University Chatbot using Artificial Intelligence Markup Language

Naing Naing Khin
Natural Language Processing Lab
University of Computer Studies,Yangon
Yangon, Myanmar
naingkhin@ucsy.edu.mm

Khin Mar Soe
Natural Language Processing Lab
University of Computer Studies,Yangon
Yangon, Myanmar
khinmarsoe@ucsy.edu.mm

***Abstract*—Chatbots are conversational systems that can do chat interactions with human automatically. It is developed to be virtual assistant, making entertainment for people, helping for answering the questions, getting driving directions, serving as human partner in smart homes etc. Most of the chatbots utilize the algorithms of artificial intelligence (AI) in order to get the required responses. In this paper, we provide the design of a University Chatbot that provides an efficient and accurate answer for any user questions about university information. This is the first University Chatbot for inquiring about school information in Myanmar Language based on Artificial Intelligence Markup Language and uses Pandorabots as the interpreter.**

Keywords—AIML, Natural Language Processsing, Pandorabots, Pattern Matching, Response Generation

I. INTRODUCTION

Conversational agents become essential by interacting of machines with the desired users to provide natural language interfaces. So, the role of chatbots in the information technology and communication is widely in used. Many chatbots are created day by day through marketing, medical, education and banking. Chatbot is also a user assistant substance that is intended to produce a communication with human through their regular language. In educational system, it is essential for teaching, learning and searching the desired information for a specific area. The obvious factor that leads us one step closer to living in our fantastic world is that it knows our messages and can respond to us. The bot would match the input sentence from the user with that pattern existed in the knowledge base. This system is simple Myanmar chatbot using AIML but it can answer the necessary information for the users. There are many chat engines with different methods and can perform chatting. Some famous chatbots are SimSimi, Mitsuku, A.L.I.C.E, and now the machine learning chatbots like Siri, Alexa, Cortana and so on. Although modern chatbots apply the power of artificial intelligence to answer complicated questions, they still need some improvements for low resource languages.

II. LITERATURE REVIEW

B. A. Shawar and E. Atwell [2] described a system to access Arabic language information using chatbot without sophisticated natural language processing or logical inference. They showed this work properly with English and European languages and how the same chatbot will react in terms of Arabic FAQs. S. A. Prasetya, A. Erwin[3] aimed to found right AIML interpreter which can be used in Indonesian language E-Commerce website. The system is built by using Artificial Intelligence Markup Language (AIML) and Pandorabots as the interpreter for the customers tending to buy things online. S. Hussain, O. A. Sianaki, and N. Ababne[4] discussed chatbots classification, design techniques used in earlier chatbots and modern ones, and how the main categories of chatbots can handle conversation context. They presented with the emergence of new technologies intelligent systems have emerged using complex knowledge-based models. B. R. Ranoliya, N. Raghuwanshi and S. Singh[5] provided the university related FAQs for students by using the Artificial Intelligence and Latent Semantic Analysis. They implemented to solve the academic needs of the visitors for Manipal University. A. Mondal, M. Dey, D. Das, S. Nagpal, K. Garda[9] focused on the design of texual communication in educational domain. They developed and processed the accurate chatbot by using random forest algorithm. R. Sharma, M. Patel[10] presented the review of design techniques on chatbot in speech conversation systems. They discussed the performance and usability of many chatbots in our everyday lives.

III. HUMAN COMPUTER INTERACTION IN CHATBOTS

Human Computer Interaction is a communication field of study focusing on the design of computer technology and the interaction between human and computers. Conversation system between a human and a computer is either chatting by typing text or speech dialogue using the voice. Thus, interaction with natural language is a feasible option for connecting machine agents and human users. The chatbot popularity has brought the new feeds for HCI as it has changed the pattern of human interaction. Human Computer Interaction may need to consider for chatbots as the main object of design, focus on services than user interfaces, and design for interaction in networks of human and machine actors[13]. The main parts which include human computer interaction in conversation systems design are (a) the techniques used to produce keywords, (b) pattern matching techniques used inside the chatbot and (c) the type of response. So, HCI is considered taking on human-chatbot interaction design as an area of research and practice.

