

A Thai-language chatbot analyzing mosquito-borne diseases using Jaccard similarity

Benjamin Chanakot¹, Charun Sanrach²

¹Department of Innovation Management and Business Information, Faculty of Management Technology, Rajamangala University of Technology Srivijaya, Nakhon Si Thammarat, Thailand

²Department of Computer Education, Faculty of Technical Education, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

Article Info

Article history:

Received Feb 18, 2023

Revised Jul 14, 2023

Accepted Aug 2, 2023

Keywords:

Aedes mosquitoes

Chatbot

Jaccard similarity

Natural language processing

Term frequency-inverse

document frequency

ABSTRACT

The objective of this study is to create a Thai-language chatbot analyzing mosquito-borne diseases using Jaccard similarity with an aim to develop an artificial intelligence (AI)-based chatbot used to analyze Aedes-borne diseases through natural language processing. The analysis occurred when the symptoms provided by users on the chatbot were assessed to select relevant words as text attributes using the term frequency-inverse document frequency (TF-IDF) before the Jaccard similarity was used to measure the similarity of the information on the mosquito-borne disease database. The Line Messaging API was applied to facilitate communication between users and the chatbot through the Line application. The chatbot applied PHP 7.2.34 and MySQL 5.7.32 for database management, with Apache 2.2.29 serving as the bot server. The performance evaluation of the chatbot revealed that the chatbot accurately understood user intentions with an intent accuracy of 85.00%. Likewise, the usability of the chatbot was assessed using the system usability scale (SUS), and it received a score of 89.75, indicating a high level of user-friendliness. Furthermore, it has been found that appropriate tokenization enables accurate feature selection. This leads to improved accuracy in measuring Jaccard similarity. Consequently, the chatbot is capable of providing precise responses that align with the user's intent.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Benjamin Chanakot

Department of Innovation Management and Business Information

Faculty of Management Technology, Rajamangala University of Technology Srivijaya

109 Village No.2, Thamyai sub-district, Thungsong District, Nakhon Si Thammarat 80110, Thailand

Email: benjamin.c@rmutsv.ac.th

1. INTRODUCTION

Dengue outbreaks are occurring across the world resulting in increasing numbers of patients, particularly in countries with tropical and subtropical climates [1]. It is a disease caused by the dengue virus which is transmitted by Aedes mosquitoes which accelerate the rapid spread of the disease [2], [3]. In addition, Aedes are also the vector of Zika, Chikungunya, and other virus infections, but dengue exclusively accounts for approximately 390 million dengue cases annually [4]. Thus, dengue is a global health threat with significant social and economic impact. Although the disease is endemic in more than 100 countries in Southeast Asia, the Americas, the Western Pacific, Africa and the Eastern Mediterranean region [5], according to the World Health Organization, it is estimated that between 2.5 and 3 billion people worldwide live in dengue-endemic areas [6], posing a risk of getting dengue. Nowadays, there is no specific treatment

for dengue and its vaccines can only provide symptomatic treatment [7], [8]. Therefore, early detection of the disease allows healthcare professionals to deliver timely medical treatment and it also lowers the mortality rates below 1%. A patient with dengue suffers flu-like symptoms [4] with a high fever of 40 °C or 104 °F and are usually accompanied with at least two of the following symptoms: i) headache; ii) pain behind the eyes; iii) nausea and vomiting; iv) glandular swelling; v) joint, bone or muscle pain; and vi) rash [9].

For patients with severe symptoms, there will be a critical period of about 3-7 days after the onset of the illnesses; the fever will not go away with the symptoms including: i) severe abdominal pain; ii) persistent vomiting; iii) bleeding gums; iv) vomiting blood; v) rapid breathing; and vi) fatigue and restlessness [9]. When critical dengue is suspected, seek medical treatment as soon as possible to prevent risks of plasma leaking, circulatory failure or severe bleeding, shortness of breath, and organ failure [10]. These severe symptoms result in hypovolemic shock and lead to a risk of death [9], when the patient does not receive timely medical treatment.

In Thailand, a country in Southeast Asia, dengue outbreaks occur throughout the nation due to the favorable climate for breeding *Aedes* mosquitoes. As a result, dengue fever poses a significant threat [11], [12] to the health of the population in the country. Furthermore, this illness also impacts the public healthcare system and the nation's economy.

