Development of AI Chatbot for Tourism Promotion: A Case Study in Ranong and Chumphon, Thailand

Nasith Laosen
Department of Digital Technology
Faculty of Science and Technology
Phuket Rajabhat University
Phuket, Thailand
nasith.l@pkru.ac.th

Kanjana Laosen
Andaman Intelligent Tourism and
Service Informatics Center
College of Computing
Prince of Songkla University
Phuket Campus
Phuket, Thailand
kanjana.l@phuket.psu.ac.th

Piyawat Sriprasert
Computer Engineering
Institute of Engineering Suranaree
University of Technology
Nakhon Ratchasima, Thailand
d6200534@g.sut.ac.th

***Abstract*—Ranong and Chumphon provinces in Thailand possess significant tourism potential due to their natural beauty and archaeological heritage. However, tourism development in the region remains limited due to inadequate information infrastructure and limited technology integration. This research presents a chatbot designed to promote tourism in these locations. The chatbot leverages a graph database built using data from the University to Tambon (U2T) project. Functioning as a virtual travel companion, it provides information on tourist attractions, accommodations, and dining options. Advanced natural language processing and machine learning enable the chatbot to understand user inquiries and effectively retrieve relevant information. Evaluation demonstrates strong performance with an F1-score of 99.56 for intent classification and 99.96 for named entity recognition. Testing with user-designed test cases confirmed the chatbot's functionality.**

***Keywords*—chatbot, intent classification, named entity recognition, tourism promotion, graph database**

I. Introduction

Thailand recognizes the tourism industry's potential to generate revenue and distribute income, and consequently prioritizes tourism diversification within its national strategic plan (2018-2037) to enhance competitiveness. This focus includes a strategic emphasis on creative and cultural tourism, which leverages Thailand's distinctive history, art, culture, and traditions to align with tourist preferences.

Ranong and Chumphon provinces, while lesser-known tourist destinations, possess unspoiled natural beauty and a rich history. Their strategic location on a narrow strip of land reveals archaeological evidence of ancient trade routes linking significant global regions since Suvarnabhumi era. Modern plans for a Land Bridge trade route across this isthmus further highlight the area's potential. However, integrated tourism development connecting the two provinces is still in its early stages. Challenges include inadequate infrastructure for information management and limited use of modern technologies for tourism promotion. Consequently, advancing tourism within this region remains an ongoing endeavor.

The rapid evolution of digital technologies and artificial intelligence offers opportunities for innovation in the tourism sector. Recognizing this potential, this research proposes a chatbot system to invigorate tourism within the Ranong and Chumphon provinces. This chatbot functions as a virtual travel companion, addressing tourist inquiries regarding attractions, accommodations, and dining options. The proposed chatbot leverages a graph database as its knowledge base, utilizing data sourced from the University to Tambon (U2T) project [1]. By employing state-of-the-art natural language processing techniques, the chatbot identifies named entities and user intent within questions. This enables effective querying of the graph database to retrieve relevant information. The main contributions of the proposed chatbot cloud be summarized as follows: (1) streamlining information access for tourists, (2) enhancing their travel experiences, and (3) promoting tourism within the target region.

The remaining sections of this paper are structured as follows: Section II reviews relevant literature. Section III details the construction of the U2T graph database. Section IV outlines the chatbot architecture and describes development of its components. Section V presents and discusses testing results. Section VI provides conclusion and future work.

979-8-3503-8176-4/24/$31.00 ©2024 IEEE

II. Literature Review

Linked data and graph databases have emerged as a prominent technique within chatbot systems due to their scalability and inherent flexibility. The core development process for chatbots of this type typically encompasses the following stages: (1) named entities and user intent are determined from natural language questions; (2) questions are transformed into a structured query language; (3) relevant information is retrieved from a graph database; (4) the retrieved answer is presented to the user, potentially augmented with additional phrasing for naturalness.

