IJCSE International Journal of Computer Sciences and Engineering Open Access

A Proposal of Chat bot for Malayalam

Sandhini S[1], Binu R[2], Dr. Rajeev R R[3], Reshma M M[4]

[1]Dept. Of Computer Science, Government Engineering College, Palakkad
[2]Dept. Of Computer Science, Government Engineering College, Palakkad
[3]Computational Linguistics. International Centre for open Source and Software
[4]Fellow, Computational Linguistics. International Centre for open Source and Software
e-mail: sandinisukumar@gmail.com , binurajappan@yahoo.com , rajeevrr@icfoss.in , reshmamm@icfoss.in

***Abstract*—** A chatbot is a conversational agent which interacts with humans via natural languages. Text, as well as speech, is used as the input to these systems. We propose a first Malayalam chatbot based on a language-independent natural language processing library with a learning mechanism. The chatbot is completely retrieval based chatbot which can converses in the Malayalam language. Malayalam is a Dravidian language talked over the Indian state of Kerala. The machine learning, as well as NLP approaches, is used to analyze user's queries and generate appropriate responses. We experiment an AIML based chatbot and a machine learning based chatbot. From the two chatbots, the machine learning-based chatbots have the best performance. Developed a domain-specific chatbot. It is a commercial product, so we can apply this to any domain.

***Keywords*—** Chatbot, Natural Language Processing, Machine Learning, Artificial Intelligence

I. INTRODUCTION

Chatbot is computer programs that communicate human through natural languages[1]. It is also known as a talkbot, chatterbot, Bot, IM bot, interactive agent, or Artificial Conversational Entity. The user can interact in a conversation via either written, oral, or mixed. The chatbot can be used in various fields such as Business, Health, Tourism, Customer Support and so on. The architecture of a chatbot integrates with the language model as well as computational algorithms.

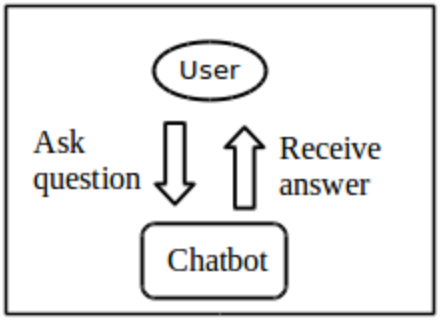

Figure 1.1 Chatbot Components

There are many English chatbots are available. Because it is easy to build chatbots in English. A lot of platforms are available for building chatbots. But a lot of people are interested in developing regional language chatbots these days. Figure 1.1 shows the simple design of the chatbot. The chatbot conversation framework in two types: retrieval based and generative based chatbot.

1. Retrieval based chatbot: The model uses a knowledge base of predefined responses and working is based on pattern matching algorithm and select an appropriate answer. The system doesn't generate new sentences they reply based on the domain of knowledge.

2. Generative model: It does not contain a knowledge base, so they generate a new response for each question.

This paper proposes a chatbot that returns a relevant answer to questions asked by the wise. This paper is organized as follows: Section I contains the introduction of the chatbot technology. Section II contains the related work of chatbot. In this, it describes the previous methods used to implement the chatbot. Section III contains the design and methods used In this. Section IV contains the comparison of previous methods used in the development of chatbot. Section V concludes research work with future directions.

II. RELATED WORK

The idea of chatbot was originated by Weizenbaum [3] implemented the first chatbot ELIZA to mimic a psychotherapist. The ELIZA chatbot is based on the keyword matching mechanism. ALICE is another chatbot developed by Abu Shawar et.al [1] that implements in various dialogues using AIML (Artificial Intelligent Markup Language) is a extension of XML to represent patterns and templates fundamental these dialogues. The categories in the AIML are the basic element. Most commonly used a mechanism to develop chatbot is an AIML based method. The pattern matching concept is used in this method. AIML contains set of predefined rules.

The AIML based chatbot are most popular because they are very easy to implement. The ALICE chatbot won Leobner Price three times. Here, each categories consits of a pattern which is the input given by the user and the template which includes the output by the chatbot to the user. Here, there are three categories such as atomic, default, and recursive categories.

Ranoliya, et.al [2] implements a FAQ chatbot based on AIML and Latent Semantic Analysis (LSA) methods. They developed an interactive University related Frequently Asked Questions (FAQs) chatbot. Initially, user post the query on the chatbot. Then, process the query that entered by the user to match the predefined format. The pattern matching is done in the query and the pattern in the knowledge base. Finally, the patterns matched answer is presented to the user. The chatbot interacts queries related to college information, admission etc. The LSA method is used to discover the similarities between words as vector representation. So that the unanswered questions by AIML will be seen as an answer by LSA.