Therefore, the researcher aims to develop a Thai-language chatbot for analyzing *Aedes*-borne diseases using Jaccard similarity. The artificial intelligence (AI)-based chatbot uses natural language processing technology to analyze users messages in order to diagnose diseases by selecting the word features in the text through an integration between the term frequency-inverse document frequency (TF-IDF) method and the Jaccard similarity measurement in the *Aedes*-borne disease database. The development of the chatbot incorporates the Line Messaging API for user-system communication through the Line application. The chatbot is developed using PHP 7.2.34 and utilizes MySQL 5.7.32 for database management, with Apache 2.2.29 serving as the bot server. The contributions of this chatbot development can be beneficial in the fields of medical services and public health, as well as serve as a guideline for enhancing various service sectors in the future.

2. METHOD

2.1. Artificial intelligence

AI refers to technology that mimics human intelligence [13], using sophisticated mathematical algorithms to process data and produce results [14] without requiring a new command for each task. Nowadays, AI, such as chatbots, facial recognition systems, virtual assistants, and more, plays a crucial role in our daily lives by performing various tasks for humans. Due to its ability for automatic learning, AI enables processing, analysis, planning, and decision-making similar to humans.

2.2. Natural language processing

Natural language processing is a branch of AI technology that leverages knowledge from various fields [15], such as linguistics, computer science, and statistics, to analyze the language humans use in daily communication. It enables computers to understand humans' intentions and meanings in their communication. Natural language processing technology can be applied in various domains, including healthcare, education, and business.

2.3. Chatbot

Chatbot is a computer program developed for communicating with humans through naturally occurring everyday language used to communicate in our everyday life. The program simulates conversations with human users the same way they interact with human beings. This program can be used in real-time, 24 hours a day [16], [17]. Chatbots are divided into 2 main types including: i) rule-based chatbots, which is a chatbot developed to process and understand human conversations according to predefined rules and conditions [18]; however, when a question is outside the predefined rules or conditions, the program will not be able to answer it or provide a wrong answer [19], and ii) AI-based chatbot or intelligent bot, a chatbot developed with natural language processing that understands natural language or the language that humans use to communicate in everyday life without relying on any predefined answers [20]. Therefore, the program comprehends the user's intent and provides correct answer to the question.

2.4. Line Messaging API

Line Messaging API is a communication channel between Line application users and service providers. Messages are received and sent between the servers of the parties via the Line Platform, enabling the development of chatbots for interactive messaging with users. This API provides a seamless and efficient means of connecting users with various services and facilitating real-time communication, as shown in Figure 1.

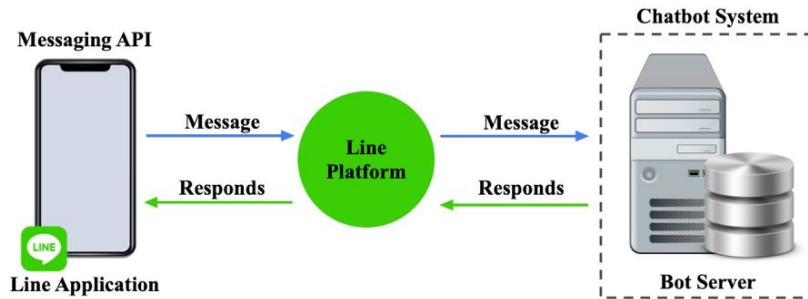


Figure 1. Line Messaging API working processes

2.5. Architecture of the Thai-language chatbot analyzing mosquito-borne diseases using Jaccard similarity

This study aims to develop a chatbot using PHP 7.2.34 and MySQL 5.7.32 to manage the database with Apache 2.2.29 functioning as the bot server. The server and Line Messaging API allow users to communicate via Line application. The architecture of the system is shown in Figure 2. The figure shows the working processes of Line Messaging API, a communication channel between Line application users and chatbot on Line Platform. Bot server must be connected to the Line platform, which transmits data between the server and the Line application users in JSON FORMAT.