Examples of chatbots leveraging linked-data structures include the TPMAP Bot [2], KBot [3], and OntBot [4] systems. TPMAP Bot facilitated natural language searches of Thai population data stored within a graph database. It employed data augmentation for dataset creation, training machine learning models for entity extraction and intent classification. Queries were generated from identified entities and question types to retrieve data. The data was then presented in both textual and graphical formats. KBot integrated Semantic Web principles, knowledge graphs, and machine learning for versatile chatbot functionality. It enabled analytical data searches and provided concise knowledge cards in response to frequently asked questions. Search results and user interactions were stored for continuous learning and improvement. OntBot leveraged an ontology-based approach for chatbot development. Utilizing mapping techniques, ontological knowledge was structured within a relational database to power chatbot interactions.

III. U2T Graph Database Construction

To facilitate comprehensive access to local businesses for tourists, the authors leveraged Thailand Community Big Data (TCD), collected under the University to Tambon (U2T) project, as the basis for the chatbot's knowledge base. The U2T graph database construction process involved the following steps: (1) Ranong and Chumphon provincial data were acquired from the U2T website; (2) Excel files containing tourist attraction, accommodation, and restaurant data underwent cleaning and transformation into CSV format; (3) a graph data structure was designed to optimize storage within the database; (4) a Python program was developed to parse the CSV documents, align the data with the designed graph structure, and ultimately store this representation within an ArangoDB graph database. The structure of the resulting graph database is depicted in Fig. 1.

From the database's structure and available data, we analyzed and identified two possible data inquiry patterns:

- *Entity Retrieval by Location*: Users can query for the names of restaurants, accommodations, and tourist attractions within specified administrative boundaries (i.e., village, subdistrict, district, or province) in the Ranong-Chumphon region.
- *Detailed Entity Information*: Users can retrieve comprehensive details associated with individual restaurants, accommodations, and tourist attractions located within Ranong and Chumphon provinces. These details may include images, coordinates, addresses, operating hours, and contact information.

These patterns serve as the foundation for the chatbot development, dictating its primary functionalities.

IV. Chatbot Design and Development

A. Chatbot Architecture

Fig. 2 shows the architecture of the proposed chatbot. From the figure, the messaging channel functions as the intermediary for communication and message exchange between the human user and the chatbot. This research leverages the LINE platform as the designated messaging channel due to its widespread adoption within the Thai population, ensuring maximal accessibility for the target user base. The question-answering process begins with processing a user question through an intent classification model. This model categorizes the question into a question type, e.g., inquiries about tourist attractions or restaurant locations. Each question type is associated with a predefined query command (PQC) template. The system then employs a named entity recognition (NER) model to identify named entities, e.g., subdistrict names, tourist attraction names, and restaurant names, presenting within the question. Subsequently, a query command is generated by integrating the named entities identified by the NER model into the PQC template associated with the question type categorized by the intent classification model. The generated query command is then executed to retrieve relevant information from the U2T graph database. This information is finally presented to the user as the chatbot's response.

B. Intent Classification Model Development

As mentioned in the previous section, an intent classification model was utilized to identify question types of user questions. The development process of this intent classification model comprised two primary stages: dataset preparation and model construction. Specific details concerning these stages are outlined below.