Kalaiyarasi, T. et.al [7] proposed a chatbot a Tamil chatbot POONGKUZHALI. In poongkuzhali the user can pick any current topic for discussion and ask in Tamil. Based on the context of the query the system generates an appropriate answer to the user. The system identifies the minimal context of the input and this can be done by using a set of decomposition rules. The response is generated by utilizing a set of reassembly rules that resided in the knowledge base.

S. Chaitrali et.al [5] proposed a Bank chatbot to handle queries related to the Bank. The user is interacting with the system by a web application. The user entered a query in the front end and then submitted the query, then the query is handled by the boot controller logic. Here, the query is preprocessed by using NLTK library. In this, the query is tokenized, then removed the unnecessary spaces, stop words and then extract the lemmas for each token. Then the query is converted to the vectorized format and classification algorithms are used to find the class it belongs to.Based on cosine similarity, the most similar answer is returned to the user as a response.

Shah et.al [6] proposed an intelligent chatbot based on Natural Language Processing (NLP) in educational systems. The database contains the topics related the educational system. The user enters the query. Then, in NLP phase the following process is applied to the query such as Tokenization, Lemmatization, POS tagging, Dependency Parsing and finally Role Labeling. The NLP and Machine Learning(ML) are applied at their respective levels.The LSTM method is used to develop the answer from the database. It allows the model to learn the encoded part and how to create the output relevant to the question asked by the use.

III. MALAYALAM CHATBOT

Chatbots are a computer program that simulates the human conversation. Developing a chatbot that converse in Malayalam. The input is a Malayalam text and the output is the response to the user input also in Malayalam. Introducing chatbot in a free order language like Malayalam is a challenging task. The attempt will be a defining moment in the Malayalam language. Regardless of whether the work begins as a domain specific, it can be brought into different domains and areas.

AIML based Malayalam Chatbot

AIML is the common technique used in the design of the chatbot. AIML is the derivation of the XML . The AIML data objects consists of two units: Topics and Categories. The purpose of the AIML chatbot is to facilitate a better conversation modeling. The data object in the AIML language has the responsibility of modeling the conversational patterns.

The important objects in the chatbots are categories, pattern, and template. The category tag is used to specify the knowledge in the conversation. The pattern contains the input from the user to the system and the template includes the output or response to the user. The structure of the AIML categories, patterns, template is given below,

```
<category>
        <pattern> Input from the User</pattern>
            <template>
                Response to the User
            </template>
</category>
```

The AIML chatbot is based on the pattern matching concept, the response is generated is based on the mapping of keywords in each request and their patterns. With the help of the AIML Interpreter the pattern matching between query and response is done. The natural language input is initially preprocessed. The keywords are extracted by removing stopwords, stemming of the words. The pattern matching is

achieved with the help of these tags such as <that>, <srai>, <topic> for remembering the previous conversation.

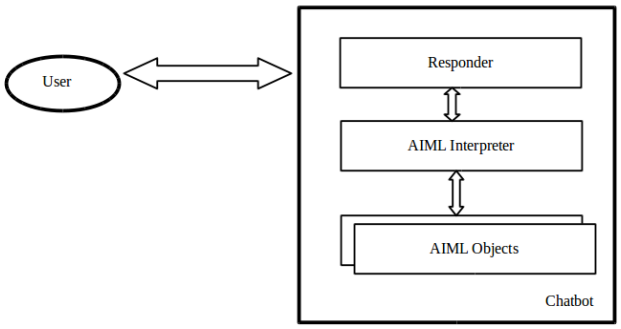

Figure 3.1 AIML based chatbot Architecture

The user inputs the question to the system and the AIML Interpreter matches the keywords and generate the response then, manage the conversation. The <random> tag is used to generate random response. The AIML language is language independent. So Malayalam chatbot can be developed based on the AIML tags.

```
<?xml version = "1.0" encoding = "UTF-8"?>

<aiml version="1.0.1" encoding="UTF-8">

<!-- basic_chat.aiml -->

<category>

    <pattern>ഞാൻ *</pattern>

    <template>

      നമസ്കാരം <set name = "username"> <star/>! </set>

    </template>

  </category>

</aiml>
```

This simple AIML file is shown above. The category part includes pattern and template tags which contains the input as well as the output. The AIML chatbot can be store the name of the user using the tag <set name>. They can remember the name of the user till the conversation ends.

