

Rule-based Myanmar Language Chatbot for Travel and Tourism Domain

Htet Htet Yadanar Oo¹, May Thae Sue², Ye Kyaw Thu^{3,4}, Hnin Aye Thant¹, Nang Aye Aye Htwe²

¹Department of Information Science, University of Technology (Yatanarpon Cyber City), Myanmar

²Department of Computer Engineering and Information Technology, Mandalay Technological University, Myanmar

³National Electronics and Computer Technology Center (NECTEC), Thailand

⁴Language Understanding Lab., Myanmar

htetyadana123@gmail.com, misue.mts@gmail.com, ka2pluskha2@gmail.com, hninayethant@utycc.edu.mm,
nangayeayhtwe@mtu.edu.mm

Abstract - Chatbot is a program that generates response based on given input to emulate human conversations in text or voice mode. This proposed system is Myanmar language (Burmese) chatbot for travel and tourism domain. This chatbot can help the users to suggest where to visit, show where are recommended places and answer what the users want to know about places to be visited. Users can also use this chatbot to find hotels, foods to be eaten and types of transportation around the user's location. This chatbot only supports Myanmar Language since it aims for the local people. This paper shows how a rule-based chatbot can be created artificially by using Regular Expression, Damerau-Levenshtein distance and Web scraping.

Keyword - Rule-based chatbot, Regular Expression, Damerau-Levenshtein distance, Web scraping

I. INTRODUCTION

A chat dialogue system or a conversational agent is a computer program designed to hold conversations in a human-like way and ideally "understand" the user's intent. These applications are designed to simulate human-human interactions. Their functioning can range from customer service, product suggestion, product inquiry to personal assistant [1]. Many of these chat agents are built using rule-based techniques, retrieval techniques or simple machine learning algorithms. Many dialogue systems are rule-based, and one of the most popular mechanisms of representing rules is AIML (Artificial Intelligence Markup Language). In retrieval based techniques, chat agents scan for keywords within the input phrase and retrieve relevant answers based on the query string. They rely on keyword similarity and retrieved text is pulled from internal or external data sources including world wide web or organizational database. Some other advanced chatbots are developed with Natural Language Processing (NLP) techniques and machine learning algorithms. Also, there are many commercial chat engines available, which help in building chatbots based on client data input. Recently there have been major increase or interest in use and deployment of dialogue generation systems. Many major technical companies are using virtual assistant or chat agent to fill the needs of customers. Some of them include Google's Google Assistant, Microsoft's Cortana and Amazon's Alexa.

The purpose of this system is to create a chatbot for helping and giving suggestions and recommendations while travelling around Myanmar. This chatbot can help the users to suggest where to visit and advise where are

recommended places for them. The users do not need to wait for the tourism services and they can go anywhere they want to. Some people want to visit with their family and they cannot decide where to go, at that time they can get suggestions from this chatbot. The main aim of this proposed system is to develop a chatbot prototype for travel and tourism in Myanmar. This chatbot system uses rule-based approach to analyze the places for the users' interested activities. In this system, Regular Expression is used for redirecting the chatbot's search direction to which city the user asked. Damerau-Levenshtein distance algorithm is used to measure between the user's input sentence and the sentence in our dataset we prepared. Web scraping is used for extracting webpages from Google, YouTube and Facebook.

II. RELATED WORK

Eliza is considered as the first chatbot which works on the pattern matching system. It is developed by Joseph Weizenbaum et. al. in 1964 [2]. ALICE is a rule-based chatbot based on the Artificial Intelligence Markup Language (AIML). It has more than 40,000 categories, where each category has a combination of pattern and its response.

Aung Myint Myat Kyaw et. al. [5] proposed a chatbot for the tourism industry in Myanmar. This research was developed by a chatbot based on travel Ontology System. This chatbot system is implemented on the Facebook messenger. It only supports English language. Dialogflow that can understand natural language that conversation with the user is used. The Ontology system is developed using Protégé which stored the knowledge while the dialog manager is handled using Dialogflow (API.ai).

