International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 11 | Nov 2019 www.irjet.net p-ISSN: 2395-0072

Review of Chatbot System in Marathi Language

Darshan Navalakha[1], Manjiri Pittule[2], Ravina Mane[3], Amit Rathod[4], Prof. N.G. Kharate[5]

[1,2,3,4] Students, Comp Dept., VIIT College, Pune, India
[5]Guide, Comp Dept., VIIT College, Pune, India

Abstract - *In the modern Era of technology, Chatbots is the next big thing in the era of interactive services. Chatbots is a virtual person who can effectively talk to any human being using collaborative textual skills. A virtual person is based on machine learning and Artificial Intelligence (AI) concepts and due to dynamic nature, there is a drawback in the design and development of these chatbots as they have built-in AI, NLP, programming and conversion services. Chatbot acts like routing agent that can categorize user context in conversation. Chatbot helped with natural language processing (NLP) to analyze the request and extract some keyword information. In this system chatbot is designed according to query-based system in Marathi language (Indian Language). User asks the question to system; the system automatically replay all answer through chatbot. OCR concept is used to fetch data from pdf or images in the text format. All the query related information is added in this system. Although the significant work is not carried out for Marathi language chatbot but many researchers and organization have started building chatbot system in this field.*

Key Words: **NLP, OCR, Data-set, Pattern Matching, Machine Translation.**

1. INTRODUCTION

A chatbot is a program intended to counterfeit a smart communication on a text or spoken ground. But this paper is based on the text only chatbot. Chatbot categorize the user input as well as by using pattern matching, access information to provide a predefined

Acknowledgement, based on the sentence given by the user. When the input is bringing into being in the database, a response from a predefined pattern is given to the user. A Chatbot is executed using pattern comparing, in which the order of the sentence is recognized and a saved response pattern is adapt to the select variables of the sentence. Chatbots are mainly developed is conversational dialogue engine which is built in Python, which makes it possible to reply based on the collections of all the known conversations. This chatbot system uses Marathi language. All the query related data is available in this system's database.

Text Recognition usually abbreviated to OCR, involves a computer system intended to interpret images of typewritten text (usually captured by a scanner) into machine editable text or to interpret pictures of characters into a standard encoding scheme representing them. OCR commenced as a field of research in artificial intelligence and computational vision. Text Recognition used in official task in which the large data have to type like post offices, banks, colleges etc., in real life applications where we want to collect some information from text written image. People wish to scan in a document and have the text of that document available in a .txt or .docx format.

2. LITERATURE SURVEY

The making and execution of chatbots is still a developing area, heavily related to artificial intelligence and machine learning, so the provided solutions, while possessing obvious advantages, have some important limitations in terms of functionalities and use cases. However, this is changing over time.

Mr. A.M.Rahman has identified some programming challenges of chatbot, For bot to work resourcefully it needs to provide vast logical resources which are I/P, O/P and entity phrases. It should be given care on singular, plural forms, Synonyms, Antonyms and the most important is the sentimental analysis [1].

Response generation using Intent Classification and Entity Recognition. Response selector selects response which should work better for the user. Bot is based on the integration, in which information "bot sends" commands into web service and gets it results [1].

Belfin R.V, Shobhana A.J, Megha Manilal, Ashly ann Mathew, Blessy Babu they have proposed A Graph Based Chatbot for cancer Patients. It is presented in text and Audio. The Continuous Communication with Bot bring positive attitude in patients. In their research work they have implemented, Knowledge Based Chatbot which fetch appropriate data from Cancer database and scrap data from different Cancer forums using Beautiful Soap. Some data pre-processing techniques like tokenization, punctuation removal, stop word removal, stemming is performed.[2]

In Ticketing Chatbot Service using Serverless NLP Technology, they have defined Specific domain chatbot which can redefine chat experienced with the automated response and also some CUI response that guided the user. Existence of more database /data result in more intellectual chatbot. Classification of intent in term of the sentence needs to be more fluent to consider assurance rate. In the future, chat history can be considered as chat experience so behaviour of the user can be analysed [3].

College Enquiry Chatbot Using Knowledge in

Database proposed that User will interact with web interface of the bot. The web application will be connected to the bot with the help of the bot connector which will create the object through which it will communicate with the bot. the query entered by the user will be sent to LUIS.ai where it will be handled and Intent and Entity of the user query will be retrieved and the equivalent response will be fetched from database[4].

