proceedings of the International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2017)*, Huahin, Prachuap Khiri Khan, Thailand: 27-29 August, 2017.