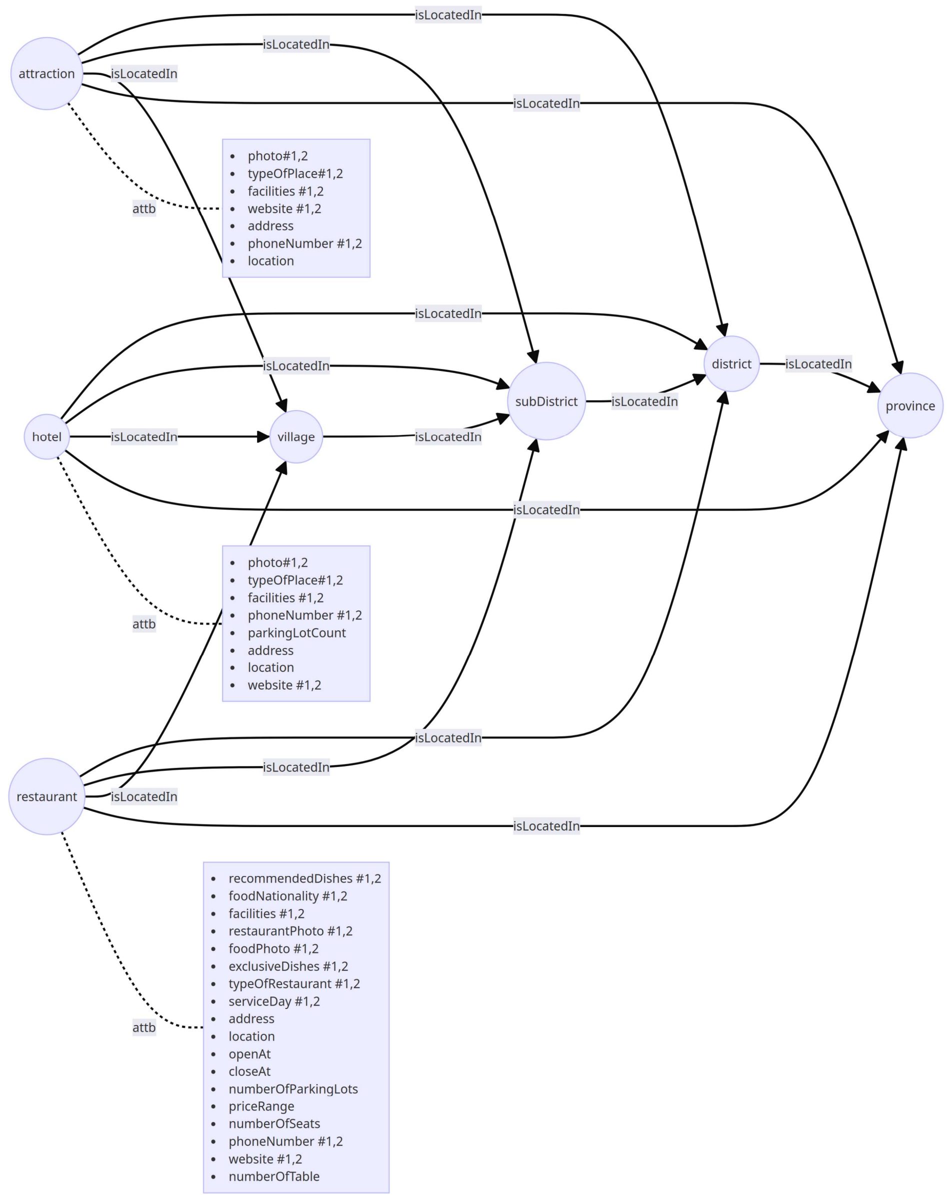


Fig. 1. The structure of the U2T graph database.

1) Intent Classification Dataset Preparation: The authors employed the data augmentation technique to create the intent classification dataset. Data augmentation involves creating diverse, synthetic data samples by applying transformations to existing data while preserving semantic integrity. This technique expands the dataset's size and diversity, potentially leading to improved classification performance [5]. To create the dataset, question types (classes) were first defined based on the possible data inquiry patterns mentioned in Section III. Table I shows the question types resulting from this step. Subsequently, question templates were created for each defined question type. Examples of these question templates are presented in Table II. Data samples were finally generated by substituting placeholders within the templates with relevant entities (e.g., sub-district, district, province, hotel, restaurant, and tourist attraction names) retrieved from the U2T graph database. Illustrative examples of the data samples resulting from this step are shown below:

- ชุมพรมีที่เที่ยวอะไรบ้าง (What attractions does Chumphon have?)
- ขอรายชื่อสถานที่ท่องเที่ยวในระนอง (Let me see the list of tourist attractions in Ranong.)
- ขอดูรูปมัลลิการีสอร์ท (Let me see the pictures of Mallika Resort.)

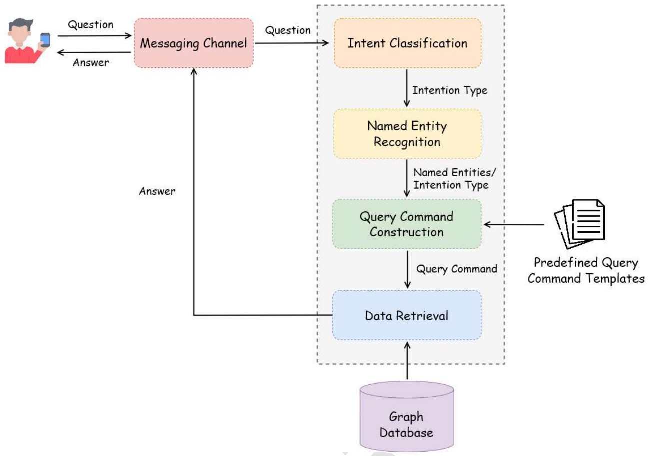


Fig. 2. The architecture of the proposed chatbot.