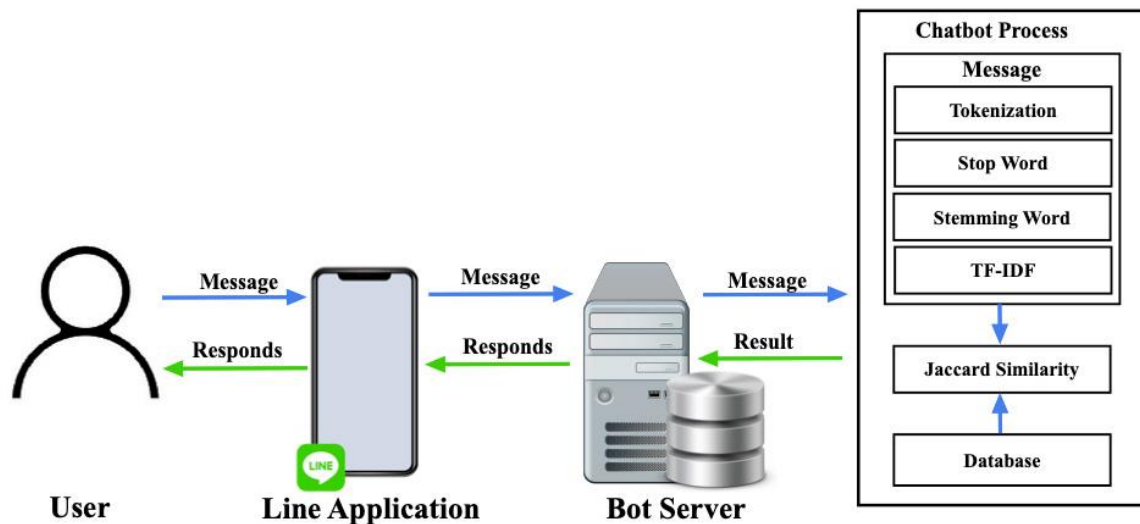


Figure 2. The architecture of the chatbot

2.6. Chatbot development process

This research aims to develop a Thai-language chatbot analyzing mosquito-borne diseases using Jaccard similarity in which the working processes of the chatbot can be divided into 6 steps as:

- Tokenization is the process of splitting a text object into smaller units [21], [22] in order to determine the boundaries. This study applied the longest word pattern matching technique in which the longest matching word in a string is separated by comparing it with the dictionary.
- Stop word is the process of removing insignificant words like prepositions, pronouns, conjunctions, and interjections, from a text. When these words are removed, the meaning will not be affected, and it also helps reduce the size of the text [23]–[25].
- Stemming word is the process of replacing words with the same root or words with the same meaning by the same token [26], [27].
- Feature selection is the process of selecting words that are significant to the text. In this research, feature selection was performed by using TF-IDF.

- e. Similarity measurement is the process of measuring the similarity of users messages through the Jaccard similarity in the Aedes-borne disease database.
- f. Responds is the process of displaying the processing output to users.

2.7. Term frequency-inverse document frequency

TF-IDF is a mathematical algorithm used to calculate the weight of significant words in a text [28], [29] on the assumption that a word appearing frequently in the text is usually with high term frequency (TF), while a low document frequency (DF) can be found when a word does not appear in any other texts, but with a high inverse document frequency (IDF) as seen in (1) for word weight calculation [30], [31].

$$tf_{ij} = \frac{n_{ij}}{\sum_k n_{kj}} \quad (1)$$

where n_{ij} is the frequency of the word t_i in statement D , and $\sum_k n_{kj}$ is the sum of the frequencies of all words appearing in statement D [30].

$$idf_i = \log \frac{|D|}{|\{j: t_i \in d_j\}|} \quad (2)$$

where $|D|$ is the total number of words in the corpus and $|\{j: t_i \in d_j\}|$ is the number of the word t_i appearing in all statements in the corpus [32].

$$TF - IDF = tf_{ij} \times idf_i \quad (3)$$

2.8. Jaccard similarity

Jaccard similarity is a statistical method used to measure similarity or coefficients. The method, used to measure the similarity between sets, is calculated by dividing the intersection of set A and B by the union of both data sets, with the result between 0-1, where 0 refers to no similarities and 1 appears when they are similar [33]–[35]. Jaccard similarity can be calculated through (4):

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (4)$$

2.9. Measuring the chatbot efficiency

Chatbot efficiency measurement refers to the accuracy evaluation of the chatbot to examine whether the interaction with users is effective enough for it to meet the needs or desire of users. In this research, purposive sampling was applied in selecting 10 experts that including 5 people of information technology experts and 5 people of medical specialists. The experts asked 120 questions to measure the chatbot efficiency, which can be calculated through (5) [36]:

$$Intent Accuracy = \frac{Number\ of\ intents\ correctly\ identified}{Total\ number\ of\ intents} \quad (5)$$

2.10. The chatbot usability assessment

The chatbot usability was tested using the system usability scale (SUS), a questionnaire used in evaluating usability of an application developed by John Brooke in 1986 [37] including 10 items as shown in Table 1. Each of them is measured at 5 levels of satisfaction using likert Scale [38] as shown in Table 2. The score for the strongly agree is equal to 5, while the strongly disagree represents the score of 1 [39], [40]. In this research, the 10 experts responding to the questionnaire were purposively selected applying Purposive Sampling. The experts consisted of 5 people of information technology experts and 5 people of medical specialists. The calculation of the score from the questionnaire can be conducted through (6):

$$SUS\ Score = X + Y \times 2.5 \quad (6)$$

In which:

X=the sum of the scores of all odd-numbered questions minus 5.

