



Received 18th January 2021,
Revised 14th March 2021,
Accepted 16th March 2021

Sentimental Analysis and Keyword Extraction from Thai Users of Facebook in COVID-19 Period

DOI: 10.14456/x0xx00000x

Boonthida Chiraratanasopha^{1*}

¹Computer Program, Faculty of Science Technology and Agriculture, Yala Rajabhat University, Mueang, Yala, 95000

*E-mail: jboontida16@gmail.com

Abstract

The new outbreak of Covid-19 leads to the confusion and has many effects on people including change in lifestyle and income opportunity. With the campaign of social distancing and stay-home, it becomes difficult to obtain people opinion using traditional methods such as questionnaire and interview. This work thus applies natural language processing tasks including keyword detection and sentimental analysis to analyze a trend of topics and their feelings towards the new outbreak of Covid-19 from Facebook posts. Detected keywords represent significant topics people were frequently mentioned while sentiment helps to reveal people feeling towards the topics. This study collected Facebook data related to Covid-19 posted between 1st November 2020 and 5th January, 2021 from those setting home location in Three Southern Border Provinces of Thailand for 3,127 posts. The keyword detection results indicate that there are four main groups of terms frequently discussed online as terms related disease prevention, terms about worry, terms about healthcare and treatment, and terms related to government policy. From results of sentimental analysis, the collected posts were separated to 80.07% negative posts, 14.42% positive posts and 15.45% neutral posts. Majority of the negative posts (46%) was about feeling frustrated of illegal activities causing the new outbreak whilst 56% of the positive posts were feeling impressive and praising towards fast attempt on prevention of Covid-19 spreading.

Keywords: Sentimental Analysis, Keyword Extraction, Facebook Post, Covid-19

1. Introduction

Social medias are an online platform for people to share their thoughts, record of activities and have interaction to one another (1, 2). Social medias such as Facebook and Twitter have become well-known and widely used online service as online community for sharing idea and topic discussion with actual social impact (2, 3). With the consistent rise of the popularity of social networking, Thai people regardless of gender, age or occupation have usually expressed their thoughts on emerging topics on such social medias, and they become a good source for understanding a trend of topics with social impact. Natural language processing (NLP) is a method for processing and analyzing large amounts of natural language data (3). One of the tasks in NLP is keyword detection (4, 5) from the text. This task is to automatically extract the most semantically significant words in a text to summarize and recognize the main topics being discussed. Another task is a sentimental analysis (3, 6, 7) which is to analyze and identify opinions expressed in a text to determine whether the writer's attitude towards a topic is positive, negative, or neutral. These two tasks

are often used to process a large amount of text data for understanding trends and frequently discussed topics in many domains. Unlike a traditional method of using questionnaires and interviews to approach people for their opinions, the use of text processing is more directive and reflective to people opinion without concerning about biases and misleading questions.

With the re-emerging of Covid-19 (8), Thai most anticipated social media, Facebook, has been used for spreading news and discussion on its many related topics such as curfews, regulations and new-normal. In this work, we apply an automated keyword detection to find discussed topics and sentimental analysis from texts to learn on how the Covid-19 has affected on Thai people. The found topics and sentiments are for learning their concerns, opinions and feelings. Moreover, the results can be used to analyze the lack of supplies for respondent organizations to make decision. In this study, we focus on the Facebook posts made by the locals in the Three Southern Border Provinces which are Yala, Pattani and Narathiwat since these provinces are a border to other countries prevalent with Covid-19,

yet less information is informed than those from North-eastern provinces. The topics of discussion will automatically be found using keyword detection while the feeling of the people will be realized through sentimental analysis. We expect that this study will lead to the findings of how people react to the emerging of Covid-19 and its related situation.