Iulia Nica et. al. [4] presented the underlying methods and technologies behind a chatbot for e-tourism that allows people textually communicate with the purpose of booking hotels, planning trips, and asking for interesting sights worth being visit. This paper showed how model-based reasoning can be used for enhancing user experience during a chat.

III. CHATBOT SYSTEM ARCHITECTURE

In this section, we describe the detailed system architecture of Myanmar Language (Burmese) chatbot system. We show an overview of the chatbot system architecture in Fig. 1.

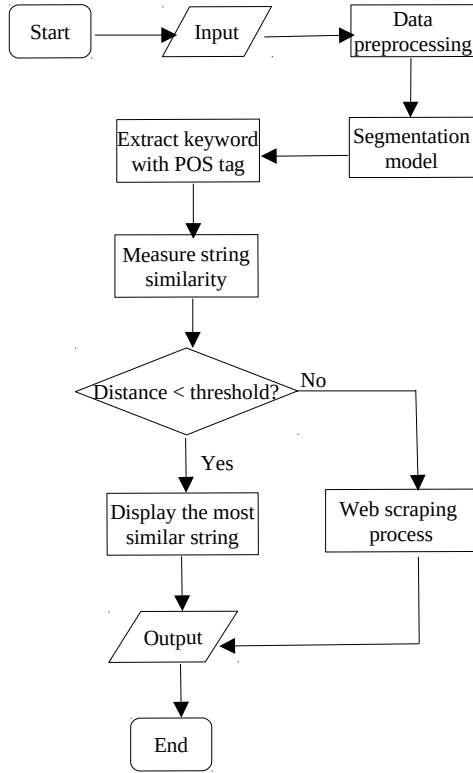


Fig. 1 An overview of Myanmar Language Chatbot system architecture

A. Regular Expression

A regular expression is a special sequence of characters that helps us match or find other strings or sets of strings, using a specialized syntax held in a pattern [13]. The module **re** provides full support for regular expressions in Python. We built regular expression patterns with python **re** module [3]. For example: this Regular Expression pattern `(.*မွန်ပြည်နယ်.*ဘုရား.*|.*စေတီ.*)` means that a sentence has these words (မွန်ပြည်နယ် and ဘုရား or စေတီ) and the other words are any characters. We first analyzed the meaning of user's input sentence and handled this input sentence is concerned with which database. In this system, we separated all dataset of Mon State into categories such as pagoda, beach, restaurant, hotel, transportation, etc of each place. In each category database, it has parallel sentences of corresponding categories. If the user asks question (မွန်ပြည်နယ် မှာ ရှိတဲ့ ဘုရားတွေကို သိချင်ပါတယ်။), the chatbot can guess this question is concerned with the pagodas of Mon State according to the above RE pattern `(.*မွန်ပြည်နယ်.*ဘုရား.*|.*စေတီ.*)` and can handle to search in the pagoda database of Mon State.

B. Damerau-Levenshtein distance

Damerau-Levenshtein distance is a String similarity metric for measuring how far two given strings are, in terms of four basic operations; deletion, insertion, substitution and transposition. This algorithm operates to find the minimum distance between two strings. It can be used for finding the closet match from sequential databases and input data sequence [9].

In this system, we first analyzed user's input dialogue what the meaning was and handled this user's input was concerned with which category databases by the above Regular Expression patterns. Then, we compared two strings; the first string was a string in the database determined by RE pattern and the second string was a

string that the user entered. For example, take the case of the strings A = "a cat" and B = "an act" and the distance of these two strings is "2"; to get from A to B requires one addition (the 'n') and two substitutions ('a' to 'c' and 'c' to 'a'). These two strings ကျိုက္ကမီ (Kyitekamee city in Mon State), ကျိုက္ခမီ (Kyitekamee city in Mon State) are the same meaning but there are two ways of spelling in Myanmar. The distance of these two strings is "1". [8]