TABLE I
QUESTION TYPES (CLASSES) FOR INTENT CLASSIFICATION

No.	Class Name
Entity Retrieval by Location	
1	AskForAttractionListInArea
2	AskForHotelListInArea
3	AskForRestaurantListInArea
4	AskForAttractionListByTypeInArea
5	AskForHotelListByTypeInArea
6	AskForRestaurantListByTypeInArea
Detailed Entity Information	
7	AskForAttractionAddress
8	AskForHotelAddress
9	AskForRestaurantAddress
10	AskForAttractionContactInfo
11	AskForHotelContactInfo
12	AskForRestaurantContactInfo
13	AskForAttractionCoordinates
14	AskForHotelCoordinates
15	AskForRestaurantCoordinates
16	AskForAttractionPictures
17	AskForHotelPictures
18	AskForRestaurantPictures
19	AskForAttractionType
20	AskForHotelType
21	AskForRestaurantType
22	AskForRestaurantOpenHour
23	AskForRestaurantOpenDay
24	AskForRestaurantFoodPictures
25	AskForRestaurantFoodTypes
26	AskForRestaurantRecommendedDishes
27	AskForRestaurantSpecialMenu

The data augmentation technique yielded a dataset of 211,761 data samples, which were subsequently utilized to train the intent classification model.

2) Intent Classification Model Construction: The authors fine-tuned the XLM-RoBERTa (XLM-R) large language model (LLM) [6] to construct the intent classification model. The architecture of the XLM-R model adapted for this downstream task is depicted in Fig. 3. The fine-tuning was performed on the prepared dataset with a data split of 80% for training, 20% for validation, and 20% for testing. The input text was tokenized into subword units (sentence pieces), and the special token [CLS] marked the beginning of each sequence. In cases where sentence pairs were present, a [SEP] token was employed as a delimiter to differentiate between the sentences. The model was trained for three rounds with the objective of minimizing cross-entropy loss, achieving an F1-score of 99.56.

TABLE II
EXAMPLES OF QUESTION TEMPLATES

Question Type	Question Templates
AskForAttractionListInArea	{area_name} มีที่เที่ยวอะไรบ้าง (What attractions does {area_name} have?)
	สถานที่ท่องเที่ยวใน{area_name} (Attractions in {area_name})
	อยู่ {area_name} ไปไหนดี (Where should I go in {area_name}?)
AskForHotelPictures	ขอดูรูป {hotel_name} (Let me see the pictures of {hotel_name})
	ตัวอย่างรูปของ {hotel_name} (Example pictures of {hotel_name})

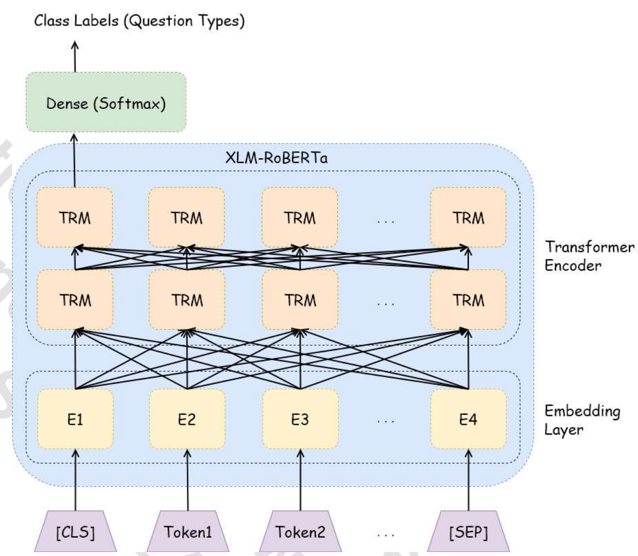


Fig. 3. The architecture of the XLM-RoBERTa model for intent classification.