2. Background and Related Works

2.1 Covid-19 in Thailand

With large numbers of tourists coming to Thailand during the New Year holiday in late December 2019, Thailand reported its first Covid-19 infected case in early January 2020 (9). The number of new cases increasingly reported in March as more than a hundred cases reported daily. Measures to contain the outbreak were implemented, such as the cancellation of public gatherings, remote working measures, the closure of gathering venues, and social campaigns to “stay home, stop the virus, save the nation”. One of the major measures in outbreaking situation is national lockdown (10) including prohibition of international inbound travel, restrictions on domestic movement of people, and temporary business closures (11). These measures though helped to maintain number of cases, but some of Thai people criticized them to make their lives more difficulty. Some may adapt to the situation while some may find the measures such as locking down and shutting down businesses to hinder their ordinary lifestyle. These measures however ended the Covid-19 transmission effectively, and it results in the end of the first wave of Covid-19 outbreak in 8 July, 2020 (12). The summary of the timeline for the first wave of Covid-19 outbreak is sketched in Figure 1.

However, some incidents caused Thai citizen to uproar from new cases of Covid-19 after months of calm situation. The incidents are such as a group of people illegally sneaking out to work in a casino in Myanmar and a smuggling of illegal workforces from neighbor countries. These caused a skyrocket of Covid-19 cases in the late 2020 (13),

and the cases were declared to be a new wave of Covid-19 in Thailand. The incidents led to a furious Thai people who patiently endure and strictly follow the recommended measures against Covid-19. Unlike the first wave in which people were panickily afraid of getting Covid-19 and tried to adapt to the situation with their lack of disease knowledge, the new wave of Covid-19 is a situation which people have learned more knowledge of the disease; thus, opinions of people in this situation are valuable to learn from. The information can be used to determine future countermeasures to the emerging disease and to find topics of interest from Thai citizen in this endemic situation.

2.2 Literature Review

Applying Natural language processing (NLP) techniques to process and analyze information from text is a widely used method to learn specific information (insight) in this data-overloaded era. Furthermore, the trend of using text-based sentimental analysis has become necessary tasks for finding users or consumers’ opinion towards their products and services. Sentiment analysis is a classification of opinions as positive, negative or neutral of text data using text analysis techniques. Sentiment analysis approaches can be categorized into two groups as knowledge-based and statistical methods. Knowledge-based technique is to analyze texts based on the presence of words which refers to the feeling (1, 7), for example happy (positive) and troublesome (negative). Thus, a pre-defined dictionary of word aligned with sentiment degree (14) or a knowledge-base of sentiment expression (15-17) is essential to sentimental result. For statistical approach, machine learning techniques, such as latent semantic analysis (LSA) (18), support vector machines (SVM) (5, 16) and deep learning (2, 19), are applied to detect a presence of terms with semantically sentimental expression automatically by using statistical model of terms frequency or co-occurrence of terms.

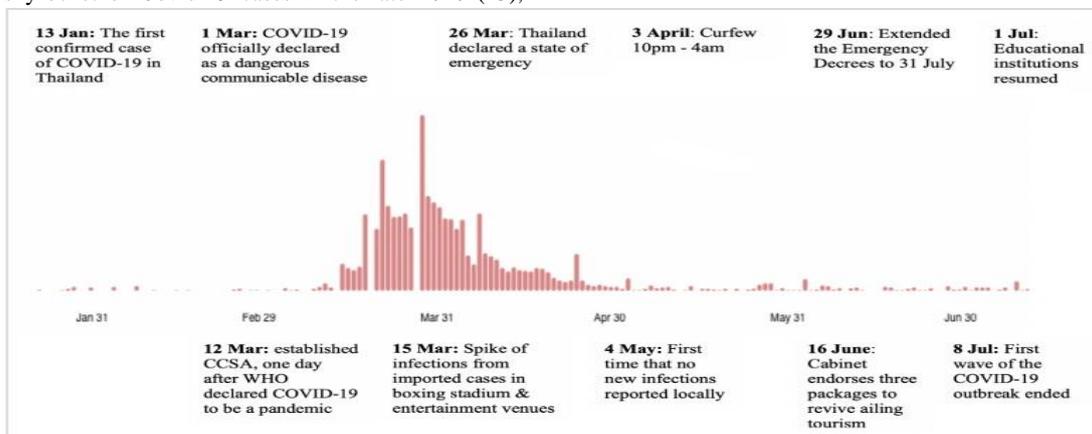


Figure 1 Timeline of COVID-19 Incidence in Thailand (Graph adapted from <https://covid19.who.int/region/searo/country/th>)