Malayalam Chatbot based on Chatterbot Python Library

Propose a retrieval based domain specific chatbot that converse in Malayalam language. The chatbot is developed based on natural language processing library with machine learning mechanism. Initially, the chatbot matches the input with the queries in the database, and then calculates a confidence value for sentences that matched. Finally, selects the appropriate response based on the highest confidence value among them. Here, the response generation is retrieval based. The retrieval based method retrieve responses from the knowledge base based on the context. The input is given to a web interface. Then process the text that given by the user. Selects the sentences which matches the input text. Returns known response to the selected matches. Then, calculates the confidence value of each responses. The system will return appropriate response as the highest confidence value. Finally, the response is presented to the user.

Chatterbot is a Python package based on Machine learning concepts. It is a language independent platform. The important module of the Chatterbot library are Chatterbot's adapters. The input is returned from the input adapter, the input is processed and stored by the logic and the storage adapters. Finally, it passed to output adapter to returns the responses to the user. Chatterbot includes training tools to make simple training process. The Chatterbot training includes loading of the dialogues in to the chatterbot database. Several training classes are in built in chatterbot package. Here, we train the data from the dialogue corpus using Corpus Trainer Class of Chatterbot. Initially, developed a Malayalam Tourism corpus in the predefined format in YAML file. Then, set a trainer for the Malayalam dialogue corpus. The chatterbot has built in storage adapter that connects different databases. The Best Match logic adapters is used here to find the best response to the closet match. Here, uses an Low confidence adapter and set a threshold value 0.5. The Best Macth adapter calculates similarity function used to compare the input text to known responses. The response is selected based on the threshold value.

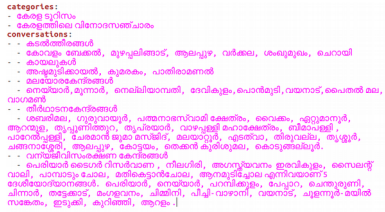

Figure 3.2 Format of creating the corpus

Manually created Malayalam tourism corpus in a specific format shown in the figure. The datas are stored as in the form of YML file. YML(*YAML Ain't Markup Language*) is a data serialization language in human readable format. The chatbot interface is created by using Flask Liabrary. The Yml file contains a categories and conversations part, the category part describes the name or category of the data. The conversations part includes the different conversations made on the topic for training the chatbot.

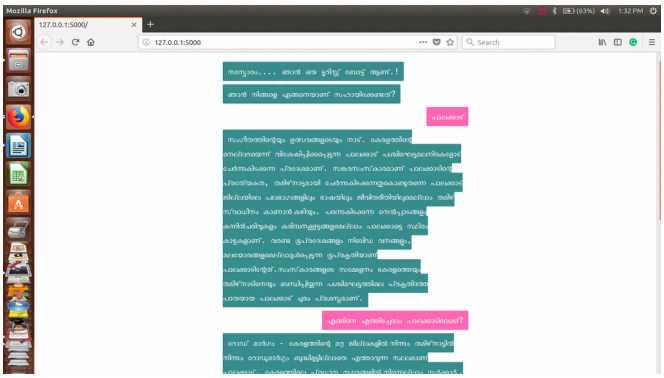

Figure 3.3 Implementation of Malayalam chatbot

Figure shows the implement of Malayalam chatbot. The principle focal point of this chatbot is to produce sentences free from linguistic errors, spelling errors and steady. It accomplishes the objective of delivering linguistically right Malayalam responses. Since the responses are on a par with its knowledge base so a great deal of work needs to be done to upgrade the knowledge base. However, amid building the knowledge base, must give an knowledge base free from mistakes.

IV. COMPARISON

Since, this is the first cahtbot in Malayalam. So, there is no another Malayalam chatbot for comparison. So compare the AIML and Chatterbot based Malayalam chatbot. The AIML based chatbot is a rule based one. The Chatterbot is integrates the Machine learning and Natural language processing concepts.

AIML based chatbot used simple patten template to represent the user input and output, and also using simple pattern matchin algorithm. The chatbot based on chatterbot starts off with no knowledge of how to communicate. Since we don't locate an appropriate database for this purpose we manually prepared the corpus in Malayalam. We take care of the issue of absence of required tools by choosing a language independent platform and picking a retrieval based model to fill the need.

We examine same questions to both chatbots.The machine learning based chatbot can answer progressively like others. Since it can take include in Malayalam and can give a response in Malayalam so we can state that the pattern matching algorithm is working great. Our chatbot answers in linguistically remedy Malayalam and it is free from spelling errors and any kind of linguistic errors. It makes some accentuation issue which can be enhanced in future.