TABLE III
BIO Labels (Classes) for Named Entity Recognition

No.	Class Name	Description
1	B-AreaName	Beginning token of an area name
2	I-AreaName	Inside token of an area name
3	B-PlaceName	Beginning token of a place name
4	I-PlaceName	Inside token of a place name
5	B-RestaurantName	Beginning token of a restaurant name
6	I-RestaurantName	Inside token of a restaurant name
7	B-HotelName	Beginning token of a hotel name
8	I-HotelName	Inside token of a hotel name
9	B-PlaceType	Beginning token of a place type
10	I-PlaceType	Inside token of a place type
11	B-HotelType	Beginning token of a hotel type
12	I-HotelType	Inside token of a hotel type
13	B-RestaurantType	Beginning token of a restaurant type
14	I-RestaurantType	Inside token of a restaurant type
15	O	Out-of-Scope token

TABLE IV
Examples of the NER Dataset

<table>
<tr><th>Question</th><th>Tokens</th><th>Labels</th></tr>
<tr><td>ผาเปิดใจอยู่ในจังหวัดอะไร</td><td>ผา|เปิดใจ|อยู่|ใน|จังหวัด|อะไร</td><td>B-PlaceName|I-PlaceName|O|O|O|O</td></tr>
<tr><td>ขอดูรูปน้ำตกสวนจันทร์</td><td>ขอ|ดู|รูป|น้ำตก|สวน|จันทร์</td><td>O|O|O|B-PlaceName|I-PlaceName|I-PlaceName</td></tr>
<tr><td>ขอพิกัดของมูนไชน์รีสอร์ท</td><td>ขอ|พิกัด|ของ|มูน|ไช|น์|รีสอร์ท</td><td>O|O|O|B-HotelName|I-HotelName|I-HotelName|I-HotelName</td></tr>
<tr><td>ขอเบอร์ติดต่อกาแฟโบราณบ้านเถ้าแก่</td><td>ขอ|เบอร์|ติดต่อ|กาแฟ|โบราณ|บ้าน|เถ้าแก่</td><td>O|O|O|B-RestaurantName|I-RestaurantName|I-RestaurantName|I-RestaurantName</td></tr>
</table>

C. Named Entity Recognition Model Development

The authors developed a NER model to facilitate the extraction of entity names from user questions. Similar to the intent classification model, the NER model development process involved two primary stages: preparation of a suitable dataset and subsequent construction of the model.

1) Named Entity Recognition Dataset Preparation: The authors utilized the data samples obtained from the data augmentation technique (detailed in the previous section) to create the NER dataset. Data samples was first tokenized using the PyThaiNLP library [7], facilitating subsequent token-level classification. Each token was then assigned a respective label according to the BIO format outlined in Table III. Table IV provides illustrative examples of the labeled NER dataset.

2) Named Entity Recognition Model Construction: The authors fine-tuned the WanchanBERTa LLM [8] to construct the NER model. The architecture of the WanchanBERTa model adapted for this NER task is depicted in Fig. 4. Consistent with the intent classification model's processing, the [CLS] and [SEP] tokens were appended to demarcate sentence boundaries. However, in contrast to predicting a single class for the entire input, the NER model assigns class labels for each individual tokens. The NER model was trained with an 80/20/20 split of the dataset into training, validation, and testing sets, respectively, achieving an F1-score of 99.96 after five rounds of training.

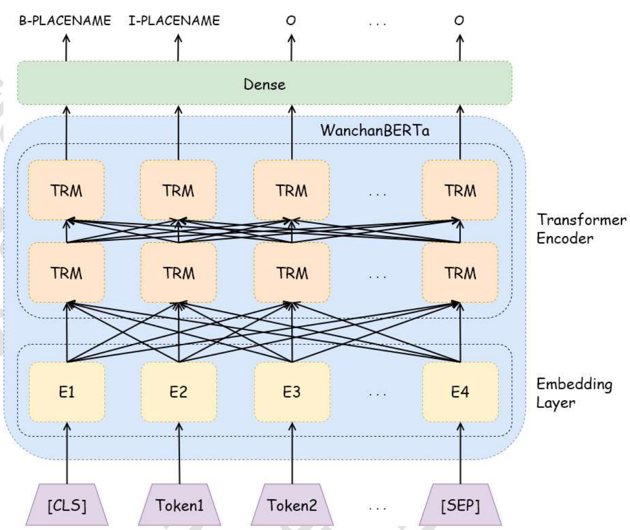


Fig. 4. The architecture of the WanchanBERTa model for named entity recognition.