Y=25 minus the sum of all even-numbered questions.

Table 1. Questions in the chatbot usability assessment form using the SUS

No.	Questions
1.	I would like to use this chatbot often.
2.	I think chatbots shouldn't be unnecessarily complicated.
3.	I think this chatbot is user-friendly.
4.	I think technical support is required to use this chatbot.
5.	I've found many functions of the chatbot work well.
6.	I think some functions of this chatbot are inconsistent.
7.	I think most people can quickly learn to use this chatbot.
8.	I find using this chatbot very cumbersome.
9.	I feel confident using this chatbot.
10.	I need to learn a lot before I can use this chatbot.

Table 2. Description of the score from the SUS calculated from (6)

Scores	Meanings
> 80.3	Excellent usability
68-80.3	Good usability
67	Moderate usability
51-66	Low usability
<51	Limited usability

3. RESULTS AND DISCUSSION

The output of the development of a Thai-language chatbot analyzing mosquito-borne diseases using Jaccard similarity processing is shown in Figure 3. The Figure 3(a) displays the use of Rich menu, the menu that facilitates the users. It also Figure 3(b) outlines an example of a natural language interaction between a user and the chatbot. The intents confusion matrix was employed to evaluate the intent accuracy of Aedes-borne disease analysis in the performance evaluation of the Thai-language chatbot analyzing mosquito-borne diseases using Jaccard similarity. The results are presented in Table 3.

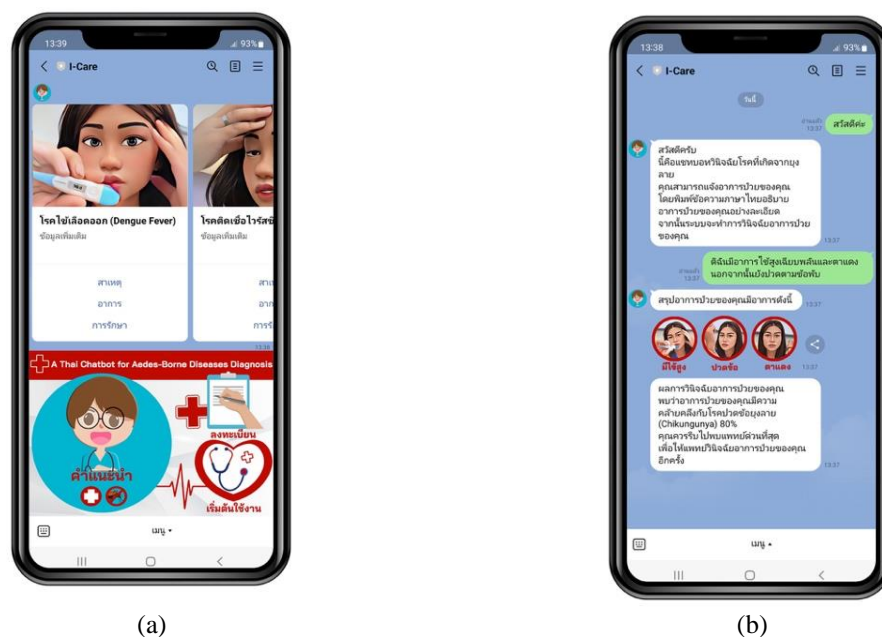


Figure 3. A Thai-language chatbot analyzing mosquito-borne diseases using Jaccard similarity, (a) an example of a Rich menu and (b) an example user-chatbot interaction

Table 3 shows the performance evaluation results of the Thai-language chatbot analyzing mosquito-borne diseases using Jaccard similarity. It can be concluded that the chatbot achieved an intent accuracy of 85.00%. This indicates that the chatbot effectively and accurately analyzes Aedes-borne diseases in line with the users intentions. Figure 4 shows the results of the usability test of the Thai-language chatbot analyzing

mosquito-borne diseases using Jaccard similarity through the SUS. The mean score of the usability was 89.75 referring to excellent usability. This indicates that the chatbot was user-friendly and straightforward. Due to the fact that the respondents were already familiar with using Line application in their daily life, it was easy for them to understand the functionality of the Thai-language chatbot without any need to additionally learn or study how to use the program.