However, we discovered that there was more accuracy result and reduced run time in measuring two strings with keywords i.e. Noun, Verb, Adjective, Adverb, etc of Myanmar sentence rather than measuring with the whole sentence. And thus, we did POS tagging on the input sentence for word-category disambiguation such as Noun, Verb, Adjective, Adverb etc. Myanmar word can usually be identified by the combination of root word, prefix and suffix. For example, ဘုရားသို့သွားမယ် ("go to pagoda" in English) can be identified as ဘုရား/n သို့/ppm သွား/v မယ်/ppm and ဘုရား (pagoda) is Noun (n), သွား (go) is Verb (v), သို့ (to) and မယ် is Post-positional Marker(ppm) [14]. We only compared keywords (Noun (n), Verb (v)) of two strings and we removed optional words such as Particle (part), Post-positional Marker (ppm). Therefore, we need to segment user's input sentence into appropriate word in order to make POS tagging because words are composed of single or multiple syllables and are not usually separated by white space in Myanmar texts. The steps of extracting keywords for String similarity measurement is shown in the following:

Unsegmented input sentence:

ကျိုက္ကမီသီးရိုးဘုရားရဲ့သမိုင်းကြောင်းကိုသိချင်ပါတယ်။
("I want to know about the history of Kyitehteyoe pagoda.")

Segmented sentence:

ကျိုက္ကမီသီးရိုး ဘုရား ရဲ့ သမိုင်းကြောင်း ကို သိ ချင် ပါ တယ် ။

POS tagged sentence:

ကျိုက္ကမီသီးရိုး/n ဘုရား/n ရဲ့/ppm သမိုင်းကြောင်း/n ကို/ppm သိ/v
ချင်/ppm ပါ/part တယ်/ppm ။/punc

Keywords for String similarity measurement:

ကျိုက္ကမီသီးရိုး/n ဘုရား/n သမိုင်းကြောင်း/n သိ/v

We measured multiple strings with Damerau-Levenshtein distance and controlled the similarity between the two strings with threshold value ("15") because we didn't measure with each keyword and we measured with the whole keyword "ကျိုက္ကမီသီးရိုး ဘုရား သမိုင်းကြောင်း သိ" from the two strings.

C. Word Segmentation

Myanmar texts are no blanks for word boundaries. Although Myanmar language are sometimes used for separating phrases for easier reading, it is not strictly necessary and these spaces are rarely used in short sentences. There are no clear rules for using spaces in Myanmar language. So, we need to make correct segmentation for user's input sentence. [10] We defined Myanmar "word" to be meaningful units and affix, root word and suffixes are separated such as "လည် ဖို့", "သွား မယ်". Here, "လည်" ("visit" in English), "သွား" ("go" in

English) is a root word and the others are suffixes word. Plural nouns of Myanmar language are identified by particle. We also put space between noun and particle, for example "ဘုရား များ" ("pagodas" in English) is segmented as root word "ဘုရား" and the particle "များ". In Myanmar grammar, the particles describe the type of noun, and used after number or text number. For example, "ဘုရားနှစ်ဆူ" ("the two pagodas" in English) is segmented as "ဘုရား နှစ်ဆူ" and "မော်လမြိုင်မြို့" ("Mawlamyine city" in English) is segmented as "မော်လမြိုင် မြို့".[6] We used Myanmar sentences of our developing travel-tour conversation data for building a word segmentation model. The followings are the example word segmented Myanmar sentences in our chatbot corpus:

- ကျိုက်ထီးရိုး ဘုရား အနီးတဝိုက် မှာ ဘာ တွေ ရှိ လဲ ။
- မော်လမြိုင် မြို့ မှာ ဘာ ဘုရား တွေ ရှိ လဲ ။
- မုဒုံ မြို့ မှာ ဘယ် စားသောက်ဆိုင် က နာမည်ကြီး လဲ ။
- ဘီးလင်း မြို့ မှာ ဘာ ဘုန်းကြီးကျောင်း တွေ ရှိ လဲ ။

We used Conditional Random Fields (CRFs) for building word segmentation model [7].