We found some works using NLP-based keyword detection and sentimental analysis methods on Thai text as follows. From Rojratnavijit et al. (1), authors aimed to develop a sentimental analysis system for learning feedbacks from customers of the Metropolitan Electricity Authority (MEA). The main text resources of the work are complaints, call logs from customers, and social media data that contain the word ‘MEA’. A dictionary between words and feeling is used to indicate customers’ feeling, and the obtained information is kept for improving its customer services. Piyatumrong et al. (4) applied a keyword detection method using term-frequency and inverse-document frequency to represent highly impact events. The work focuses on three major circumstances as ‘Ratchaprasong’ incident, ‘Bike for mom’ event and ‘The Voice Thailand’ show from Twitter social media. The work can find the significant terms including the topic and important characteristics of the events. Chumwatana (7) applied statistical based sentimental analysis on product reviews given in Facebook and Twitter. 3348 Thai reviews of well-known products and shops such as MK Restaurant, Shabushi Restaurant, JBL Speaker and Nescafe Red Cup Machine were sentimentally analyzed using a classification model learned from sentiment-tagged corpus regarding SentiWordNet. The results of this research provide a summary of feedback trends of products and shops to users to choose the right product or services among competitive brands. Last, Songram et al. (20) presents a system to determine opinions of people regarding government and its revolution from 467 Facebook posts. Words from the posts are annotated for positive or negative sentiment from a specified dictionary, and the tagged posts are trained for a classification model using machine learning techniques such as support vector machines (SVM), Naïve Bayes (NB) and K-nearest neighbor (KNN). The results show that the system has a capability to analyze the new post for its sentimental degree, and KNN model from selected lexical features yielded highest accuracy among all tested models.

In a summary, statistical methods are often used to generate both sentimental analysis model and keyword detection such as term-frequency (TF) (4, 16), support vector machines (SVM) (5, 16, 21), K-nearest neighbor (KNN) (16, 20) and decision tree (DT) (5, 16, 20, 21). These techniques though have different own advantages and disadvantages that fit to their own dataset, and they are accountable to accomplish the task to realize users’ opinion from the data. The difficult and tedious part of the statistical method however is to develop sentiment-annotated data as a training dataset since it requires accurate tag to ascertain prediction quality and accuracy.

3. Research Methods

3.1 Data Collection

This work aims on finding effects of new outbreaking of Covid-19 to Thai people living in the Three Southern Border Provinces which are Yala, Pattani and Narathiwat. The input data are the Facebook posts made in 1st November 2020 and 5th January, 2021. The collected posts were text-based posts excluding image and video. The data were collected anonymously, and only posts set as ‘public’ were retrieved. Since we focus on effects of Covid-19, we pruned the collected data with the term ‘Covid-19’ in both Thai and English language as well as their spelling variants such as ‘โควิด19’, ‘โควิด19’, ‘โควิด-19’ and ‘covid2019’. The posts containing such terms thus were used as materials for text processing.

Since the post data from Facebook is natural Thai language, they contain the issues including unintentional misspelling (1, 20), emerging terms (4) and term transformation (1, 4, 20, 21). These issues reduce the capability to analyze and process the post in terms of accuracy and data distribution; hence, it is important to tackle the issue in prior. The first issue is handled by manually checking and editing since they can be a hint of their original words by two means: common misspelling terms and mistyping using nearby typing buttons such as using ‘u’ instead of ‘a’ since the typing letters are near to one another. For statistics, the data have the following characteristics.

- Amount of collected Facebook posts: 3,127
- Collecting post date: 1st November, 2020 to 5th January, 2021
- Post owner: those having home location set to Yala, Pattani and Narathiwat
- Type of Posts: text data only excluding text given in an image

3.2 Methods

There are two NLP tasks in this work. First, keyword detection (KD) is applied to find significant topic in a post that post owner is discussed. Second, sentimental analysis (SA) is to analyze the sentiment of the post owner via the terms in the post. An overview of the processes is given in Figure 2.