Table 3. The intents confusion matrix of the Thai-language chatbot analyzing mosquito-borne diseases using Jaccard similarity

Intents class	True class			
	Dengue virus	Zika virus	Chikungunya fever	Other
Dengue virus	27	1	1	1
Zika virus	1	26	2	1
Chikungunya fever	2	3	24	1
Other	2	2	1	25

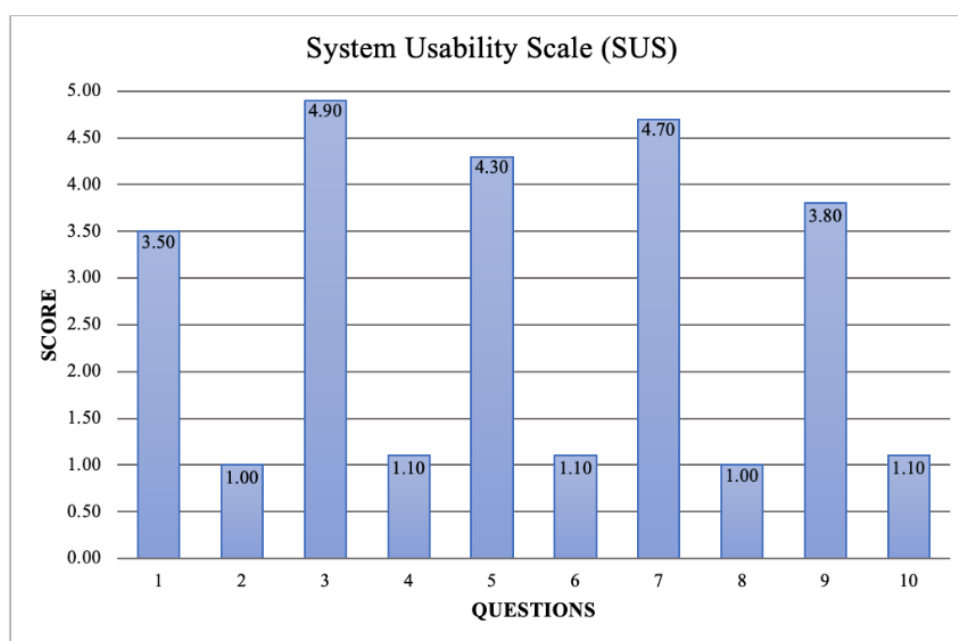


Figure 4. The usability testing result of the Thai-language chatbot analyzing mosquito-borne diseases using Jaccard similarity through the SUS

4. CONCLUSION

This research focuses on develop a Thai-language chatbot analyzing mosquito-borne diseases using Jaccard similarity. It utilizes natural language processing to understand the intentions of users interacting with the chatbot. The selection of text attributes from a text is done using the TF-IDF before the Jaccard similarity is then used to measure similarity against a database of mosquito-borne diseases, providing appropriate responses to users. The chatbot is developed using PHP 7.2.34 and MySQL 5.7.32 for database management in which Apache 2.2.29 operates as a bot server using and incorporates the Line Messaging API for communication with users via the Line application. The research findings indicated that the Thai-language chatbot achieved an intent accuracy of 85.00%, accurately capturing user intentions. The SUS assessment also indicated a high usability score of 89.75, demonstrating that the chatbot was user-friendly.

Therefore, it can be concluded that the Thai-language chatbot, which analyzes mosquito-borne diseases using Jaccard similarity, provides accurate and user-friendly interactions. The research found that the performance of the chatbot in engaging in conversations and providing precise responses aligns correctly with the user's intentions. This performance relies on tokenization to define appropriate boundaries for morphemes, as the Thai language lacks clear word boundaries. Therefore, when appropriate boundaries are assigned to the words, it impacts the accuracy of term weighting for feature selection using TF-IDF and the measurement of Jaccard similarity, resulting in increased accuracy.

ACKNOWLEDGEMENTS

The authors sincerely thank the Office of Disease Prevention and Control Region 11 Nakhon Si Thammarat, Department of Disease Control, Ministry of Public Health for their invaluable guidance and assistance in finding the research information. Finally, we thank the Faculty of Management Technology, Rajamangala University of Technology Srivijaya for the support throughout this study.




