Microblog Entity Detection for Natural Disaster Management

Wannapa Phopli, Maleerat Maliyaem, and Choochart Haruechaiyasak
Department of Information Technology and Digital Innovation, KMUTNB, Bangkok, Thailand
Email: wannapa.phopli@gmail.com

Hathairat Ketmaneechairat
The College of Industrial Technology, KMUTNB, Bangkok, Thailand

***Abstract*—The information from social network can be used to report crisis situation especially natural disaster events. This paper aims to present the utilization of information from twitter in the natural disaster event topic in order to detect the place of event that occurred base on Named Entity Recognition (NER). The place will extract from microblog – Twitter using three techniques: Degree, Betweeness, and Closeness Centrality and then using majority vote for the end of result. In preprocessing step, data only Thai language was collected from twitter using #hagibis in the topic of Super Typhon Hagibis blowing in Japan. Then using three techniques as mention to select only top 5 of words that related to the event. The experimental result show that the word "Japan, ญี่ปุ่น" is the first word of three methods (Degree, Betweenness, Closeness Centrality) with score of 0.57, 0.55, and 0.65 respectively It showed that the message from twitter can be trusted and indicated the event location.**

***Index Terms*—Twitter message analysis, microblog analysis, entity detection**

I. INTRODUCTION

Natural Disaster is an event that humans cannot be inevitable. How are we doing when encountering those dire situations? Resolving this situation is important. Then it needs to find the best solution for solving problems. In 2019, Thailand faced many natural disasters such as in early 2019, smoke and dust problem were happened in Bangkok and nearby provinces. Moreover, at the end of August 2019, it was started rainy season. Unfortunately, the people have problems with storm blows and floods. Aside from Thailand, there were many countries have-encounter from natural disasters. For instance, Super Typhon Hagibis blowing in Japan that were reported in social media frequently in the meantime. Notification and helps need to do as soon as the event occurs in order to mitigate effect of this situation. Gathering and discovering knowledge data from social media such as twitter can be used to support the decision making for natural disaster management. Nowadays, social network is popularly used in many fields and many researchers used content for their research [1], [2]. Twitter is one of the popular social networks media content especially Thai people used because it includes short messages that people can share their moment. It can be used for personal discussions. The highlight of twitter is fastness to disseminating information which called retweet: the user can retweet a message to another users. It sounds interesting to utilize in the crisis situation. The objective of this paper to present the utilization of information from twitter in the natural disaster event topic in order to detect the place of event that occurred base on Named Entity Recognition.

This paper is organized as follows. Section II introduces Theory and related works. Section III research methodology, Section IV present experimental results and finally conclusion in Section V.

Manuscript received November 21, 2020; revised July 21, 2021.

II. THEORY AND RELEATED WORKS

A. Twitter

The Twitter is one of the social media applications. It has become the most popular micro-blogging social networking website in which users share information or event in the form of very short message limited to 140 characters. It was called "tweet". The outstanding of this application is an event update or data sharing which can update or share in fast. It was called "retweet". Twitter is also being used for many other purposes such as marketing promotions. Moreover, Twitter is also being used by the users to express their opinions and views about prominent issues of daily life which may be social, political, entertainment, or even natural disasters event [3]. According to these attributes, it had the researchers utilize the data from twitter in many aspects. For instance, Kelly Y. Itakura and Noboru Sonehara [2] presented utilization from the message on Twitter by creating a graph for disaster alarm systems in a critical situation. Shih-Feng Yang and Julia Taylor Rayzthe [4] proposed an event detection approach that applied hashtags in tweets. They adopted the feature extraction used in STREAMCUBE for K-mean clustering.

B. Natural Language Processing

Natural Language Processing (NLP) is a part of Artificial Intelligence that involves human language processing. It is the approach of information extraction

doi: 10.12720/jait.12.4.351-356

which is developed for automatic extraction from documents or messages [5], [6]. The challenge of this research is "how can a computer understand human communication?" This topic is popular for the researcher to work on [7]. While the main task of the development of NLP consists of many activities such as Corpus development, Word Segmentation, and Named Entity Recognition. It found that NLP researchers try to develop an approach that can be used for analyzing data effectively with the amount of information that is increasing every day and to take advantage of those data. For example, Rodrigo Agerri *et al*. [8] they designed and deployed a complete chain of NLP modules within virtual machines and present the NLP modules included in the virtual machines. Furthermore, they provide empirical performance results when applied to realistic volumes of news within hard time constraints.