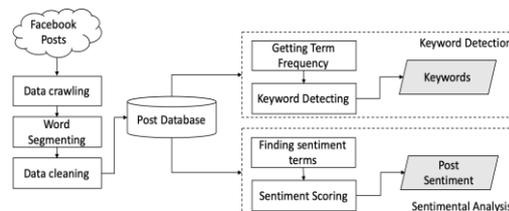


Figure 2 An overview of the processes to Sentimental Analysis and Keyword Extraction from Thai Users of Facebook in COVID-19 Period

3.2.1 Keyword Detection

The targeted input data of this work are Thai text from the Facebook post. Thus, text cleansing and word segmentation are essential as a pre-processing for statistically analyzing the text effectively. For automated Thai word segmentation, free service of Longest matching function from Lexto (LongLexTo) (22) is deployed since we want the term to be semantically understandable instead of a short term. Furthermore, the automatic word segmentation is prone to produce some errors in assigning a word boundary from ambiguity, typos and unknown words; hence, post-edition is required to maintain text data quality. For cleansing, non-terms including ordinal markers, symbols and emoticons are removed since they are not related to a content or provide a little to no semantic meaning. To keep the keywords at a semantic level, name-entities and stop-words (grammatical words) are also removed.

For determining significance of terms, term frequency (TF) and inverse-document frequency (IDF) is exploited. The TF and IDF is defined as given in (3.1) and (3.2), respectively.

$$TF = N(w, p) \quad (3.1)$$

$$IDF = N(w, p) \times \log(|P|/N(p, w)) \quad (3.2)$$

where $N(w, p)$ refers to the number of occurrences of each word (w) in a post (p), while IDF is logarithmic scale value of the collection of the entire posts (P) divided by the number of posts that contained the word (w). The TF-IDF score is a multiplication of TF and IDF, and the higher the score shows the higher the significance.

3.2.2 Sentimental Analysis Method

Sentimental analysis is a method to systematically identify a sentiment (positive, negative and neutral) of a text to learn how the content creator feels towards a specific topic. The common used method for sentimental analysis is machine-learning approach in which trains a model from supervised text data to generate a prediction model for classifying future contents regarding the supervised tag set. Sentiment analysis is widely applied to learn a feedback from comments, reviews and survey responses via online and social media. The approach of this method is supervised-learning; thus, it requires text data with sentiment tag to supervise the machine as training. In fact, it is necessary that the training data are needed to be large to cover all possible words and reducing a chance of overfitting problem. The main issue is that, for the language with low processed text resources such as Thai, sentiment-tagged data are apparently scarce or have the legal right protection. Hence, we apply the free service of sentimental analysis from AI for Thai (23). The service of sentimental analysis provides the sentimental classification model specifically for Thai ready for analyzing a text into

three sentiments, i.e. positive, negative and neutral. For the details of its analyzing method, it consists of three analysis modules as follows.

- Language usage analysis: this module is to realize text about the use of offensive language and the use of informal language.

- Intention analysis: this module aims to categorize text into four intentional types as announcement, request, question and sentiment

- Sentiment analysis: this module assesses a text regarding its sentiment, i.e., positive or negative.

With the combination of these three analysis aspects, an input text is classified by comparing the use of words to a threshold of the prediction model, and the model returns the sentiment result of each input sentence.

In this work, we thus apply the model and method from the service of sentimental analysis from AI for Thai (23) to perform the task since they have proper model learned from valid and appropriate data. The tool is applied to the Covid-19 related posts from Thai Facebook users to find how they feel towards the topic. The sentiment result from the system then is used to reveal overall feeling of participants and represent their feeling towards specific existing issues in Covid-19 pandemic period.

4 Experiment

4.1 Results of Keyword Detection

With the keyword detection, we found the major topics of the posts. Since the terms are segmented in concept level, the found keywords thus are long and have meaning for understanding post owners. The found keywords in the top-20th rank are given in Table 1.

From the results, the terms can be categorized into 4 groups as follows:

- First, a group of terms about disease prevention. This group consists of 1, 6, 7, 9 and 11.

- Second, a group of terms about worry. Terms in this group include 2, 5, 10, 12, 13, 17, 18, 19 and 20.