In this system there are two modules:

A. Online Enquiry:

Students can enquire about faculties and query related to exams

B. Online Chat Bot:

1) The result can be show in images, cards format.

2) The query will be answered on the basis of question asks and language model built in LUIS and responses store [4].

In AMBER Chatbot and Detection of Paraphrases for Devnagari implemented by Shah Manthan, Jigneshkumar, Arnav Mediratta, Akshay Kundale,

Shubham Nangare. Has proposed Natural Language Processing tasks such as Transliteration, Tokenization, Part of Speech. AMBER uses Sentence Generation. It uses Paraphrasing detection Library. It also uses several frameworks for determining semantic similarity DKPro Similarity. Word Alignment Algorithm is used to calculate alignment amongst idioms and phrasal chunks. RDRPOST Tagger is used for obtaining POS Tags. It uses phonetic and stripping algorithm to find out suffix.[5]

Partrick Jansson, Shuhua Liu proposed detection of Novel and Emerging Named Entities from Social Media by using LSTM Model. Its framework contains of the components like Word Embedding, POS Tagging.

Word embeddings are present standard for many text applications as they are found to be able to appropriately capture not only syntactic regularities but also for semantic regularities of word. The GATE Twitter POS tagger is used to assign POS tags for each word [6].

In Question-Answering System surveyed by Ali Mohamed Nabil Allam and Mohamed Hassan Haggag, worked in the field of information retrieval (IR). The QA Systems are concerned with providing appropriate answer in reply to question proposed in natural language. The system is composed of three mechanisms as question classification, information retrieval and answer extraction [7].

In survey of machine translation for Indian languages to English and its approaches by Namrata G. Kharate and Dr. Varsha H. Patil surveyed on machine Translation based on several approaches such as Dictionary based, Rule based, Corpus based, Knowledge based and Hybrid based. Which showed us the work required and way to frame answer for asked question relatively with the work of shallow parser [8].

Researched done by M. Dahiya who has implemented a text based chatbot, which recognize the user input with pattern matching and access the given information for creating the answers. The paper covers the fundamental techniques to design the chatbot. Such as creating the dialog, creating database, creating a chat and conversation [9].

Table 1: Previous Survey

Sr no	*YEAR*	*TITLE*	*Language*	*METHODOLOGY*	*OBSERVATION*
1.	2017	Programming Challenges of Chatbot.	English	Response generation using Intent Classification and Entity Recognition. Response selector selects response which should work better for the user. Bot is based on the integration, in which information "bot sends" commands into web service and gets it results.	For bot to work efficiently it needs to provide vast logical resources which are I/P, O/P and entity phrases. It should be given attention on singular, plural forms, Synonyms, Antonyms and the most important is the sentimental analysis.
2.	2019	A Graph Based Chatbot for cancer Patients	English	1.Knowlege Based Chatbot a. Fetch Appropriate Data from Cancer database b. Scrap Data from different Cancer forums using Beautiful Soap. 2.Data Pre-processing a. Tokenization b. Punctuation Removal c. Stop word Removal d. Stemming 3.Data Modelling a. Data converted into graph model using Neo4j. 4.Detecting type of cancer and show	1.It is available in text and Audio 2. The Constant Communication with Bot bring positive attitude in patients. 3. Helpful for job Schedulers because they did not need to wait in Queue for doctors counselling, they directly use chatbot for their help. 4. Data collected from various forums makes chatbot more dynamic and specific.
3.	2018	Ticketing Chatbot Service using Serverless NLP Technology	English	1.Node JS Webhook: a. Webhook is HTTP call-back. b. It work on post and get request, then request is routed by node JS. c. It used in NLP to make data process more flexible. 2.Wit.AI NLP Services: a. It is used for ML and AI purposed. b. Wit.AI is trained with understanding i.e. combination of entity and intent Recognition. c. It trained with some keyword that user usually writes in chat.	1.Specific domain chatbot can redefine chat experienced with the automated response and also some CUI response that guided the user. 2. Presence of more database /data result in more intelligent chatbot. 3. Classification of intent in term of the sentence needs to be more fluent to consider confidence rate. In the future, chat history can be considered as chat experience so the behaviour of the user can be analysed.
4.	2018	College Enquiry Chatbot Using Knowledge in Database	English	A. Online Enquiry 1) Students can enquiry about faculties and query related to exams B. Online Chat Bot 1) The result can be show in images, cards format 2) The query will be answered on basis of question asks and language model built in LUIS and responses are stored.	User will interact with the web interface of the bot. the web application will be connected to the bot with the help of the bot connector which will create the object through which we will communicate with bot. the query entered by the user will be sent to LUIS.ai where it will be processed and Intent and Entity of the user query will be retrieved and the corresponding response will be fetched from database.