V. CONCLUSION and FUTURE SCOPE

Conversational AI is a big part of the future. The users can easily enter their question in natural language and retrieve data. In this paper, propose chatbot for Malayalam language. A chatbot is an important tool for communicating with the user. General purpose chatbot must be straightforward, easy to use, must be effectively comprehended and the database must be smaller. Although some of the commercial items have developed, enhancements must be made to find a typical approach for designing a Chatbot. The Malayalam language has its own characteristic features, that make it different from other languages. This experiment is a spearheading work in the field of conversation framework in Malayalam. The principle test of this work is to make a chatbot in light of the precise learning base. Because of need of a substantial dataset, we actualized a retrieval based closed domain chatbot which will speak with the user based on the pattern matching algorithm and will enhance its performance measure by gaining from the cooperation. Our work will give a Malayalam conversation corpus which will help in the improvement of tools for Malayalam Language Processing research.

For future work, we can implement a voice enabled chatbot system and add pictorical representation for better understanding for people.

REFERENCES

[1] Shawar, Bayan Abu, and Eric Atwell.”Chatbots: are they really useful?." *Ldv forum*. Vol. 22. No. 1. 2007.

[2] Ranoliya, Bhavika R., Nidhi Raghuwanshi, and Sanjay Singh. "Chatbot for University Related FAQs.", 2017.

[3] Weizenbaum, Joseph. "ELIZA—a computer program for the study of natural language communication between man and machine." *Communications of the ACM* 9.1, 36-45, 1966, .

[4] Marietto, Maria das Gracas Bruno, et al.”Artificial intelligence markup language: A brief tutorial.”*arXiv preprint arXiv:1307.3091*, 2013.

[5] S. Chaitrali, Kulkarni, U. Amruta, Bhavsar, Savita Chaitrali S Pingale. “BANK CHAT BOT – An Intelligent Assistant System Using NLP and Machine Learning”, International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 05, 2017.

[6] Shah, Rishabh, Siddhant Lahoti, and K. Lavanya. "An intelligent chat-bot using natural language processing." *International Journal of Engineering Research* 6.5 : 281-286. 2017.

[7] Kalaiyarasi, T., Ranjani Parthasarathi, and T. V. Geetha. "Poongkuzhali-an intelligent tamil chatterbot." *SIXTH TAMIL INTERNET 2003 CONFERENCE*. Vol. 1. sn, 2003.

[8] Abdul-Kader, Sameera A., and John Woods. "Survey on chatbot design techniques in speech conversation systems." *International Journal of Advanced Computer Science and Applications* 6.7 : 72-80.2015.

[9] E. Loper, and S. Bird, "NLTK: The natural language toolkit." pp. 63-70, 2002.

[10] S. Bird, "NLTK: the natural language toolkit." pp. 69-72, 2006.

[11] A. S. Lokman, and J. M. Zain, "An architectural design of Virtual Dietitian (ViDi) for diabetic patients." pp. 408-411, 2009.

[12] A. M. Galvao, F. A. Barros, A. M. Neves, and G. L. Ramalho, "Persona aiml: An architecture developing chatterbots with personality." pp. 266-1267, 2004.

[13] Mujeeb, Sana, Muhammad Hafeez Javed, and Tayyaba Arshad. "Aquabot: A Diagnostic Chatbot for Achluophobia and Autism." *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS* 8.9 : 209-216, 2017.

[14] Bani, Balbir Singh, and Ajay Pratap Singh. "College Enquiry Chatbot Using ALICE."

[15] S. J. du Preez, M. Lall and S. Sinha, "An intelligent web-based voice chat bot," EUROCON 2009, EUROCON '09. IEEE, St. Petersburg, 2009.

[16] Wailthare, Sumit, et al. "Artificial Intelligence Based Chat-Bot." *Artificial Intelligence* 5.03 (2018).

[17] TIWARI, AMEY, RAHUL TALEKAR, and SM PATIL. "College Information Chat Bot System."

[18] Hatwar, Nikita, Ashwini Patil, and Diksha Gondane. "Ai based chatbot." *International Journal of Emerging Trends in Engineering and Basic Sciences* 3.2 85-87. 2016.

[19] Sarthak V. Doshi “Artificial Intelligence Chabot in Android System using Open Source” Program International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 6, Issue 4, April 2017.

[20] Bayu Setiaji “ Chatbot Using A Knowledge in Database” International Conference on Intelligent System, Modling and Simulation 2016.