- Third, a group of terms about healthcare and treatment. This includes 3, 4 and 14.

- Forth, a group of relevant things related to government policy terms such as 8, 15 and 16.

With grouping, the topics being discussed the most in Facebook post from those living in the Three Southern Border Provinces were about worry. This reflects that the people had the concern on disease spreading and economic status. Moreover, they also frequently discussed on how to prevent the disease. This indicates that the government and media successfully provided important information about how serious and easy for the disease to be spread and gained sufficient attention for people to be aware of the disease control and prevention.

Table 1 Top-20 ranked keywords from Facebook posts regarding Covid-19 situation

Rank	Word	TF-IDF score
1	อยู่บ้าน	535.46
2	คิดเชื่อ	497.84
3	วัคซีน	405.69
4	ผู้ป่วย	331.92
5	ลักลอบข้ามเมือง	319.55
6	กักตัว	309.72
7	ป้องกัน	298.53
8	เชี่ยวชาญ	281.29
9	หยุดเชื้อ	280.54
10	ระบาด	274.80
11	หน้ากากอนามัย	260.21
12	อดตาย	253.66
13	ระบาดระลอกใหม่	249.14
14	บุคลากรทางการแพทย์	237.50
15	ด่านชายแดน	229.43
16	หมอชนะ	222.73
17	หมดตัว	222.24
18	ไม่พอกิน	213.35
19	เสี่ยง	211.86
20	บ่อน	208.19

4.2 Results of Sentimental Analysis

From all the collected posts, a result of sentimental analysis in overall is given in Figure 3.

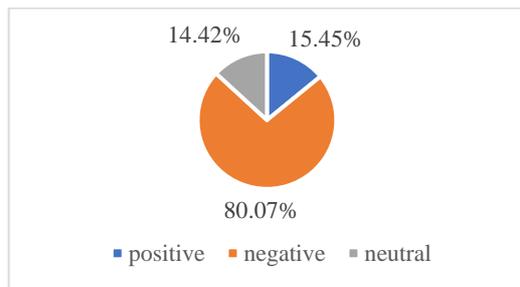


Figure 3 Sentiment results of the posts

From the result, around 80% of the collected data represented negative feeling while 14% were mentioned positively. By further grouping in each sentimental type, we found that about 46% of negative posts were about feeling frustrated from causes of Covid-19 spreading such as country smuggling and illegal casino, 28% were about raging on losing household income, and 21% were feeling down or frustrated from Government policy. For positive Facebook posts, 56% were on a topic of admiring of attempt and sacrifice on prevention of Covid-19 spreading, especially on the Governor of Samut Sakhon province while 29% were happy about vaccine news. The rest are miscellaneous topics.

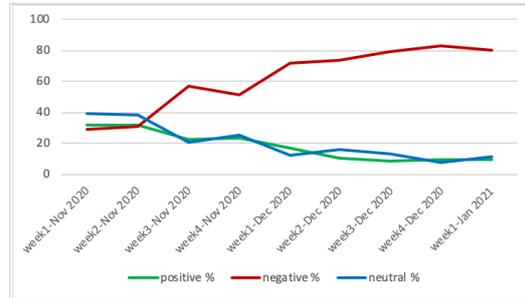


Figure 4 Sentiment results based on weekly timeline

Based on timeline as shown in Figure 4, we found rapid increase on negative sentiment around the third week of November in which according to the beginning of new outbreak, and the percentage of negative sentiments had constantly been over 50% since then. Moreover, we also found that the amount of posts related to Covid-19 increased relatively to the number of illegal incidents that may cause a spreading of Covid-19. In addition, when comparing to a number of confirmed cases, a number of negative sentiment posts reached over 60% after the confirm cases (shown in Figure 5) rose to over 15 cases in daily average per week.