C. Named Entity Recognition

Named Entity Recognition (NER) is one of the main tasks for NLP Development. That is computer development for learning Named Entity of words. If we can improve this task, it can be used to learn better human language. The task of NER relates to identify name of word. The Named Entity is divided into 4 main groups, including name of person, place, organization, and time. NER is a type of information extraction task that has an important role in improving the performance of NLP applications. NER can use in various NLP tasks such as Machine Translation, Question Answering, and Text clustering [9]. For the research in this area, there are researchers who develop methods or algorithms to improve performance to be able to label name in words correctly such as Atefeh Zafarian *et al*. [10]. They used a bilingual corpus and transliteration technique to make a model with low costs on the bilingual corpora. In addition, they used a graph-based semi-supervised method to extract the reliable sentences from an unlabeled data to expand the training set and enhance the quality of NER model.

D. Graph Theory

A graph is a set of nodes or vertices. It is connected by a set of edges that can be represented by G = V, E. It is a graph therefore each edge is a pair of vertices such as (v1, v2, v3) ∈ V (G) and e ∈ E (G) [11] which showed in Fig. 1.

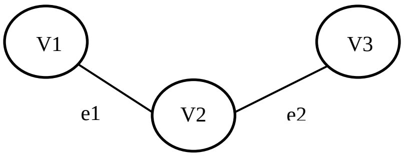

Figure 1. Example of graph.

Graph theory can apply to present relations of the data or events. It is the popular method to represent the data [12]-[14]. Each node is linked by edges to represent relativity that occurs as showed in Fig. 2.

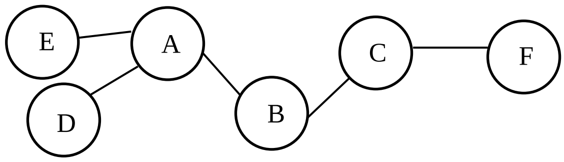

Figure 2. Relation of nodes in graph.

Fig. 2, node A is the most important node because it interrelated with 3 nodes (node B, node D, and nod E). Node B and node C are interrelated. In addition, node C and node F are relevant nodes. It implied that if nodes A, B, C, D, E, and F are event nodes so, event A is the most important event regarding to amount of links that connected to other nodes.

E. Centrality Measure

The centrality is an important index since it indicates which node takes up a critical position in one whole network. It represents the important information from the whole data. For example, if the network comprises of nodes which is the person's name, a person's name node is called centrality if the most influential or the most important among those persons [15].

The centrality is a judgment of the relation of nodes. There are 3 methods for centrality measure.

1) Degree centrality

Degree Centrality (Cd) is measured by the total amount of direct links with the other nodes. The degree centrality for a node is simply its degree. A node with 10 social connections would have a degree centrality of 10. A node with 1 edge would have a degree centrality of 1. The higher values of Degree Centrality mean that the node is more central. Therefore each centrality measure indicates a different type of importance. [14]-[16]. Degree centrality is calculated as follows:

$$d = \sum_{j=1}^{n} Xij\ (i \neq j) \tag{1}$$

2) Betweenness centrality

Betweenness Centrality (Cb) is a measurement of ability of nodes that undertake mediation one node to other nodes. If the node constitutes in the only way which other nodes must go through, then this node should be important and very likely have a high Betweenness Centrality. Technically, it measures the percentage of shortest paths that must go through the specific node. [14]-[16]. Betweenness Centrality is calculated as follows:

$$Cb(Ni) = \sum_{j<k} \frac{Gjk(Ni)}{Gjk} \tag{2}$$

3) Closeness centrality

Closeness centrality finds the node that is closest to all other nodes. Recall that a path is a series of steps that go from one node to another. Closeness centrality for a node is the average length of all the shortest paths from that one node to every other node in the network. If the length of node N's shortest paths with other nodes in the network is small, then the node N has a high closeness centrality. It can show the form of an equation as follows:

$$Cc(Ni) = \frac{1}{[\sum_{i=a}^{n} d(Ni,Nj)]} \ (i \neq j) \tag{3}$$

III. Research Methodology

A. Proposed Framework

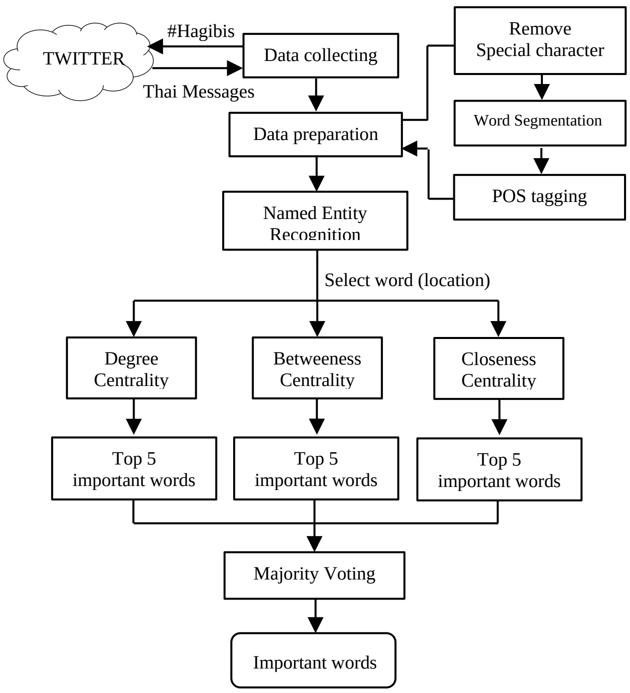

Figure 3. Proposed framework.

Fig. 3 illustrated the framework for Microblog Entity Detection for Natural Disaster Management. From the figure, there are five steps for finding important word as follows:

Step 1: Data collecting: This process used the Twitter API in the rapid miner for gathering the data which was the Thai messages that talked about the super Typhon Hagibis storm blow in Japan which occurred between 9/10/2019 and 12/10/2019. The message was collected by using #hagibis as a keyword in the query message process. The total of data is 1974 messages that were stored in the database system.

Step 2: Data preparation: It is a method for preparing the data before sending these data to the next process: Named Entity Recognition. This process can be divided into 3 subprocess

1) Removing special character or symbol from the messages. It is the approach for filtering unwanted data such as special characters for example $!@* or URL from the messages.

2) Word Segmentation: It is the method for segmentation messages into words. In this paper word_tokenize in Pythainlp [17] was applied for segmentation twitter messages into words. For instance, วันนี้ที่จังหวัดเพชรบุรีจะมีฝนตกหนัก (Today Petchburi province it is raining heavily) when it is separated into words. It will be 'วันนี้', 'ที่', 'จังหวัด', 'เพชรบุรี', 'จะ', 'มี', 'ฝน', 'ตกหนัก'.

3) POS tagging: It was used in the process for labeled part of speech word. When the words were labeled by part of speech. The result of this process will be (word, POS). The process of POS tagging applied pos_tag in Pythainlp [17] For example, the message is 'วันนี้คาดว่ากรุงเทพฝนจะตกหนัก' (Today it is expected that Bangkok is going to rain heavily.) When this message passed POS tagging process the result of the process will be ('วันนี้', 'NCMN'), ('คาด', 'VACT'), ('ว่า', 'JSBR'), ('กรุงเทพมหานคร', 'NPRP'), ('จะ', 'XVBM'), ('มี', 'VSTA'), ('ฝน', None), ('ตกหนัก', None).

The example of Thai message that passed the data preparation process as follows:

The raw data: เมื่อวานที่ผ่านมา (8 ตุลาคม) กรมอุตุนิยมวิทยาญี่ปุ่นเผยว่า #ไต้ฝุ่นฮากิบิส (Hagibis) ไต้ฝุ่นหมายเลข 19 กำลังเคลื่อนตัวมุ่งหน้าสู่ภาคตะวันตกของแผ่นดินญี่ปุ่นและอาจพัฒนาขึ้นเป็นซูเปอร์ไต้ฝุ่นเตรียมพัดถล่มประเทศในช่วงสุดสัปดาห์นี้ด้วย #TheStandardCo https://t.co/BP4SwD4dkV

When the massage passed remove special character process the result of this massages can showed that เมื่อวานที่ผ่านมา8ตุลาคมกรมอุตุนิยมวิทยาญี่ปุ่นเผยว่าไต้ฝุ่นฮากิบิสไต้ฝุ่นหมายเลข19กำลังเคลื่อนตัวมุ่งหน้าสู่ภาคตะวันตกของแผ่นดินญี่ปุ่นและอาจพัฒนาขึ้นเป็นซูเปอร์ไต้ฝุ่นเตรียมพัดถล่มประเทศในช่วงสุดสัปดาห์นี้ด้วย. According to the raw data, this message #, (Hagibis), and (#TheStandard Co https://t.co/BP4SwD4dkV) are unwanted data so when this massage passes the data preparation process this message will be removed.

