
ACKNOWLEDGEMENT

I bow down to the almighty for His blessing, Reverend Karnataka Ratna Dr. Sree Sree Sree Shivakumara Maha Swamigalu and His Holiness Sree Sree Siddalinga Swamigalu.

My special thanks to Dr. H.P.VeerabhadraSwamy, Principal, Sree Siddaganga College of Arts, Science and Commerce, Tumkur for providing us all the facilities to carry out the project work.

I express my gratitude to Prof. K.S.Lingadevarappa, Head of the Department, Department of Computer Science, Sree Siddaganga College of Arts, Science and Commerce, B H Road, Tumkur for his valuable guidance, support and encouragement throughout the period of this project work.

It is a matter of pride to express my deep sense of gratitude to Prof.MAMATHA.M., Department of Computer Science, Sree Siddaganga College of Arts, Science and Commerce, Tumkur for his guidance indispensable support, encouragement and valuable suggestions.

I thank my classmates, who helped us throughout the project work through their valuable discussions and suggestions.

I express my deep gratitude towards my beloved parents for their encouragement, inspiration and blessings without which I would not have carried out my studies.

CONTENTS:

- 1. ABSTRACT**
 - 2. INTRODUCTION**
 - 2.1. DOMAIN INTRODUCTION**
 - 2.2. OBJECTIVES**
 - 2.3. PROBLEMS**
 - 2.4. SCOPE**
 - 3. LITERATURE SURVEY**
 - 4. EXISTING SYSTEM**
 - 5. PROPOSED SYSTEM**
 - 6. TECHNOLOGY**
 - 7. REQUIREMENTS**
 - 8. DESIGN**
 - 9. ARCHITECTURE**
 - 10. USE CASE DIAGRAM**
 - 11. DATA FLOW DIAGRAM**
 - 12. CODEING**
 - 13. CONCLUSION**
 - 14. BIBLIOGRAPHY**
-

1.ABSTRACT:

A chatbot is AI computer software that can act as a conversation through textual or auditory methods. The core of chatbots analyses a customer's data using the artificial intelligence which integrates the response with them. Different tasks can be replaced with AI-powered bots as they are much more powerful—and are capable of performing multiple tasks at once. Machine Learning techniques are basically used in the process of understanding the input that we get from the user and replying to the user. Natural language processing allows a bot to have a conversation as naturally as possible. The ideal interaction between user and chatbot is a balanced mix of Innovative technology and human Intervention.

Keywords: Chatbot, Artificial Intelligence, Human conversational partner.

2.INTRODUCTION

2.1 DOMAIN INTRODUCTION

A chatbot is a computer program that uses Artificial Intelligence (AI) and Natural Processing Language (NLP) to understand customer questions and automate responses to them, simulating human conversation. Artificial Intelligence chatbots are chatbots trained to have human-like conversations using a process known as Natural language processing (NLP). With NLP, the AI chatbot is able to interpret human language as it is written, which enables them to operate more or less on their own. In other words, can understand language outside of pre-programmed commands and provide a response based on existing data. This allows site visitors to lead the conversation, voicing their intent in their own words. Chat has become the centre of focus in this current era; thus, the bots are being utilized to deliver information engagingly and conveniently. A chatbot is stand out amongst the most progressive and promising tools of communication among people. Chatbot are not a recent development. They are simulations which can understand human language process it and interact back with humans while performing specific tasks.

2.2 OBJECTIVIES OF CHATBOT

A Chatbot can communicate with a real person behaving like a human. Let's list down objectives of chatbots. You can create chatbots for any business the same as you recruit a person for any department of your company. Whether you are :

- Wedding Planner
- Insurance Assistant
- Education Consultant
- Legal Assistant
- A real estate business
- Recruiter
- Travel Agency
- Hospital or a Beautician!

1. Chatbot are mainly used to provide customer support.
 2. IT helps in creating a huge amount of target audience at the same time 24/7.
-

2.3 PROBLEMS WITH CHATBOT DEVELOPMENT

One of the struggles of simple chatbot technology is the inability to complete backend processes. Instead, they rely on the end-user to take further actions themselves.