D. Predefined Query Command Templates and Query Command Construction

The authors constructed 27 predefined query command templates using the ArangoDB query language (AQL). This number directly corresponds to the 27 question types in Table I, ensuring a one-to-one mapping between question types and query command templates. Illustrative examples of these query command templates (implemented as functions in the Python language) are shown in Fig. 5, Fig. 6, and Fig. 7.

To query data, the system constructs a query command by filling the recognized named entities (identified by the NER model) into the query command template of the question type (categorized by the intent classification model). This finalized query command is then executed to retrieve data from the U2T graph database.

```
def query_attractions_in_area(area_name):
  cursor = AIMHTDB.aql.execute(
    f"""FOR place IN A_placeName
        FILTER place.address.province == "{area_name}" ||
        place.address.district == "{area_name}" ||
        place.address.subDistrict == "{area_name}" ||
        place.address.village == "{area_name}" ||
        place.address.street == "{area_name}"
        SORT place.numberOfTouristPerYear DESC
            RETURN place
    """,
        batch_size=65535)
  result = [doc for doc in cursor]
  return result
```

Fig. 5. The predefined query command template for retrieving attractions in an area.

```
def query_accomodation_in_area(area_name):
  cursor = AIMHTDB.aql.execute(
    f"""FOR accommodation IN P_placeName
        FILTER accommodation.address.province == "{area_name}" ||
        accommodation.address.district == "{area_name}" ||
        accommodation.address.subDistrict == "{area_name}" ||
        accommodation.address.village == "{area_name}" ||
        accommodation.address.street == "{area_name}"
            RETURN accommodation
    """,
        batch_size=65535)
  result = [doc for doc in cursor]
  return result
```

Fig. 6. The predefined query command template for retrieving hotels in an area.

V. Testing Results and Discussion

In addition to evaluating performance of the NER and intent classification models (described in Section IV), the authors also evaluated the proposed chatbot by using test cases. These test cases and their expected results are shown in Table V. They were designed based on the question types presented in Table I. During the testing process, the chatbot was tested with various test data derived from the test cases. Testing results indicated that the chatbot operates effectively, meeting the specified criteria and fulfilling all functions. However, three key areas for potential improvement were also revealed. First, expanding the training dataset for the machine learning models would enhance the system's ability to recognize diverse question syntax. While current models demonstrate flexibility to some extent, accommodating all writing variations remains challenging. Second, incorporating techniques such as advanced typo correction mechanisms or speech-to-text input would mitigate the system's sensitivity to misspellings, particularly in proper nouns that directly impact database queries. Finally, extending system capabilities to handle complex data searches (e.g., accommodations filtered by specific features or graphical result displays) would significantly enhance its utility. Fig. 8 provides screen captures of the completed chatbot.

```
def query_pictures_of_attraction(place_name):
  cursor = AIMHTDB.aql.execute(
    f"""FOR place IN A_placeName
      FILTER place.placeName == "{placeName}"
        FOR rel IN REL_A_placeName_photo
          FILTER rel._from == place._id
            FOR photo IN A_photo
              FILTER rel._to == photo._id
                RETURN photo.googleBucket
    """,
        batch_size=65535)
  result = list(cursor)
  return result
```

Fig. 7. The predefined query command template for retrieving pictures of an attraction.