D. Web Scraping

Web scraping is a computer software technique of extracting information from websites. This technique mostly focuses on the transformation of unstructured data (HTML format) on the web into structured data (database or spreadsheet). There are several ways to extract information from the web. Use of API is probably the best way to extract data from a website. Almost all large websites like Facebook, Google, Twitter, StackOverflow provide APIs to access their data in a more structured manner. However, we used Python's free libraries (BeautifulSoup4, Urllib2) for scraping data from websites. We applied a web scraping process for replying Google maps, YouTube videos and website links that are corresponding to the user's question when the user's input question is not similar to prepared strings in the dataset.

IV. IMPLEMENTATION

A. Corpus Statistics

For the experiment, we collected 4,422 Question and Answer parallel corpus based on the travel and tourism domain including well-known sightseeing places of Mon State, Pyin Oo Lwin and Mandalay city. We searched for information on each place from Facebook, Google and prepared possible questions from chatbot users and suitable answers reply to users. It contains family destinations (pagoda, beach, waterfall, etc), event destinations, hotel destinations, historical and cultural destinations, mountain destinations, food destinations (traditional food, restaurant, etc), transportation destination and weather destination. In detail, prepared of 2,978 parallel sentences relating to Mon State, 1,117 parallel sentences relating to Mandalay city and 327 parallel sentences relating to Pyin Oo Lwin respectively. However, we only used 2,978 sentences relating to Mon State for the chatbot prototype development. We used 4,422 parallel sentences for word segmentation model and 4,000 parallel sentences for training data and 442 parallel sentences for test data of word segmentation model.

B. Conditional Random Fields

Conditional Random Fields (CRFs) are models that consider dependencies among the predicted segmentation labels. It can obtain probably distribution over sequences. It is suitable for segmenting and labelling structured data. We used the CRFsuite toolkit [12] to build the CRF models. The feature set used in the models (up to character/syllable tri-grams) was as follows (where t is the index of the character/syllable being labeled) [7]:

Syllable unigrams: $\{w_{t-2}, w_{t-1}, w_t, w_{t+1}, w_{t+2}\}$
 Syllable bigrams: $\{(w_{t-1}, w_t), (w_t, w_{t+1})\}$
 Syllable trigrams: $\{(w_{t-2}, w_{t-1}, w_t), (w_{t-1}, w_t, w_{t+1}), (w_t, w_{t+1}, w_{t+2})\}$

These n-grams were combined with label unigrams and bigrams to produce the feature set for the model. The word segmentation can be done from both character segmented data and syllable segmented data. We used syllable segmented data for this chatbot prototype. [11] We used predicted tag set label with each syllable segmented data. An example of tagging label of each syllable for word segmentation model is shown in the following:

Unsegmented input: မွန်ပြည်နယ်မှာဘာဘုရားတွေရှိလဲ။
 Syllable broken output: မွန် ပြည် နယ် မှာ ဘာ ဘု ရား တွေ ရှိ လဲ ။
 Label for each syllable: - - | | | - | | |

We used two tag labels “-” and “|”. The tag label “-” means other syllable word, “|” means word segmented symbol. Thus, the output is “မွန်ပြည်နယ် မှာ ဘာ ဘုရား တွေ ရှိ လဲ ။” because it will only segment the word that is tagged with label “|”.

C. System Implementation

This system was developed by Python (version 3.7.3) and the Flask framework for Web interface. Our chatbot system designed not only text message replies but also required web-based services such as Google Maps, YouTube videos, and website links. A sample of chatting between user and chatbot interaction is shown in the following:

User : ဟိုင်း

Chatbot : ဟယ်လို မင်္ဂလာ ညနေခင်းလေးပါ ခင်ဗျာ

User : ဒီအချိန် ခရီးထွက်ရင် ဘယ်ကို သွားရင် ကောင်းမလဲ ။

Chatbot : အခုချိန် က မိုးရာသီ ဆိုတော့ ခရီးထွက်တာ တွေ တော့ သိပ်မရှိဘူး။ ဒါပေမယ့် အညာဒေသ ဖြစ်တဲ့ ပုဂံ၊ မုံရွာ၊ မန္တလေး စတဲ့ နေရာ မျိုးကို သွားလည်ကြတာ များ ပါတယ် ခင်ဗျာ ☺

User : မွန်ပြည်နယ် သွားလည်ရင် ကောင်းမလား ။

Chatbot : မွန် ပြည်နယ် မှာ ယဉ်ကျေးမှု မြင့်မား တဲ့ မွန် လူမျိုး တို့ ရဲ့ ရာမညမြေ လေး တစ် ခု ဖြစ် ပါ တယ် ။ ကျိုက်ထီးရိုး ၊ ကျိုက္ခမီ ၊ ကျိုက်သလုံ စတဲ့ တန်ခိုးကြီး ဘုရား များ နှင့် ဘုရား စေတီ ပေါင်း များစွာ တည်ရှိ ရာ ပြည်နယ် လေး တစ် ခု ဖြစ်လို့ သွားလည်သင့်ပါတယ် ။

User : ကျိုက်ထီးရိုးဘုရားက ဘယ်မြို့ မှာ ရှိတာလဲ ။

Chatbot : ကျိုက်ထို မြို့ မှာ ရှိ ပါ တယ် ။

User : မွန်ပြည်နယ်မှာ ဘာ အထင်ကရဘုရား တွေ ရှိ လဲ ။

Chatbot : ကျိုက်ထီးရိုး ဆံတော်ရှင် စေတီတော် ၊ ကျိုက္ခမီ ရေလယ် ဘုရား ၊ ကျိုက်သလုံ စေတီတော် ၊ ရွှေစာရုံ စေတီတော် ၊ ကျိုက်ပေါလော မွဲရှင်တော် ၊ မော်လမြိုင် တောင်ဝိုင်း ဘုရား ၊ ကျိုက်ထီးဆောင်း စေတီတော် ၊ ကျိုက်မရော ဘုရား တို့ ရှိ ပါတယ် ။

User : ကျေးဇူးပါ

Chatbot: ရပါတယ် ဗျာ။ အခု လို ကူညီ ခွင့် ရ တာ ကို ပဲ

ကျေးဇူးတင် ပါ တယ်😊

In Fig. 2, we show the user interface of our chatbot prototype.



Fig. 2 User interface of chatbot prototype

V. EVALUATION

In our experiments, we examined the effectiveness of Regular Expression and String similarity approach for the Myanmar language chatbot system. We also examined with web scraping techniques as an intelligence of chatbot system at the state of out of database.

Evaluation is crucial in dialog system design. To get a more fine-grained idea of user happiness, we can compute a user satisfaction rating, having users interact with a dialog system to perform a task and then having them complete a questionnaire. We used a Likert scale evaluation. A Likert scale is a rating scale, often found on survey forms, that measures how people feel about products, services, or experiences. It includes a series of questions that we ask people to answer and ideally 5-7 balanced response people can choose from [15]. We conducted a Likert scale evaluation test of our chatbot, involving 16 users (13 female and 3 male), the age range is between 22 and 26, 8 undergraduate and 8 graduate students. We prepared 8 user satisfaction survey questions and defined each question with range 1 to 5; 1-highly dissatisfied, 2-dissatisfied, 3-neutral, 4-satisfied and 5-highly satisfied. We then created 10 conversation scenarios based on several categories such as pagoda, hotel, beach, waterfall, mountain, monastery, market, restaurant, transportation, and general conversation. Moreover, we also take free comments and feedback from user suggestion and requirement for our chatbot.

We evaluated our chatbot system with the two versions. The first version is measuring String similarity with the whole dataset without separating categories and

the second version is with separating the dataset into several categories such as beach categories of Mon State, pagoda categories of Mon state, etc. The evaluation of the chatbot's first version was tested with 8 people (6 female and 2 male) and the second version was tested with 8 people (7 female and 1 male).