Major problems with chatbots and chatbot development:

1. Discovering bots is still an issue for potential
- 2.No established business models
- 3.Most chatbots aren't that great at having normal conversation

2.4. SCOPE OF THE PROJECT:

Innovations and better technology will open up many more applications over the next decade. Businesses, in particular, will benefit from these advancements. For example, new AI chatbots with advanced translation capabilities could help companies expand globally and improve international customer service.

Chatbots are Now Based on Natural Language processing (NLP).

The goal is to allow users and Artificial Intelligence to communicate naturally and understand complex requests. This would mean that customer service agents would be able to focus on other tasks while the AI takes care of customers' queries.

3. LITERATURE SURVEY:

In this section we outline the following main aspects of chatbots based on our finding from the literature review: implementation approaches, available public database used in previous data- driven approaches to chatbot implementation, the main evaluation methods for measuring the performance of chatbots and the application of chatbots in different domains. AI models, contrary to Rule-based models, are based on Machine Learning algorithms that allow them to learn from an existing database of human conversations. In order to do so, they need to be trained through Machine Learning algorithms that can train the model using a training dataset. Through the use of Machine Learning algorithms, there is no longer the need to manually define and code new pattern matching rules, which allows chatbots to be more flexible and no longer dependent on domain specific knowledge. As stated, AI models can be further categorized into Information Retrieval based models and generative models. Information Retrieval Models. Information Retrieval based models are designed so that given a dataset of textual information, the algorithm will be capable of retrieving the information needed based on the user's input. The algorithm used is usually a Shallow Learning algorithm; nonetheless, there are also cases of Information Retrieval models that use Rule based algorithms and Deep Learning ones. Information Retrieval based models include a pre-defined set of possible answers; the chatbot processes the user query and based on this input it picks one of the answers available in its set. The knowledge base for this kind of model is usually formed by a database of question- answer pairs. A chat index is constructed from this database, in order to list all the possible answers based on the message that prompted them. When the user provides the chatbot with an input, the chatbot treats that input as a query, and an Information Retrieval model into those used for web queries is used to match the user's input to similar ones in the chat index. The output returned to the user is thus the answer paired with the selected question among those present in the chat index. The main advantage of this model is that it ensures the quality of the responses since they are not automatically generated. These models have seen a surge in popularity with the advent of the Web 2.0 and the increase in available textual information that could be retrieved on social media platforms, forums, and chats.

4.Existing system:

A chat bot (also known as a talk bot, Bot, chatterbox, Artificial Conversational Entity) is a computer program which conducts a conversation via auditory or textual methods. Such programs are every time designed to determine how a human would behave as a conversational partner, thereby passing the Turing test.

1. Emanuela Haller and Traian Rebedea, “Designing a Chat- bot that Simulates an

Historical Figure”, IEEE Conference Publications, July 2013. There are many applications that are incorporating a human appearance and intending to simulate human dialog, but in most of the cases the knowledge of the conversational bot is stored in a database created by a human experts. However, very few researches have investigated the idea of creating a chat-bot with an artificial character and personality starting from web pages or plain text about a certain person. This paper describes an approach to the idea of identifying the most important facts in texts describing the life (including the personality) of an historical figure for building a conversational agent that could be used in middle-school CSCL scenarios.

Rule based chatbots can’t learn on their own, they only provide answers your legal team provides from a predefined set of rules. That means your clients experience with rule based bots is very linear. In other words if your client asked questions outside its pre-set understanding, they fail and need human intervention.

5. Proposed system:

Context Identification: Pre-processing is applied to the input text to standardize the input as per the system's requirement. Based on the keywords used in the text, appropriate context is recognized.