REFERENCES

- [1] T. Fourié *et al.*, “Dengue virus type 1 infection in traveler returning from benin to France, 2019,” *Emerging Infectious Diseases*, vol. 26, no. 8, pp. 1946–1949, 2020, doi: 10.3201/eid2608.200055.
- [2] C. H. Huang, Y. T. Tsai, S. F. Wang, W. H. Wang, and Y. H. Chen, “Dengue vaccine: an update,” *Expert Review of Anti-Infective Therapy*, vol. 19, no. 12, pp. 1495–1502, 2021, doi: 10.1080/14787210.2021.1949983.
- [3] C. Shah, G. Vijayaraghavan, and C. C. Kartha, “Spectrum of cardiac involvement in patients with dengue fever,” *International Journal of Cardiology*, vol. 324, pp. 180–185, 2021, doi: 10.1016/j.ijcard.2020.09.034.
- [4] World Health Organization, “Dengue and severe dengue,” *World Health Organization*, 2019. [Online]. Available: <https://www.who.int/news-room/questions-and-answers/item/dengue-and-severe-dengue>, access date: Nov. 09, 2022.
- [5] M. G. Guzman and E. Harris, “Dengue,” *The Lancet*, vol. 385, no. 9966, pp. 453–465, 2015, doi: 10.1016/S0140-6736(14)60572-9.
- [6] W. H. Wang *et al.*, “Dengue hemorrhagic fever – A systemic literature review of current perspectives on pathogenesis, prevention and control,” *Journal of Microbiology, Immunology and Infection*, vol. 53, no. 6, pp. 963–978, 2020, doi: 10.1016/j.jmii.2020.03.007.
- [7] K. Kinoshita *et al.*, “A case of dengue fever that should be considered as imported infectious disease with digestive symptoms,” *Clinical Journal of Gastroenterology*, vol. 14, no. 4, pp. 1115–1120, 2021, doi: 10.1007/s12328-021-01391-w.
- [8] European Centre for Disease Prevention and Control, “Factsheet about dengue,” *European Centre for Disease Prevention and Control (ECDC)*, 2021. [Online]. Available: <https://www.ecdc.europa.eu/en/dengue-fever/facts>, access date: Nov. 07, 2022.
- [9] World Health Organization, “Dengue and severe dengue,” *World Health Organization*, 2022. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>, access date: Nov. 09, 2022.
- [10] R. M. Weerakkody, J. A. Patrick, and M. H. R. Sheriff, “Dengue fever in renal transplant patients: A systematic review of literature,” *BMC Nephrology*, vol. 18, no. 1, pp. 1–6, 2017, doi: 10.1186/s12882-016-0428-y.
- [11] J. T. Lim, B. S. L. Dickens, and A. R. Cook, “Modelling the epidemic extremities of dengue transmissions in Thailand,” *Epidemics*, vol. 33, 2020, doi: 10.1016/j.epidem.2020.100402.
- [12] S. Polwiang, “The time series seasonal patterns of dengue fever and associated weather variables in Bangkok (2003-2017),” *BMC Infectious Diseases*, vol. 20, no. 1, pp. 1–10, 2020, doi: 10.1186/s12879-020-4902-6.
- [13] J. Skrebeca, P. Kalniete, J. Goldbergs, L. Pitkevica, D. Tihomirova, and A. Romanovs, “Modern Development Trends of Chatbots Using Artificial Intelligence (AI),” in *ITMS 2021 - 2021 62nd International Scientific Conference on Information Technology and Management Science of Riga Technical University, Proceedings*, 2021, pp. 1–6, doi: 10.1109/ITMS52826.2021.9615258.
- [14] P. I. Prayitno, R. P. Pujo Leksono, F. Chai, R. Aldy, and W. Budiharto, “Health Chatbot Using Natural Language Processing for Disease Prediction and Treatment,” in *Proceedings of 2021 1st International Conference on Computer Science and Artificial Intelligence, ICCSAI 2021*, 2021, pp. 62–67, doi: 10.1109/ICCSAI53272.2021.9609784.