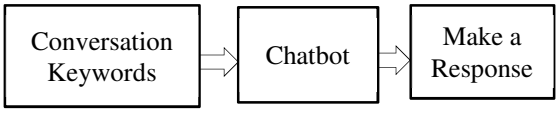

Fig. 1. Human Computer Interaction

IV. Artificial Intelligence Markup Language

AIML is an XML based markup language for specifying chatbot content. It was created by the ALICE bot free software community in 1995-2000 for the people to input dialogue pattern knowledge into chatbots based on the ALICE free software technology. An AIML Interpreter is able to load and run the bot, then, it provides the bot's responses in a chat session with a user. AIML consists of data objects called AIML objects, which are made up of units called topics and categories. The topic is called an optional top-level element, it has a name and a set of categories related to that topic. Categories are the basic units of knowledge in AIML. One category is a rule for matching an input and converting to an output, and consists of a pattern, which represents the user input, and a template, which responses the answer. The AIML pattern is simple and consists of words, spaces, and the wildcard symbols _ and *.

```
<?xml version="1.0" encoding="UTF-8"?>
<aiml version="2.0">
<!—AIML code goes here -->
</aiml>
```

A. *AIML Categories*

There are three AIML types: (a) atomic categories, (b) default categories, and (c) recursive categories[1].

(a) Atomic categories are those with patterns that do not have wildcard symbols, _ and *.

```
<category>
<pattern>မင်္ဂလာပါ</pattern>
<template>ဟုတ်ကဲ့ မင်္ဂလာပါ ရှင်</template>
</category>
```

(b) Default categories include wildcard symbols * or _.

```
<category>
<pattern>ကျောင်းချိန်် က ဘယ်အချိန်မှာ * </pattern>
<template>ကျောင်းချိန် က မနက် ၉ နာရီမှာ စပါတယ်
</template>
</category>
```

For such situation, if the user enters ကျောင်းချိန် က ဘယ်အချိန်မှာ စတာလဲ then the AIML class will search until the wild symbol (*) and if there is a match it will accompany response.

(c) Recursive categories are the categories with templates <srai> and <sr> tags, which represent recursive artificial intelligence and symbolic reduction. Applying a combination of wild cards and srai, the stop words of the sentences can be carefully checked out from the user input. Recursive categories involve many applications: **(i) symbolic reduction** which reduces the complex grammatical forms to simpler ones; **(ii) divide and conquer** category splits an input into two or more subparts and add the responses to one; **(iii) synonyms resolution** is possible to appear different words with the same meanings depending on the consisting text; **(iv) keyword detection** is possible to find the same response when a definite keyword is found in the user input [6]. These are some examples of different AIML categories.

(i) symbolic reduction

```
<category>
<pattern>တီချယ်သီ ကို သိလား</pattern>
<template>ကွန်ပျူတာသိပ္ပံမဟာဌာန မှ ဌာနမှူး
ဖြစ်ပါတယ် </template></category>
<category>
<pattern>တီချယ်မွန်း ကို သိလား</pattern>
<template>သူတသိပ္ပံမဟာဌာန မှ ဌာနမှူး ဖြစ်ပါတယ်
</template></category>
<category>
<pattern>သင် * ကို သိလား</pattern>
<template><srai><star/> ကို သိလား</srai>
</template></category>
```

(ii) divide and conquer

```
<category>
<pattern>ကျေးဇူးပါ ရှင့်</pattern>
<template>ဟုတ်ကဲ့ ပြန်လည် တွေ့ဆုံဖို့ မျှော်လင့်ပါတယ်
</template> </category>
<category>
<pattern> ကျေးဇူးပါ *</pattern>
<template> <srai>ကျေးဇူးပါ ရှင့် </srai>
</template></category>
```

(iii) synonyms resolution

```
<category>
<pattern>ကျောင်းလိပ်စာ သိချင်လို့ပါ</pattern>
<template>အမှတ်-၄ လမ်းမကြီး၊ ရွှေပြည်သာမြို့၊
ရန်ကုန်တိုင်း ဖြစ်ပါတယ်</template>
</category>
<category>
<pattern>ကျောင်းက ဘယ်မှာ ရှိတာလဲ</pattern>
<template><srai>ကျောင်းလိပ်စာ သိချင်လို့ပါ </srai>
</template> </category>
```

(iv) keyword detection

```
<category>
<pattern>သင်ယူခြင်း</pattern>
<template> သင်ယူခြင်းသည် ပြောင်းလဲခြင်း ဆီသို့
ဉီးတည်စေသော လုပ်ငန်းစဉ် ဖြစ်သည်</template>
</category>
<category>
<pattern> _ သင်ယူခြင်း</pattern>
<template><srai>သင်ယူခြင်း</srai></template>
</category>
<category>
<pattern>သင်ယူခြင်း *</pattern>
<template><srai>သင်ယူခြင်း</srai></template>
</category>
<category>
<pattern>_ သင်ယူခြင်း *</pattern>
<template><srai>သင်ယူခြင်း</srai></template>
</category>
```

V. PROPOSED SYSTEM

We have implemented a Myanmar interactive chatbot for university frequently asked questions. AIML is defined with general inquiries and messages which are replied by applying AIML formats. According to the Artificial Intelligence Markup Language, we have used different AIML tags to get the user required information from the bot.