After that, the message was segmented into words. It showed the result of word segmentation as follows:

'เมื่อวาน', 'ที่ผ่านมา', '8', 'ตุลาคม', 'กรมอุตุนิยมวิทยา', 'ญี่ปุ่น', 'เผย', 'ว่า', ''ไต้ฝุ่น', 'ฮา', 'กิ', 'บิ', 'ส', ''ไต้ฝุ่น', 'หมายเลข', '19', 'กำลัง', 'เคลื่อนตัว', 'มุ่งหน้า', 'สู่', 'ภาค', 'ตะวันตก', 'ของ', 'แผ่นดิน', 'ญี่ปุ่น', 'และ', 'อาจ', 'พัฒนา', 'ขึ้น', 'เป็น', 'ซูเปอร์', ''ไต้ฝุ่น', 'เตรียม', 'พัด', 'ถล่ม', 'ประเทศ', 'ใน', 'ช่วง', 'สุดสัปดาห์', 'นี้', 'ด้วย'

Finally of the data preparation process, when the message was segmented into words. These words will bring to POS tagging process. The result of this process as

('เมื่อวาน', None), ('ที่ผ่านมา', None), ('8', 'DONM'), ('ตุลาคม', 'NCMN'), ('กรมอุตุนิยมวิทยา', None), ('ญี่ปุ่น', 'NPRP'), ('เผย', None), ('ว่า', 'JSBR'), (''ไต้ฝุ่น', None), ('ฮา', None), ('กิ', 'NCMN'), ('บิ', None), ('ส', 'NCMN'), (''ไต้ฝุ่น', None), ('หมายเลข', 'NCMN'), ('19', 'NCNM'), ('กำลัง', 'XVBM'), ('เคลื่อนตัว', None), ('มุ่งหน้า', None), ('สู่', 'RPRE'), ('ภาค', 'NCMN'), ('ตะวันตก', 'NCMN'), ('ของ', 'RPRE'), ('แผ่นดิน', 'NCMN'), ('ญี่ปุ่น', 'NPRP'), ('และ', 'JCRG'), ('อาจ', 'XVMM'), ('พัฒนา', 'VACT'), ('ขึ้น', 'XVAE'), ('เป็น', 'VSTA'), ('ซูเปอร์', None), (''ไต้ฝุ่น', None), ('เตรียม', 'VACT'), ('พัด', None), ('ถล่ม', None), ('ประเทศ', 'NCMN'), ('ใน', 'RPRE'), ('ช่วง', 'NCMN'), ('สุดสัปดาห์', None), ('นี้', 'DDAC'), ('ด้วย', 'RPRE')

Step 3 Named Entity Recognition: In this step it used the ThaiNameTagger module for labeling words. The output in this step will select only the word that was labeled by location. It will be explained the approach in the section B.

Step 4 Finding Important word: For the process of finding important word it used three methods: Degree, Betweenness, and Closeness Centrality to find the important words. After that, it will rank words by score. It described in section C.

Step 5 Majority Voting: The process of finding the most important word is conducted by majority vote model. The word that has the highest score will be important word.

Finally, the output will be the word that has the most important word and can be used to specify the location.