VI. Conclusion

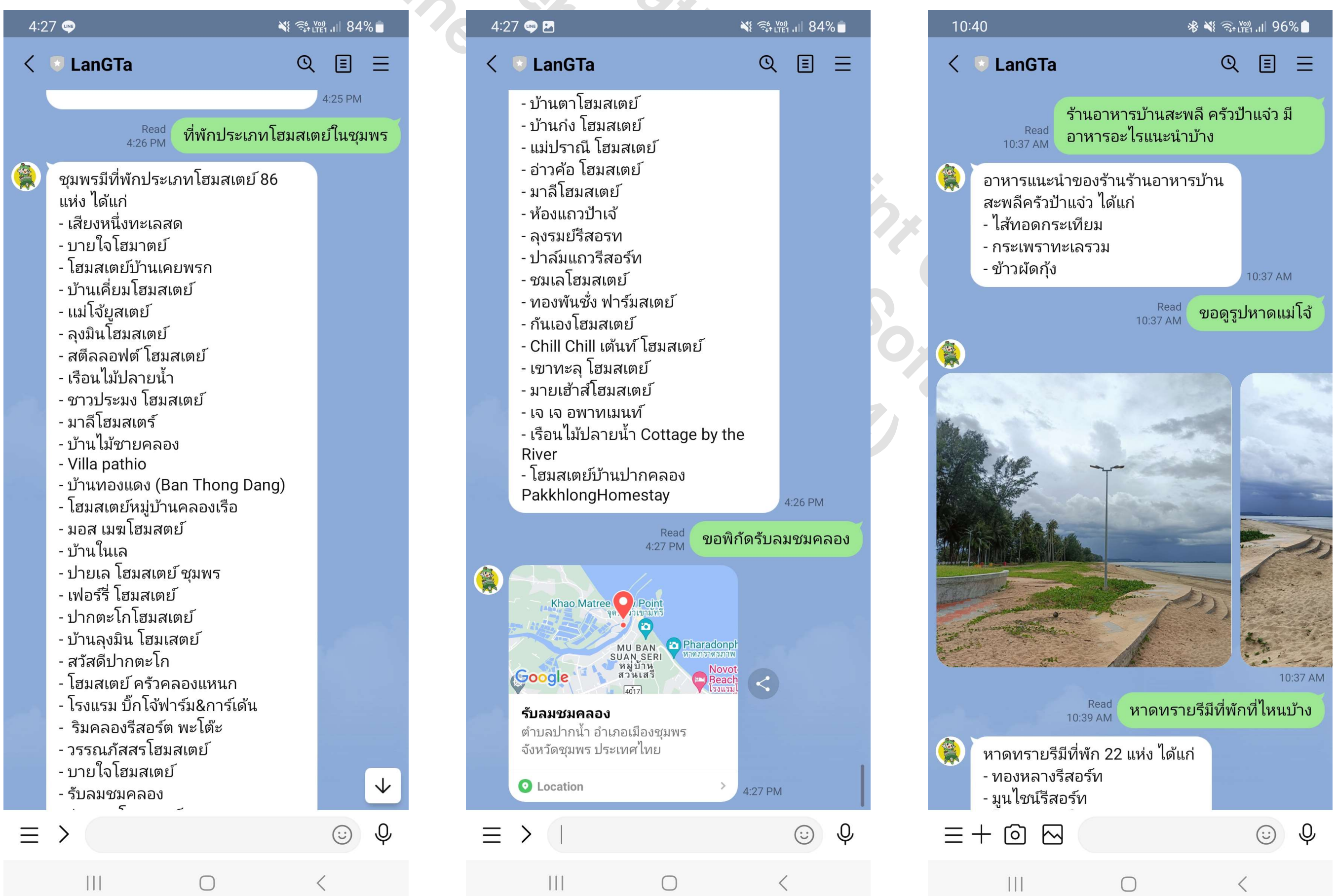
This paper presents a LINE chatbot designed to enhance tourism experiences within the Ranong and Chumphon provinces. The chatbot facilitates information searches on accommodations, restaurants, and tourist attractions, leveraging the Thailand Community Big Data (TCD) under the University to Tambon (U2T) project as its knowledge base. The chatbot development process integrated various artificial intelligence methodologies. Specifically, natural language processing and machine learning techniques were employed to enable the chatbot to interpret question intent and extract named entities for querying a graph database.

System testing with diverse test cases indicated satisfactory performance within the defined scope and functionality. However, evaluation suggests potential improvements. This includes expanding the machine learning model's training data to recognize a wider range of sentence patterns, enhancing the system's flexibility and accuracy. Additionally, incorporating mechanisms to handle misspelling and typos, particularly in proper nouns crucial for database queries, would further mitigate errors and promote a more seamless user experience.