TABLE I. AIML TAGS USED IN SYSTEM

No.	Tags used for AIML Categories
1.	<topic> </topic>
2.	<category></category>
3.	<pattern></pattern>
4.	<template> </template>
5.	<srai></srai>
6.	<random> </random> with
7.	<set> </set>
8.	<get> </get>
9.	<that> </that>
10.	<think> </think>
11.	<condition> </condition>

The system operation is divided in three steps. In the first step, the question is entered by the user. In the second step, the system performs word processing actions to match the user's input to a pre-defined format and do the pattern matching between user input and the Knowledge Base. Finally, the answer is presented to the user in the third step.

A. Knowledge Base

Artificial Intelligence Markup Language is a well-known XML derived language to build chatbot knowledge base. Users' frequent asked question sets are defined semantically the knowledge domain given to the chatbot. The questions are available from the university academic center related to nine topics and manually collect from the teachers, students and their parents that they want to ask about the university. We have used 970 question-answer pairs as data distribution. A well-designed knowledge base can positively impact the effectiveness of chatbots that will improve the interaction between users.

B. Workflow of the System

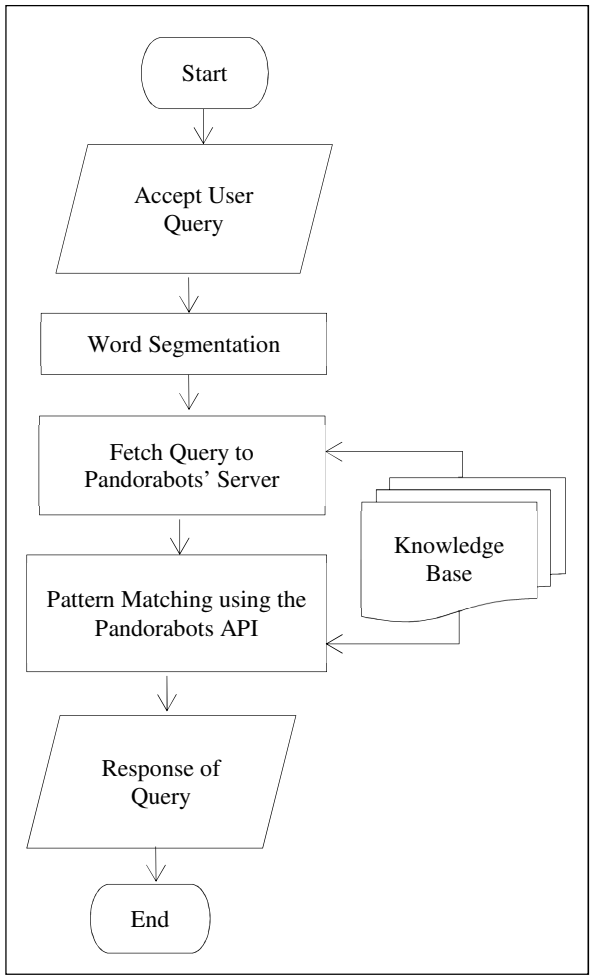

Fig. 2. Flow of the System

VI. IMPLEMENTATION OF THE PROPOSED CHATBOT

Rule-based chatbot contains a faster time-to-relevance, delivering a faster impact on user interaction. This chatbot is one of rule-based chatbot and developed on AIML language for the University of Computer Studies, Yangon. We have purposed this system to have a support for university routine. All the questions files need to be uploaded to Pandorabots server. The files include the university related questions and information that the students, teachers and parents frequently asked. The number of question-answer pairs in the system that

are utilized for different topics and type of categories are as given:

TABLE II. THE TOPICS AND NUMBER OF QA PAIRS USED IN THE SYSTEM

Topics	Atomic	Default	Recursive
Greeting	25	12	23
Location & Address	13	15	7
Academic	172	200	61
Brief History	18	20	15
Conference	39	35	11
Faculties & Staff	50	25	22
Library	40	20	10
Research & Lab	34	38	20
Alumni	12	25	8