- [15] L. Ceuca, A. Rednic, and E. S. Chifu, “Safer museum guide interaction during a pandemic and further Using NLP in human interactive museum visits: Museum guide Chatbot,” *Proceedings - 2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing, ICCP 2021*, pp. 313–318, 2021, doi: 10.1109/ICCP53602.2021.9733712.
- [16] I.-C. Hsu and J.-D. Yu, “A medical Chatbot using machine learning and natural language understanding,” *Multimedia Tools and Applications*, vol. 81, no. 17, pp. 23777–23799, 2022, doi: 10.1007/s11042-022-12820-4.
- [17] J. J. Sophia and T. P. Jacob, “EDUBOT-A Chatbot for Education in Covid-19 Pandemic and VQAbot Comparison,” in *Proceedings of the 2nd International Conference on Electronics and Sustainable Communication Systems, ICESC 2021*, 2021, pp. 1707–1714, doi: 10.1109/ICESC51422.2021.9532611.
- [18] M. Jaiwai, K. Shiangjen, S. Rawangyot, S. Dangmanee, T. Kunsuree, and A. Sa-Nguanthon, “Automatized Educational Chatbot using Deep Neural Network,” in *2021 Joint 6th International Conference on Digital Arts, Media and Technology with 4th ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering, ECTI DAMT and NCON 2021*, 2021, pp. 85–89, doi: 10.1109/ECTIDAMTCON51128.2021.9425716.
- [19] T. M. and S. K. S. Meshram, N. Naik, M. VR, “Conversational AI,” in *2021 International Conference on Intelligent Technologies (CONIT)*, 2021, pp. 1–6, doi: 10.1007/978-3-031-02176-3.
- [20] T. H. Nguyen, L. Waizenegger, and A. A. Techatassanasoontorn, “Don’t Neglect the User!” – Identifying Types of Human-Chatbot Interactions and their Associated Characteristics. Springer US, 2021, doi: 10.1007/s10796-021-10212-x.
- [21] J. Chen, J. Li, W. Yao, and D. Chen, “Research on Early Warning Word Segmentation under Domain Corrector Model,” in *Proceedings of 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications, AEECA 2020*, 2020, pp. 842–848, doi: 10.1109/AEECA49918.2020.9213487.
- [22] M. Alksasbeh, T. Abukhalil, B. A. Y. Alqaralleh, and M. Al-Kaseasbeh, “Smart job searching system based on information retrieval techniques and similarity of fuzzy parameterized sets,” *International Journal of Electrical and Computer Engineering*, vol. 11, no. 1, pp. 636–645, 2021, doi: 10.11591/ijece.v11i1.pp636-645.
- [23] S. R. Manalu, “Stop words in review summarization using TextRank,” in *ECTI-CON 2017 - 2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 2017, pp. 846–849, doi: 10.1109/ECTICon.2017.8096371.
- [24] A. Sharaff, H. Shrawgi, P. Arora, and A. Verma, “Document Summarization by Agglomerative nested clustering approach,” in *2016 IEEE International Conference on Advances in Electronics, Communication and Computer Technology, ICAECCT 2016*, 2017, pp. 187–191, doi: 10.1109/ICAECCT.2016.7942580.
- [25] M. Jaiswal, S. Das, and Khushboo, “Detecting spam e-mails using stop word TF-IDF and stemming algorithm with Naïve Bayes classifier on the multicore GPU,” *International Journal of Electrical and Computer Engineering*, vol. 11, no. 4, pp. 3168–3175, 2021, doi: 10.11591/ijece.v11i4.pp3168-3175.
- [26] Gunawansyah, R. Rahayu, Nurwathi, B. Sugiarto, and Gunawan, “Automated essay scoring using natural language processing and text mining method,” in *Proceeding of 14th International Conference on Telecommunication Systems, Services, and*