B. Entity Detection

After the data preparation process, the data was sent to the Named Entity method. In this method, Pythainlp will be used in this process. It is developed for using natural language processing in the Thai language. Thai Name Tagger is one of the functions that serviced in the Pythainlp [17]. It used Inside-outside- beginning (IOB) format to tag NER for each word. For instance, a given sentence 'ประยุทธ จันโอชา เป็นนายกรัฐมนตรี', it would be tag the tokens 'ประยุทธ์', 'จันโอชา', 'เป็น', 'นายกรัฐมนตรี' as 'B-PERSON', 'I-PERSON', 'O', and 'O' respectively. The B-prefix indicates the beginning token for a chunk of person name, 'ประยุทธ จันโอชา' and I- prefix indicates the intermediate token. Moreover, the term O indicates that a token not belong to any NER chunk. The results of this process will be displayed in categories including organization, person, location, date and time. For other words that cannot identify to be Named Entity. It will also not use in the next process. It can be showed result of Named Entity in Table I.

TABLE I. MEANING OF ENTITY LABEL

Label	Meaning
B-LOCATION	Beginning of word Place
I-LOCACTION	After of word Place
B-PERSON	Beginning of word Person
I-PERSON	After of word Person
B-ORGANIZATION	Beginning of word Organization
I-ORGANIZATION	After of word Organization
B-DATE	Beginning of word Date
I-DATE	After of word Date
B-TIME	Beginning of word Time
I-TIME	After of word Time

For example จังหวัดเพชรบุรี<LOCATION> as the result from Named Entity process where จังหวัดเพชรบุรีมีฝนตกหนัก is the twitter message. According to this message it can show the result in NER process as จังหวัด <B-LOCATION> เพชรบุรี <I-LOCATION> มี<O> ฝนตกหนัก<O>. After that, the result that was labeled as <LOCATION> will include together such as ประเทศญี่ปุ่น (Japan Country) it will be ประเทศ was labeled as ประเทศ<B-LOCATION> and ญี่ปุ่น was labeled as ญี่ปุ่น <I-LOCATION> after the word was labeled already. It will bring to the word that combine with as one word, for example, ประเทศ<B-LOCATION>, ญี่ปุ่น<I-LOCATION> it will be ประเทศญี่ปุ่น <LOCATION>.

C. Important Words

In this process, Graph theory is used in three methods: Degree, Betweenness, and Closeness Centrality for testing importance of words. Output of this process showed that the word that given the highest score as the most importance or the most influence among the words. After that, the majority voting method will be applied to vote the most important word of three methods. It can be shown the approach as follows:

If Method 1 were A, B, and C
Method 2 were A, C, and B
Method 3 were A, B, and C

So, the orders of important word were A, B, C from majority vote from three methods.

For example of the words that pass the testing of importance score from Degree, Betweenness, and Closeness Centrality can show in the table.

TABLE II. TOP 5 WORDS FOR EACH METHODS

No.	Degree Centrality	Betweenness Centrality	Closeness Centrality
1	Japan	Thailand	Thailand
2	Thailand	Japan	Japan
3	Osaka	Osaka	Tokyo
4	Thai	Thai	Osaka
5	Tokyo	Tokyo	Thai

According to the table, it showed the order of words that had the highest score from the three methods. When all words passed the process of testing the important word. It will be sorted in descending order of score. After that, the word that had the highest score (5 words) will be sent to the majority voting method for finding the most important word.

As you can see in Table II, the most important word from the majority was “Thailand”. The next words were “Japan”, “Osaka”, “Thai”, and “Tokyo”.

IV. EXPERIMENTAL RESULT

In this experiment, it used Name Entity Recognition to label named words. The selected words are only related to location. After that, Graph theory: Degree, Betweenness, and Closeness Centrality were used to compare the important values of words and majority vote was used to select the most important word. The experimental result shows in the table and can explain meaning of the word in the table as follows:

TABLE III. TOP 5 WORDS FROM DEGREE CENTRALITY METHOD

Words Rank	Pronoun	English	Score
ญี่ปุ่น	Yipun	Japan	**0.57**
ประเทศญี่ปุ่น	Prathet Yipun	Japan country	0.29
ไทย	Thai	Thai	0.23
ญี่ปุ่นANA&JAL	Yipun ANA&JAL	Japan ANA&JAL	0.21
ประเทศไทย	Prathet thai	Thailand country	0.13

TABLE IV. TOP 5 WORDS FROM BETWEENNESS CENTRALITY

Words Rank	Pronoun	English	Score
ญี่ปุ่น	Yipun	Japan	**0.55**
ประเทศญี่ปุ่น	Prathet Yipun	Japan country	0.09
ญี่ปุ่นANA&JAL	Yipun ANA&JAL	Japan ANA&JAL	0.09
ไทย	Thai	Thai	0.04
เกาะกวม	Gor guam	Guam island	0.01