The service requires internet connection to access the system. The users can interact with the chatbot in every time if there is a connection. The user needs to input the questions as Myanmar Language. The segmentation is done by using the UCSY word segmentor (http://www.nlpresearch-ucsy.edu.mm/NLP_UCSY/wordsegmentation.html). The input from the user is normalized and processed on the Pandorabots server. The AIML files, which are separated into several categories and the chatbot's knowledge is uploaded into the server. After pattern matching, the inquired user can ask the university related questions about academic services and activities. These are some sample results of our system:

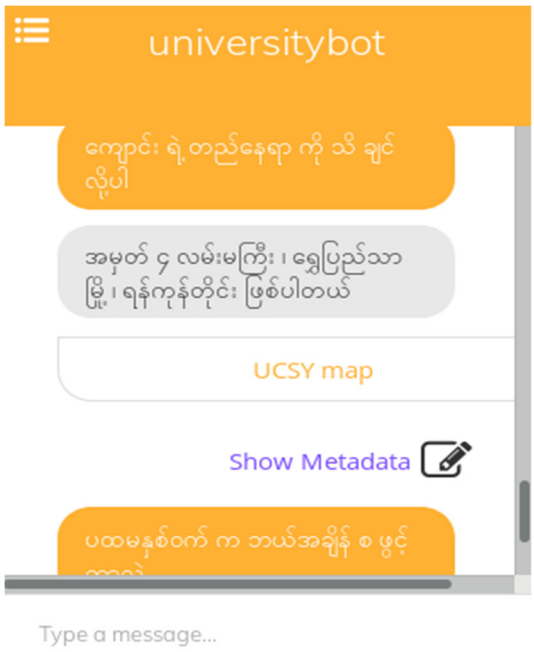

Fig. 3. Sample Outputs of the System

VII. TESTING AND EVALUATION

Testing can be made to measure the quality of chatbot. The steps included to conduct the chatbot experiments are (i) getting the overviews of questions that can be asked, (ii) inquiring the user questions related to the nine topics show in Table 2, (iii) asking the questions to get the feedback about the system, (iv) the experimental analysis whether they have correct or wrong responses [12].

We have made user testing with the closed user group by students, parents and staff. The system still needs improvement due to some patterns that mismatch with the chatbot knowledge. We also have adopted the dialog efficiency matrix to evaluate the chatbot.

A. *Dialogue Efficiency Metric*

The proposed system is measured the efficiency with four sample dialogues in terms of atomic categories, default categories, recursive categories. Adopted learning mechanism is used to see the ability to find answers for user as shown in table III.

TABLE III. TESTING IN FOUR DIALOGUES TYPES

Matching Type	D1	D2	D3	D4
Atomic Categories	8	13	20	10
Default Categories	25	9	7	19
Recursive Categories	14	10	15	20
Total	47	32	42	49
Mismatch Pairs	3	2	3	6

In the above table, we have tested the system with four dialogues as D1, D2, D3 and D4 for three matching categories. The bot can answer most of user questions correctly. We have also found some miss-match questions and answers as in figure 4. The error may occur by the user input or may be the chatbot knowledge. When we have multiple categories with the same pattern, AIML interpreter

chooses categories from bottom to top within each individual files. Unexpected bot responses can cause due to normalization error. This can be solved by the substitution file in the system. We also need to prepare more data for chatbot's knowledge.

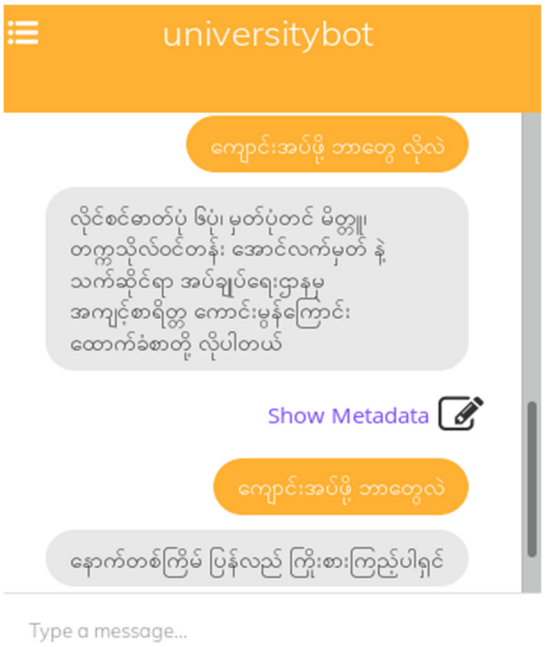

Fig. 4. Mismatch Question Example

The frequency of matching type in each dialogue generated between user and chatbot was calculated in figure 5. These absolute frequencies are normalized to relative probabilities.