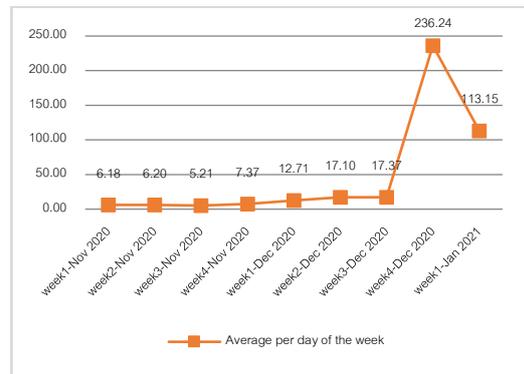


Figure 5 Confirmed Covid-19 cases based on weekly timeline

The results thus show the feeling regarding of concerns and admires during the pandemic revealed by Facebook users. The major topics including negative feelings of being afraid and frustration, and positive feelings of being generous and hard working. From an aspect of lacking supplies, we found that participants rarely mentioned the lacking of hygienic masks or consumables, but they mostly were frustrated on losing of or gaining lower income.

5. Conclusions and Remarks

During the new outbreak of Covid-19 in Thailand, Thais mostly communicate and express their feeling via Social Media with the restriction from social distancing. This work analyzes the

Facebook posts regarding Covid-19 using natural language processing methods called keyword detection and sentimental analysis aiming to understand a trend of topics and feeling of the Facebook users. The automated natural language analyzing methods used on online posts help on finding insight information from people, especially with the new normal of social distancing campaign and lowering the risk of spreading the disease. The found keywords represent trendy topics people were frequently mentioned. The sentimental result reveals their feeling towards the topics. These two aspects thus help us to realize the major concerns of the people during this suffering time. In the study, collected Facebook data were posts that related to Covid-19 during 1st November, 2020 to 5th January, 2021 from those having home location set to Yala, Pattani and Narathiwat. The total amount of collected posts were 3,127 posts which were processed to remove emoticons, stickers and images. The keyword detection results found the top 20 ranked frequently mentioned words grouped into four topics as terms related disease prevention, terms about worry, terms about healthcare and treatment, and terms related to government policy. From sentimental analysis, we found that the posts comprised of 80% negative feeling and 14% positive feeling. The major topic (46%) of the negative posts was about feeling frustrated about illegal activities leading to the new outbreak. About half of the positive posts were feeling admiring of government attempt and sacrifice on prevention of Covid-19 spreading, especially Governor of Samut Sakhon province.

The results give the suggestion that people from the Three Southern Provinces concerns on illegal activities and want the respondent government sectors to tackle the problem as soon as possible. Furthermore, the Covid-19 outbreak also affects their lifestyle and economical status as they were often worriedly expressed; hence, communication is suggested to ensure their stable morale. Last, there is no sign of lacking supplies and consumables from detection regarding Covid-19 outbreak, but news about vaccines was one of the frequently discussed topics that should be clarified on criteria to receive.

Acknowledgements

This research is financially supported by the Rotary Club of Yala, District 3330, Thailand, 2020. The authors would like to thank the Rotary Club of Yala District 3330 Thailand for the publication grants.

Declaration of conflicting interests

The authors declared that they have no conflicts of interest in the research, authorship, and this article's publication.