5.	2017	AMBER Chatbot and Detection of Paraphrases for Devnagari	Marathi	1.AMBER stands for A Marathi Based Eliza Chatterbot 2. AMBER carries out various Natural Language Processing tasks such as Transliteration, Tokenization, Part of Speech. 3.AMBER uses Sentence Generation 4.It uses Paraphrasing detection Library 5.It uses several frameworks for measuring semantic similarity DKPro Similarity. 6. Word Alignment Algorithm is used to compute alignment between idioms and phrasal chunks. 7. RDRPOST Tagger is used for obtaining POS Tags.	1. Paraphrase is a process of computing the semantic similarity between sentences which are not lexicographically similar. 2. Chatbot is used to experience chat with automated response. 3. We can use paraphrase Identification for Machine Translation, Question Answering Machine, and Multiple Text Summarization.

So, according to the survey studied we get to know that there is no as such efficient work done for Marathi language.

3. PROPOSED SYSTEM

In this system chatbot is designed for question-answering purpose. The Marathi language chatbot is used to interact with user. All the data important for answering the question is available on the portal. It is a communiqué simulating computer program. It is all about the conversation with the user. The chat with a Chatbot is very simple. It answers to the questions asked by the user. During designing a Chatbot, how does the Chatbot will communicate to the user? And how will be the conversation with the user and the Chatbot is very important topic. For creating a Chatbot, a program has to be written in python programming language is used for programming. The Chatbot is created in such a way to help the user by answering their all queries, improve the communication and amuse the user. The chat is formed using a pattern that is known to the user and could be easy to understand. Chat dialog box show up to create conversation. This dialog box is created using python coding. Pattern Matching is a method of artificial intelligence used in the design of a Chatbot. The input is compared with the inputs saved in the database and matching response is returned. Text recognition classification is used for large size of data or files. It is easy to extract text data from such large files using OCR technology. Text detection, localization and tracking modules are correlated to each other and constitute the most stimulating and difficult part of extraction process. By using tesseract OCR with python system, we extract the text from images, pdf or other multiple files.

4. SYSTEM ARCHITECTURE

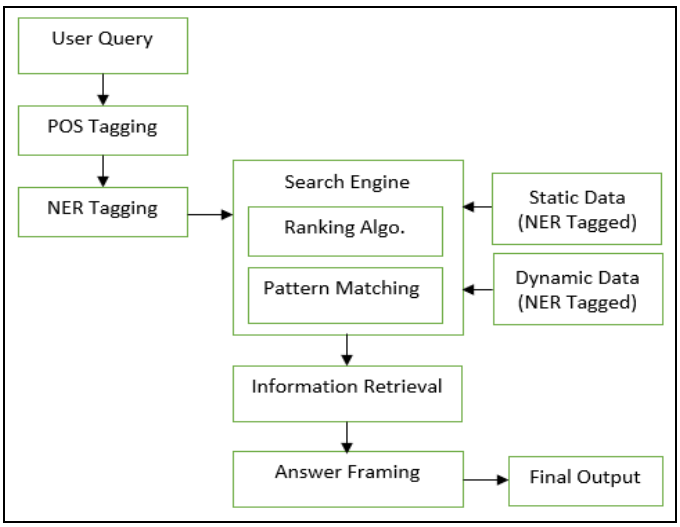

Figure 1: system architecture

In System architecture System Modules are Chatbot, OCR Text Recognition, Related Data, System Interface. The Chatbot Module uses the python language for user interface and pattern matching. In OCR python uses tesseract OCR library for text detection and character recognition. User input data as a text and chatbot system response in text format in chatbot dialog. The pdf, Doc, Image file given as input to the OCR system for text data. The system provides all data to the user on given portal.

5. Modules

There system is divided into four parts according to functions performed by individuals. In this system we need to collect the chatbot, data from intended resources.