- Applications, TSSA 2020*, 2020, pp. 1–4, doi: 10.1109/TSSA51342.2020.9310845.
- [27] M. A. Thalor, "A Descriptive Answer Evaluation System Using Cosine Similarity Technique," in *Proceedings - International Conference on Communication, Information and Computing Technology, ICCICT 2021*, 2021, pp. 56–59, doi: 10.1109/ICCICT50803.2021.9510170.
- [28] H. Zhou, "Research of Text Classification Based on TF-IDF and CNN-LSTM," in *Journal of Physics: Conference Series*, 2022, vol. 2171, pp. 218–222, doi: 10.1088/1742-6596/2171/1/012021.
- [29] R. K. Mishra, S. Urolagin, and A. A. J. Jothi, "A Sentiment analysis-based hotel recommendation using TF-IDF Approach," *Proceedings of 2019 International Conference on Computational Intelligence and Knowledge Economy, ICCIKE 2019*, pp. 811–815, 2019, doi: 10.1109/ICCIKE47802.2019.9004385.
- [30] L. Yao, Z. Pengzhou, and Z. Chi, "Research on news keyword extraction technology based on tf-idf and TextRank," *Proceedings - 18th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2019*, pp. 452–455, 2019, doi: 10.1109/ICIS46139.2019.8940293.
- [31] S. Al-Otaibi *et al.*, "Cosine similarity-based algorithm for social networking recommendation," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 2, pp. 1881–1892, 2022, doi: 10.11591/ijece.v12i2.pp1881-1892.
- [32] Y. Yang, "Research and Realization of Internet Public Opinion Analysis Based on Improved TF - IDF Algorithm," *Proceedings - 2017 16th International Symposium on Distributed Computing and Applications to Business, Engineering and Science, DCABES 2017*, pp. 80–83, 2017, doi: 10.1109/DCABES.2017.24.
- [33] M. Alian and A. Awajan, "Arabic sentence similarity based on similarity features and machine learning," *Soft Computing*, vol. 25, no. 15, pp. 10089–10101, 2021, doi: 10.1007/s00500-021-05754-w.
- [34] A. A. Amer and H. I. Abdalla, "A set theory based similarity measure for text clustering and classification," *Journal of Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00344-3.
- [35] E. Da Costa, H. Tjandrasa, and S. Djanali, "Text mining for pest and disease identification on rice farming with interactive text messaging," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 3, pp. 1671–1683, 2018, doi: 10.11591/ijece.v8i3.pp1671-1683.
- [36] K. Deepika, V. Tilekya, J. Mamatha, and T. Subetha, "Jollity Chatbot- A contextual AI Assistant," in *Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020*, 2020, pp. 1196–1200, doi: 10.1109/ICSSIT48917.2020.9214076.
- [37] I. E. Ismail, R. E. Nalawati, and A. Putra, "System Usability Scale and Net Promoter Score on Donation Application of Toddlers Equipment," *Proceedings - 2021 4th International Conference on Computer and Informatics Engineering: IT-Based Digital Industrial Innovation for the Welfare of Society, IC2IE 2021*, pp. 170–174, 2021, doi: 10.1109/IC2IE53219.2021.9649186.
- [38] S. Ratnawati, L. Widianingsih, N. Anggraini, I. M. Shofi, N. Hakiem, and F. E. M. Agustin, "Evaluation of Digital Library's Usability Using the System Usability Scale Method of (A Case Study)," *2020 8th International Conference on Cyber and IT Service Management, CITSM 2020*, pp. 1–5, 2020, doi: 10.1109/CITSM50537.2020.9268801.
- [39] R. S. Pradini, R. Kriswibowo, and F. Ramdani, "Usability Evaluation on the SIPR Website Uses the System Usability Scale and Net Promoter Score," in *Proceedings of 2019 4th International Conference on Sustainable Information Engineering and Technology, SIET 2019*, 2019, pp. 280–284, doi: 10.1109/SIET48054.2019.8986098.
- [40] A. C. Wijaya, M. W. A. Munandar, and F. Utaminigrum, "Usability Testing of Augmented Reality for Food Advertisement Based on Mobile Phone Using System Usability Scale," *Proceedings of 2019 4th International Conference on Sustainable Information Engineering and Technology, SIET 2019*, pp. 266–269, 2019, doi: 10.1109/SIET48054.2019.8986118.

BIOGRAPHIES OF AUTHORS



Benjamin Chanakot    received the M.Sc. Master of Science (Management of Information Technology) from Walailak University, Thailand, and the Ph.D. (Computer Education) from King Mongkut's University of Technology North Bangkok, Thailand. He has been a Head of Innovation Management and Business Information Program, Faculty of Management Technology, Rajamangala University of Technology Srivijaya. He is interested in natural language processing, gesture control, data mining and big data. He can be contacted at email: benjamin.c@rmutsv.ac.th.



Charun Sanrach    is an Assistant Professor, Ph.D. in the Department of Computer Education, Faculty of Technical Education King Mongkut's University of Technology North Bangkok. He received Diplôme de Docteur (Informatique) avec la mention très honorable from INPL, Nancy, France. His research interests include computer education, information technology, apply artificial intelligence in computer education, data mining and bigdata. He can be contacted at email: charun.s@fte.kmutnb.ac.th.