VI. RESULT AND DISCUSSION

The Likert scale score testing with the first version is shown in Table 1. Here, “u” stands for the user.

TABLE I

USER SATISFACTION SURVEY BY LIKERT SCALE WITH THE FIRST VERSION

	u1	u2	u3	u4	u5	u6	u7	u8	Overall
Easy to use	4	3	3	4	3	3	4	5	3.6
Easy to understand	3	3	3	2	3	2	3	3	2.7
Fast to reply	3	3	3	2	3	3	2	3	2.7
Satisfied on response	2	3	3	3	2	2	3	3	2.6
Expected response	3	3	3	3	2	2	2	3	2.6
Give advice	3	3	2	3	3	3	3	3	2.8
Reasonable response	3	2	3	2	3	2	2	2	2.5
Future use	3	3	3	4	4	3	3	4	3.2

Some common free comments and feedback from users are as follows:

Feedback 1 – Need to train more data (eg: hotels, food, famous places, fruits, traditional present, souvenir shops, stations, weather, etc)

Feedback 2 – It should reply how to reach the destination when the user asks a question; eg: Mandalay to Mon (by car, by bus, by train or by plane).

Feedback 3 – I want to suggest that more transportation data should be added in this chatbot such as distance and time from one place to another.

Feedback 4 – To add all main cities of Myanmar to get more usable suggestions.

Feedback 5 – Myanmar word segmentation of your chatbot is sometimes wrong and it will add more training data for building a word-segmentation model.

Feedback 6 – It is required to handle getting a reasonable response. (e.g: When the user is asking about pagoda of Mon state, the chatbot response should not be relating to a beach).

For example:

User: မွန်ပြည်နယ်မှာ ဘုရား ဘယ်နှဆူ ရှိလဲ။

(“How many pagodas in Mon State?” in English)

Chatbot: မွန်ပြည်နယ်မှာ အရမ်းလှတဲ့ကမ်းခြေတွေ အများကြီး ရှိ ပါ တယ် ။

(“There are many beaches in Mon State.” in English)

According to Feedback number 6, we updated our chatbot system by separating the whole dataset based on the categories such as pagoda database of Mon State, beach

database of Mon State, etc. The response algorithm is also retrieving a suitable answer with the corresponding database rather than processing with the whole dataset. Table 2 shows the Likert scale score with the second version.

From the result of the Table 2, we got more accuracy than the 1st user study (refer Table I). We also found that the response time of the chatbot prototype is also significantly increased. It is clearly showed the effect of separating the chatbot dataset is based on the categories.

According to the Likert scale score and user's feedback, we need to add more detailed data relating to food, hotels, transportation, tradition, events, etc. We examined our chatbot system with data from Google, Facebook . Thus, we need to go each region and collect data or connect with travel blogger to get more detailed information about all regions in Myanmar. To keep the continuous conversation between the chatbot and the user, we need to consider more conversation patterns of Myanmar language and the process of analyzing collected user chatting history. Moreover, a good word segmentation model is also required.

TABLE II
USER SATISFACTION SURVEY BY LIKERT SCALE WITH THE SECOND VERSION

	u1	u2	u3	u4	u5	u6	u7	u8	Overall
Easy to use	4	4	5	5	5	4	4	3	4.3
Easy to understand	4	3	3	3	4	2	2	2	2.9
Fast to reply	5	2	5	5	5	3	4	4	4.1
Satisfied on response	4	4	3	3	4	2	2	3	3.1
Expected response	4	3	3	3	4	2	2	3	3
Give advice	4	4	3	3	4	2	5	4	3.2
Reasonable response	4	4	3	3	3	3	3	3	3.2
Future use	4	5	3	4	5	3	5	4	4.1