The proposed chatbot serves as a prototype, demonstrating the application of AI technologies for tourism promotion. This research offers a model with potential for adaptation and expansion to regions beyond Ranong and Chumphon. Future work will focus on integrating historical textual data from LSTPedia [9] and a generative AI model, e.g., OpenThaiGPT [10], to promote historical tourism in Ranong and Chumphon, enabling even more engaging and informative natural conversations with the chatbot.

TABLE V
TEST CASES

No.	Test Case Description	Expected Result
1	Users inquire about restaurants in an area (subdistrict, district, or province).	The system shows a list of restaurants in the specified area.
2	Users inquire about accommodations in an area (subdistrict, district, or province).	The system shows a list of accommodations in the specified area.
3	Users inquire about tourist attractions in an area (subdistrict, district, or province).	The system shows a list of tourist attractions in the specified area.
4	Users ask for pictures of a place (restaurant, accommodation, or tourist attraction).	The system shows pictures of the specified place.
5	Users ask for the coordinates of a place (restaurant, accommodation, or tourist attraction).	The system shows the coordinates of the specified place.
6	Users ask for the address of a place (restaurant, accommodation, or tourist attraction).	The system shows the address of the specified place.
7	Users ask for contact information of a place (restaurant, accommodation, or tourist attraction).	The system shows the contact information of the specified place.
8	Users ask for a restaurant's recommended menu items.	The system shows recommended menu items of the specified restaurant.
9	Users ask about another topic or uses a sentence that the system is not familiar with.	The system shows suggestions for transforming sentences.



(a) (b) (c)

Fig. 8. Chatbot's screen captures.

Acknowledgment

This research is supported by (i) National Research Council of Thailand (NRCT) under the "Artificial Intelligent Platform Development for Metaverse Historical Tourism: AI-MHT" and (ii) Super AI Engineer season III program.

References

[1] Ministry of Higher Education, Science, Research and Innovation, "University to Tambon (U2T)," 2021. [Online]. Available: https: //www.u2t.ac.th/

[2] T. Simud, S. Ruengittinun, N. Surasvadi, N. Sanglerdsinlapachai, and A. Plangprasopchok, "A conversational agent for database query: A use case for thai people map and analytics platform," in *Proceedings of the 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, 2020, pp. 155–160.

[3] A. Ait-Mlouk and L. Jiang, "KBot: A knowledge graph based chatbot for natural language understanding over linked data," *IEEE Access*, vol. 8, pp. 149 220–149 230, 2020.

[4] H. Al-Zubaide and A. A. Issa, "OntBot: Ontology based chatbot," in *Proceedings of the International Symposium on Innovations in Information and Communications Technology*, 2011, pp. 7–12.

[5] N. Lhasiw, T. Tanantong, and N. Sanglerdsinlapachai, "Thai conver-sational chatbot classification using BiLSTM and data augmentation," in *Data Science and Artificial Intelligence*, C. Anutariya and M. M. Bonsangue, Eds. Springer Nature Singapore, 2023, pp. 127–141.

[6] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," 2020. [Online]. Available: https://arxiv.org/abs/1911.02116

[7] W. Phatthiyaphaibun, K. Chaovavanich, C. Polpanumas, A. Suriyawongkul, L. Lowphansirikul, P. Chormai, P. Limkonchotiwat, T. Suntorntip, and C. Udomcharoenchaikit, "PyThaiNLP: Thai natural language processing in Python," in *Proceedings of the 3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS 2023)*, Dec. 2023, pp. 25–36. [Online]. Available: https://aclanthology.org/2023.nlposs-1.4

[8] L. Lowphansirikul, C. Polpanumas, N. Jantrakulchai, and S. Nutanong, "WangchanBERTa: Pretraining transformer-based Thai language models," 2021. [Online]. Available: https://arxiv.org/abs/2101.09635

[9] National Electronics and Computer Technology Center (NECTEC), "LSTPedia," 2024. [Online]. Available: https://lstpedia.org/aimht

[10] Artificial Intelligence Entrepreneur Association of Thailand (AIEAT), "OpenThaiGPT," 2023. [Online]. Available: https://openthaigpt.aieat.or. th/