5.1. Chatbot:

Chatbot: For creating a Chatbot, a program has to be written in python programming language is used for programming. The Chatbot is created in such a way to assist the user, improve the communication and amuse the user.

Chat: The chat is created using a pattern that is known to the user and could be easy to understand. Chat dialog box show up to create conversation. This dialog box is created using python design and is added on the portal as plugin.

POS Tagging: It is the procedure of identifying correct tag like noun, adjective, verb, adverb etc. to each word of the input sentence.

NER Tagging: Named entity recognition is the procedure of distinguishing named entities which are present in database. NER mostly used and applied for information retrieval.

Ranking Algorithm: Once we get our question with NER tags, then we can give the weightage to our same tagged dataset as we can find answer to the query in that dataset.

Pattern Matching: It is a method of artificial intelligence used in the design of a Chatbot. The input is matched with the inputs saved in the database and consistent response is given back.

Simple: The architecture of a Chatbot is very simple. It just answers to the questions queried by the user, if the question is found in the database.

5.2. OCR Text Recognition:

Text Detection: This phase takes image or video frame as input and decides it contains text or not. It also identifies the text regions in image.

Text Localization: Text localization merges the text regions to formulate the text objects and define the tight bounds around the text objects.

Text Tracking: This phase is applied to files data only. For the readability purpose, text embedded in the files appears in more than thirty consecutive frames. Text tracking phase exploits this temporal occurrence of the same text object in multiple consecutive frames. It can be used to rectify the results of text detection and localization stage. It is also used to speed up the text extraction process by not applying the binarization and recognition step to every detected object.

Character Recognition: The last module of text extraction process is the character recognition. This module converts the binary text object into the ASCII text.

5.3. Related Data: All the schedule data for year is collected as system data for user help. This portal help user with all data on a single platform in Marathi language.

5.4. System Interface: The system interaction with internal module is done for better user interface.

6. Future Scope

The proposed chatbot can answer to only textually typed question which can also be implemented for voice-based question answering system and can also be implemented for other reginal language.

7. Conclusion

This paper enlightens all the issue, challenges and issue faced by system to provide efficient help for student, college, teachers and all other sub parts. The advance chatbot system with Marathi language is used for better user interface for user to get all data and answer in Marathi language. The system makes simple user interface so they easily interact with system. All the schedule data is available on single portal. Chatbot with OCR for large amount of data processing is available in this system.

REFERRENCES

[1] A.M Rahman, Alma Islam, "Programming Challenges of Chatbot.", 2017 IEEE Region 10 Humanitarian Technology Conference, (21-23 Dec 2017).

[2] Belfin R.V, Shobhana A.J, Megha Manilal, Ashly ann Mathew,Blessy Babu, "A Graph Based Chatbot for cancer Patients",5th international conference on advanced computing & communication systems (ICACCS),(2019).

[3] Eko Handoyo, M.Arfan,Yosua Alvin Adi Soetrisno, Maman Somantri, Aghus Sofwan, Enda Wista Sinuraya, "Ticketing Chatbot Service using Serverless NLP Technology", Proc. Of 2018 5th Int. conf. on information Tech., Computer and Electrical Engineering (ICITACEE), (2018).

[4] Harsh Pawar, PranaPrabhu, Ajay Yadav, Vincen Mendonca, Joyce Lemos, "College Enquiry Chatbot Using Knowledge in Database", International journal for research in applied science & engineering technology (IJRASET), (April 2018).

[5] Shah Manthan, Jigneshkumar, Arnav Mediratta, Akshay Kundale, Shubham Nangare, "AMBER Chatbot and Detection of Paraphrases for Devnagari", Vishwakarma Journal of engineering research, (4 Dec 2017).

[6] Partrick Jansson, Shuhua Liu, "Detection of knowledge and emerging named entity from social media by using LMST Model", International Conference on Informatics, Electronics & Vision, India, pp. 1-5, (2013).

[7] Mohamed Nabil Allam and Mohamed Hassan Haggag, "Question-Answering System: A survey", International Journal of Research and Reviews in Information Sciences (IJRRIS), (September 2012).

[8] Namrata G. Kharate and Dr. Varsha H. Patil, "survey of machine translation for Indian languages to English and its approaches", "International journal of scientific research in computer science, engineering and information technology", (2018).

[9] M Dahiya, "A tool for conversation: Chatbot", "International journal of computer science and engineering", (2017).