TABLE V. TOP 5 WORDS FROM CLOSENESS CENTRALITY

Words Rank	Pronoun	English	Score
ญี่ปุ่น	Yipun	Japan	**0.65**
ประเทศญี่ปุ่น	Prathet Yipun	Japan country	0.53
ญี่ปุ่นANA&JAL	Yipun ANA&JAL	Japan ANA&JAL	0.52
ไทย	Thai	Thai	0.51
ประเทศไทย	Prathet thai	Thailand country	0.47

From Table III-V, the highest score for the first word is ญี่ปุ่น with 0.57, 0.55, and 0.65 for Degree Centrality, Betweenness Centrality, and Closeness Centrality respectively.

The next process will use words that are the result of the three methods to compare the most important word by use majority voting. The result of the experiment can show in Table VI.

TABLE VI. MAJORITY VOTE RESULTS

Degree	Betweenness	Closeness	Majority vote (words)	score (%)
ญี่ปุ่น	ญี่ปุ่น	ญี่ปุ่น	**ญี่ปุ่น**	**100**
ประเทศญี่ปุ่น	ประเทศญี่ปุ่น	ประเทศญี่ปุ่น	ประเทศญี่ปุ่น	100
ไทย	ญี่ปุ่นANA&JAL	ญี่ปุ่น ANA&JAL	ญี่ปุ่นANA&JAL	67
ญี่ปุ่นANA&JAL	ไทย	ไทย	ไทย	67
ประเทศไทย	เกาะกวม	ประเทศไทย	ประเทศไทย	67

From Table VI, it can see that ranking of the important word using majority vote. The word (ญี่ปุ่น) is the word with score of vote 100 % from all methods. So, it can imply that ญี่ปุ่น was the most important word and can be used to indicate a location that had a natural disaster occurred.

V. CONCLUSION

In this paper, presented the utilization of information from twitter in the natural disaster event topic in order to detect the place of event that occurred base on Named Entity Recognition (NER). The place will extract from microblog - Twitter using three techniques: Degree, Betweeness, and Closeness Centrality and then using majority vote for the end of result. In preprocessing step, data only Thai language was collected from twitter using #hagibis in the topic of Super Typhon Hagibis blowing in Japan. Then using three techniques as mention to select only top 5 of words that related to the event. The experimental result shows that the word "Japan, ญี่ปุ่น" is the first word of Degree, Betweenness, Closeness Centrality methods with the score equal 0.57, 0.55, and 0.65 respectively. However, based on experimental result, it shows that the message from twitter can be used to indicate location. By this event: super Typhon Hagibis, ญี่ปุ่น (Japan) is the word that indicated the place from the messages.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Wanapa Phopli and Maleerat Maliyeam conducted the research and data analysis; Choochart Haruechaiyasak and Hathairat Ketmaneechairat corpus technique. All authors wrote the paper and had approved the final version.

ACKNOWLEDGMENT

This paper had been supported by the Graduate College, King Mongkut's University of Technology North Bangkok (KMUTNB), THAILAND.

REFERENCES

[1] N. D. Doulamis, A. D. Doulamis, P. Kokkinos and E. M. Varvarigos, "Event detection in Twitter microblogging," *IEEE Transactions on Cybernetics*, vol. 46, no. 12, pp. 2810-2824, Dec. 2016.

[2] K. Y. Itakura and N. Sonehara, "Using Twitter's mentions for efficient emergency message propagation," in *Proc. International Conference on Availability, Reliability and Security*, Regensburg, 2013, pp. 530-537.

[3] N. Azam, M. Abulaish, and N. A. Haldar, "Twitter data mining for events classification and analysis," in *Proc. Second International Conference on Soft Computing and Machine Intelligence*, Hong Kong, 2015, pp. 79-83.

[4] S. F. Yang. (2016). An event detection approach based on Twitter hashtags. Open Access Theses. [Online]. Available: https://docs.lib.purdue.edu/open_access_theses/909

[5] I. Ahsan, M. A. Ahmed, S. Rehman, M. Abbas, and M. A. Khan, "A novel NLP application to automatically generate text extraction concepts from textual descriptions," in *Proc. the 5th International Conference on Computing and Artificial Intelligence*, New York, NY, USA, 2019, pp. 55-58.