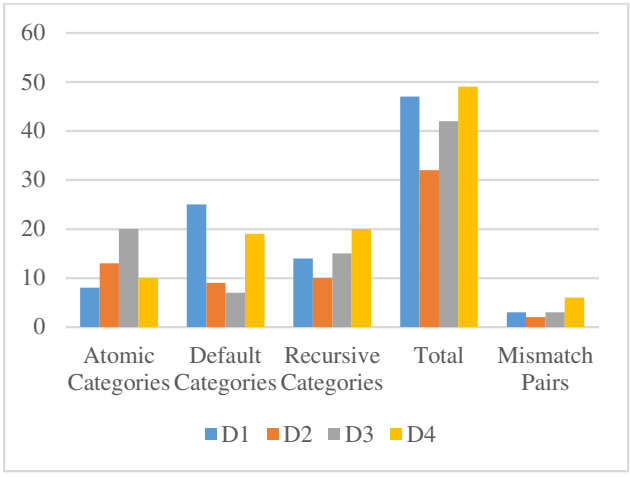

Fig. 5. Dialogue Efficiency of the Tested Dialogues

VIII. Conclusion

Chatbots can interact with people in effective ways. There are many chatbots in English and other languages by using different algorithms and models but there is little chatbot using Myanmar language. This is one of the University Chatbots using Myanmar Language to fulfill the information gaps between the university and its related users. Now, we have implemented a chatbot for the University of Computer Studies, Yangon. This is simple chatbot using Artificial Intelligence Markup Language and implemented on the Pandorabots server. The user can ask the useful questions about the university related the academic sectors through chatbot. The bot will help people to save time and get the information every time. We still need some improvements for the bot and we will develop this by using machine learning techniques in coming jobs.

References

[1] *AIML Foundation, June 2018.*

[2] *B. A. Shawar, E. Atwell, "Arabic question-answering via instance based learning from an FAQ corpus", 2009.*

[3] *S. A. Prasetya, A. Erwin, M. Galinium, " IMPLEMENTING INDONESIAN LANGUAGE CHATBOT FOR ECOMMERCE SITE USING ARTIFICIAL INTELLIGENCE MARKUP LANGUAGE (AIML)", 2018.*

[4] *S. Hussain, O. A. Sianaki, N. Ababneh, "A Survey on Conversational Agents/Chatbots Classification and Design Techniques", 2019.*

[5] *B. R. Ranoliya, N.Raghuwanshi, S. Singh, "Chatbot for university related FAQs," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 1525-1530.*

[6] *AIML - Sets and Maps in AIML 2.0, https://docs.google.com/document/d/1DWHiOOcda58CflDZ0Wsm1CgP3Es6dpicb4MBbbpwzEk/pub*

[7] *B. A. Shawar, E. Atwell ,"Chatbots: Are they Really Useful?", 2007.*

[8] *A. Hajare, P. Bhosale, R. Nanaware, G. Hiremath, "Chatbot for Education System", Vol.3, Issue 2, April 2018.*

[9] *A. Mondal, M. Dey, D. Das, S. Nagpal, K. Garda, "Chatbot: An automated conversation system for the educational domain", IEEE, 2018.*

[10] *R. Sharma, M. Patel, "Review on Chatbot Design Techniques in Speech Conversation Systems", Vol.5, Issue 9, September 2018.*

[11] *A. S. Lokman, J. M. Zain, "Chatbot Enchanced Algorithms: A Case Study on Implementation in Bahasa Malaysia Human Language", 2010, pp.31-44.*

[12] *B. A. Shawar, "Different measurements metrics to evaluate a chatbot system", 19 May 2014.*

[13] *A. Følstad, P. B. Brandtzaeg, "Chatbots and the new world of HCI", 15 April 2018.*

[14] *Giovanni De Gasperis, Isabella Chiari, "AIML Knowledge Base Construction from Text Corpora", 5 June 2014.*