VII. CONCLUSION

This paper contributes to the process of creating a rule-based Myanmar Language chatbot system with a set of rules based on Regular Expression and String similarity. The Likert scale results with the developed chatbot prototype show that human conversation for the specific domain is possible to manage by rules. However, we have to handle wrong spelling mistake of the user input (e.g. စသောကဆိုင်, ကျိုက်ထီးရိုး, မုန့်ဟင်းခါး, ဘုန်းကြီးကျောင်း) and

paraphrases (e.g. ကျိုက်ထီးရိုး, ကျိုက်ခမ်း). Using the general RE pattern (.*)မုန့်ဟင်းခါး(.*)ဘုန်းကြီးကျောင်း(.*)စသောကဆိုင်(.*) is also difficult to get the best answer. In our current chatbot prototype, it is clear that the responses are often irrelevant and not reasonable. And thus, we need to increase the corpus size of our system. In the future, we plan to study the conversational dialogue act for Myanmar language, developing spelling checking and paraphrase handling facilities to the Myanmar Language chatbot. Moreover, we also plan to explore the combination of a chatbot with machine translation technologies for handling both English and Myanmar languages.

ACKNOWLEDGMENT

We would like to express grateful thank to the University of Computer Studies (Banmaw) students who are working together at Language Understanding Lab, Pyin Oo Lwin, Myanmar as an internship and students from NLP class in University of Technology Yatanarpon Cyber City (UTYCC) for their help in user studies of our chatbot.

REFERENCES

- [1] B. AbuShawar and E. Atwell, "ALICE chatbot: trials and outputs," *Computación y Sistemas*, vol. 19, no.4, pp. 625–632, 2015.
- [2] B. A. Shawar and E. Atwell, "A comparison between Alice and Elizabeth chatbot system", University of Leeds, School of Computing research report 2002.19
- [3] Jan Goyvaerts, "Regular Expression Complete Tutorial", 2007, pp. 137-138.
- [4] I. Nica, O. A. Tazl and F. Wotawa, "Chatbot-based Tourist Recommendations Using Model-based Reasoning", Aug. 2018.
- [5] A. M. M. Kyaw, "a chatbot for recommending tourist attractions in myanmar", M. Eng. thesis, Asian Institute of Technology, School of Engineering and Technology, Thailand, Dec. 2018.
- [6] Thazin Myint Oo, Ye Kyaw Thu, Khin Mar Soe "Statistical Machine Translation between Myanmar (Burmese) and Rakhine (Arakanese)", Natural Language Processing Lab., University of Computer Studies, Yangon, Myanmar
- [7] Win Pa Pa, Ye Kyaw Thu, Andrew Finch, Eiichiro Sumita and Yoshinori Sagisaka. Word Boundary Identification for Myanmar Text using Conditional Random Fields
- [8] Khaing Hsu Wai, Ye Kyaw Thu, Hnin Aye Thant, Swe Zin Moe, Thepchai Supnithi "String Simialrity measures for Burmese (Myanmar Language)"
- [9] Fred. J. Damerau, "A Technique for Computer Detection and Correction of Spelling Errors," *Communications of ACM*, Vol. 7, No. 3, 1964, pp. 171-176.
- [10] CHENCHEN DING, YE KYAW THU, MASAO UTIYAMA, and EIICHIRO SUMITA, National Institute of Information and Communications Technology, "Word Segmentation for Burmese (Myanmar)"
- [11] Dr. Ye Kyaw Thu. Sylbreak. Available: <https://github.com/ye-kyaw-thu/sylbreak>
- [12] Naoaki Okazaki. CRFSuite. "A fast implementation of Conditional Random Fields" [Online]. Available: <http://www.chokkan.org/software/crfsuite/>
- [13] Stephen Cole Kleene. Regular Expression. Available: https://en.wikipedia.org/wiki/Regular_expression
- [14] Ye Jan Goyvaerts, 2007, "Regular Expression", Available: <https://www.regular-expressions.info/print.html>
- [15] "An introduction of Likert scale". Available: <https://wpforms.com/beginners-guide-what-is-a-likert-scale-and-how-to-use-it/>