[6] S. A. Al-Ghamdi, J. Khabti, and H. S. Al-Khalifa, "Exploring NLP web APIs for building Arabic systems," in *Proc. Twelfth International Conference on Digital Information Management*, Fukuoka, 2017, pp. 175-178.

[7] P. J. Rani, J. Bakthakumar, B. P. Kumaar, U. P. Kumaar, and S. Kumar, "Voice controlled home automation system using Natural Language Processing (NLP) and Internet of Things (IoT)," in *Proc. Third International Conference on Science Technology Engineering & Management*, Chennai, 2017, pp. 368-373.

[8] R. Agerri, X. Artola, Z. Beloki, G. Rigau, and A. Soroa, "Big data for natural language processing: A streaming approach," *Knowledge-Based Systems*, vol. 79, pp. 36-42, 2015.

[9] K. Shahina, P. Jyothsna, G. Prabha, B. Premjith, and K. P. Soman, “A sequential labelling approach for the named entity recognition in Arabic language using deep learning algorithms,” in *Proc. International Conference on Data Science and Communication*, 2019.

[10] A. Zafarian, A. Rokni, S. Khadivi, and S. Ghiasifard, “Semi-supervised learning for named entity recognition using weakly labeled training data,” in *Proc. The International Symposium on Artificial Intelligence and Signal Processing*, Mashhad, 2015, pp. 129-135.

[11] J. Susymary and R. Lawrance, “Graph theory analysis of protein-protein interaction network and graph based clustering of proteins linked with zika virus using MCL algorithm,” in *Proc. International Conference on Circuit, Power and Computing Technologies*, 2017, pp. 1-7.

[12] R. E. Bacha and T. T. Zin, “Ranking of influential users based on user-tweet bipartite graph,” in *Proc. IEEE International Conference on Service Operations and Logistics, and Informatics*, Singapore, 2018, pp. 97-101.

[13] S. Apreleva and A. Cantarero, “Predicting the location of users on Twitter from low density graphs,” in *Proc. IEEE International Conference on Big Data (Big Data)*, Santa Clara, CA, 2015, pp. 976-983.

[14] J. Golbeck. (2015). *Introduction to Social Media Investigation*. Elsevier. [E-book]. [Online]. Available: https://www.elsevire.com/books/introduction-to-social-media-investigation/golbeck/978-0-12-801656-5

[15] J. Zhang and Y. Luo, “Degree centrality, betweenness centrality, and closeness centrality in social network,” *Advances in Intelligent Systems Research*, vol. 132, pp. 300-303, 2017.

[16] S. Noopataya, S. Thongratsakul, and C. Poolkhet, “Social network analysis of cattle movement in Sukhothai Province, Thailand: A study to improve control measurements,” *Veterinary Medicine International*, p. 587252, 2015.

[17] PyThaiNLPdocumentation. [Online]. Available: https://thainlp.org/pythainlp/docs/2.0/api/tag.html

Wannapa Phopli is Ph.D. candidates at the Department of Information technology and Digital Innovation, King Mongkut’s University of Technology North Bangkok, Thailand. Her research areas are Natural Language Processing, Machine Learning, and Artificial Intelligence.

Maleerat Maliyaem received PhD in Information Technologyfrom King Mongkut’s University of Technology North Bangkok, Thailand. Currently, lecturer at the Faculty of Information Technology, King Mongkut’s University of Technology North Bangkok. Her research areas are Natural Language Processing and Information Retrieval, Machine Learning and Artificial Intelligence.

Choochart Haruechaiyasak received PhD in Computer Science & Engineering from University of Miami, U.S.A.. Currently, researcher at National Electronics and Computer Technology Center (NECTEC), Thailand. His Research areas are Information Retrieval, Information Filtering, Web and Text Mining, Natural Language Processing, Sentiment Analysis and Opinion Mining

Hathairat Ketmaneechairat received PhD in Electrical Engineering from King Mongkut’s University of Technology North Bangkok, Thailand. Currently, lecturer at the College of Industrial Technology, King Mongkut’s University of Technology North Bangkok. Her Research areas are Natural Language Processing and Data Mining, Machine Learning and Artificial Intelligence.