References

1. Rojratnavijit J, Eiamsithipan, C. Analysing Thai Social Media Content to Improve Customer Satisfaction. In The 25th

- International Conference on Electricity Distribution 2019 June 3-6. pp.1-5.
2. Katchapakirin K, Wongpatikaseree K, Yomaboot P, Kaewpitakkun Y. Facebook social media for depression detection in the Thai community. In 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE) 2018 Jul 11 (pp. 1-6). IEEE.
3. Liebeskind C, Nahon K, HaCohen-Kerner Y, Manor Y. Comparing Sentiment Analysis Models to Classify Attitudes of Political Comments on Facebook (November 2016). *Polibits. Sci.* 2017;55:17-23.
4. Piyatumrong A, Sangkeetrakarn C, Haruechaiyasak C, Kongthon A. Finding Key Terms Representing Events from Thai Twitter. In International Symposium on Natural Language Processing 2016 Feb 10 (pp. 73-87). Springer, Cham.
5. Thanasopon B, Buranapanitkij J, Netisopakul P. Mining Social Media Crowd Trends from Thai Text Posts and Comments. *Malays. J. Comput. Sci.* 2019 Dec 23:96-108.
6. Dey L, Chakraborty S, Biswas A, Bose B, Tiwari S. Sentiment analysis of review datasets using naive bayes and k-nn classifier. arXiv preprint arXiv:1610.09982. 2016 Oct 31.
7. Chumwatana T. Comment Analysis for Product and Service Satisfaction from Thai Customers' Review in Social Network. *J. of Inform. and Commun. Technology.* 2018 Mar 28;17(2):271-89.
8. Wang H, Wang Z, Dong Y, Chang R, Xu C, Yu X, Zhang S, Tsamlag L, Shang M, Huang J, Wang Y. Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China. *Cell Discov.* 2020 Feb 24;6(1):1-8.
9. Minister of Public Health (MOPH). Thailand's experience in the COVID-19 Response. (Accessed March 10, 2021, at https://ddc.moph.go.th/viralpneumonia/eng/file/pub_doc/LDoc9.pdf)
10. Bloomberg. (2020, March 25). Thailand to Impose Broad Lockdown to Fight Novel Coronavirus. (Accessed March 11, 2021, at <https://www.bloomberg.com/news/articles/2020-03-25/thailand-to-impose-broad-lockdown-to-fight-spread-of-coronavirus>)
11. Department of Disease Control (DDC). Guidelines for hotel business operators, service place, or resorts for tourists (Latest announcement on March 22, 2020) (Accessed March 14, 2021, at https://ddc.moph.go.th/viralpneumonia/file/im_commands/im_commands02.pdf) (In Thai)
12. Thaibpsworld. (2020, July 8). DCD says first wave of COVID-19 pandemic in Thailand has ended. (Accessed March 12, 2021, at

- <https://www.thaipbsworld.com/dcd-says-first-wave-of-covid-19-pandemic-in-thailand-has-ended/>)
13. Bangkokpost. (2020, Sep 5). Six caught sneaking in from Myanmar (Accessed March 13, 2021, at <https://www.bangkokpost.com/thailand/general/1980375>)
 14. Thong-iad K, Netisopakul P. Comparison of Thai Sentence Sentiment Tagging Methods Using Thai Sentiment Resource. In International Conference on Computing and Information Technology 2019 Jul 4 (pp. 89-98). Springer, Cham.
 15. Srinilta C, Sunhem W, Tungjitnob S, Thasanthiah S, Vatathanavaro S. Lyric-based sentiment polarity classification of Thai songs. In Proceedings of the International Multi Conference of Engineers and Computer Scientists 2017.
 16. Nomponkrang T, Sanrach C. The Comparison of Algorithms for Thai-Sentence Classification. Int. J. Inf. Educ. Technol. 2016 Oct;6(10):801-808.
 17. Chanlekha H, Damdoug W, Suktarachan M. The Development of Semi-automatic Sentiment Lexicon Construction Tool for Thai Sentiment Analysis. In International Symposium on Natural Language Processing 2016 Feb 10 (pp. 97-111). Springer, Cham.
 18. Inrak P, Sinthupinyo S. Applying latent semantic analysis to classify emotions in Thai text. In 2010 2nd International Conference on Computer Engineering and Technology 2010 Apr 16 (Vol. 6, pp. V6-450). IEEE.
 19. Vateekul P, Koomsubha T. A study of sentiment analysis using deep learning techniques on Thai Twitter data. In 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE) 2016 Jul 13 (pp. 1-6). IEEE.
 20. Songram P, Jareanpon C. Opinion Mining of Thai Politics on Facebook Status Updates. In Proceedings of the 3rd IIAE International Conference on Intelligent Systems and Image Processing 2015 September 2-5. pp. 240-245.
 21. Sarakit P, Theeramunkong T, Haruechaiyasak C, Okumura M. Classifying emotion in Thai youtube comments. In 2015 6th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES) 2015 Mar 22 (pp. 1-5). IEEE.
 22. LexTo-Thai Lexeme Tokenizer (LongLexto). [online] Source: <http://www.sansarn.com/lexto/>
 23. APIs and Services Sentiment Analysis [online] Source: https://aiforthai.in.th/service_sa.php