1. Personal Query Response System (Module-1): Upon receiving personal queries like CGPA, attendance, etc., the authenticity of the user is checked through user-id and password. If the user detail is invalid, an appropriate response is sent. If the user authenticates successfully, the input text is processed to extract keywords. Based on the keywords, information required by the user is understood and the information is provided from the database.

2. AIML Response System (Module-2): If the user is trying to make a normal conversation with the bot, the input is mapped to an appropriate pattern in Artificial Intelligence Modeling Language (AIML) files. If the response is available, it is sent to the user. Other data provided to the chatbot such as username, gender, etc. are also saved. If the pattern is not available in AIML files, a random response is sent suggesting "Invalid Input".

3. Query Analysis and Response System (Module-3): When a user wants some information pertaining to college, the response will be provided through this module. If the input matches a pattern in the AIML files, the appropriate response will be sent to the user. If the AIML files have no entry for that particular query pattern, keywords are fetched from the input. An algorithm to check sentence similarity (NLP) is applied to the modified input to check its similarity with the questions of a predefined question-set, whose answers are available. If a sentence is retrieved with confidence

6. TECHNOLOGY:

ARTIFICIAL INTELLIGENCE:

Artificial intelligence (AI) is intelligence—perceiving, synthesizing, and inferring information—demonstrated by machines, as opposed to intelligence displayed by non-human, animals and humans. Example tasks in which this is done include speech recognition, computer vision, translation between (natural) languages, as well as other mappings of inputs.

MACHINE LEARNING:

The recent advancement in machine learning has made it possible to develop more intelligent chatbots. Chatbots that adopt machine learning approaches use machine learning algorithms to extract information and generate responses and are able to improve through previous conversations. An extensive training set is required for machine-learning-based chatbots. Two types of models can be used, retrieval or generative. Retrieval-based models involve choosing the optimal response from a set of responses, and generative models, on the other hand, use deep learning techniques to generate the response.

NATURAL LANGUAGE PROCESSING:

Natural Language Processing (NLP) is the study of letting computers understand human languages. Without NLP, human language sentences are just a series of meaningless symbols to computers. Computers don't recognize the words and don't understand the grammars. NLP can be regarded as a "translator", who will translate human languages to computer understandable information. Traditionally, users need to follow well-defined procedures accurately, in order to interact with computers.

7. REQUIREMENT:

Hardware Requirement:

- Hardware - Pentium
- Speed - 1.1 GHz
- RAM - 1GB
- Hard Disk - 20 GB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - SVGA

Software Requirements:

- Operating System : Windows / Linux / Mac
 - Technology : ML,AI
-

8. DESIGN:

In this section, we describe a general architectural design and delve deep into the important parts of each component. A general chatbot architecture consists of five main components, Interface component, a Natural Language Understanding (NLU) component, a Dialogue Management (DM) component, a Backend component, and a Response Generation (RG) Component. **Design Principles**

There are two approaches that can be used to develop a chatbot depending on the algorithms and techniques adopted: rule-based approach and machine learning approach.

Machine Learning: The recent advancement in machine learning has made it possible to develop more intelligent chatbots. Chatbots that adopt machine learning approaches use machine learning algorithms to extract information and generate responses and are able to improve through previous conversations. An extensive training set is required for machine-learning-based chatbots. Two types of models can be used, retrieval or generative. Retrieval-based models involve choosing the optimal response from a set of responses, and generative models, on the other hand, use deep learning techniques to generate the response.

Rule-based: A rule-based chatbot processes information and provides responses based on a set of predefined rules with the use of pattern matching algorithms. Although the pattern matching techniques vary in complexity, the basic idea is the same. The user input is classified as a pattern, and the chatbot selects a predefined answer by matching the pattern with a set of stored responses. The pattern and response matching algorithms are handcrafted. Pattern matching is adopted by many chatbots and is especially popular among the early chatbots like ELIZA, PARRY, and ALICE. The advantage of the rule-based approach is its speed as it does not require any deep analysis of the input text. However, the responses are repeated and lack flexibility and originality as the knowledge is set by the developer in advance.

8. DESIGN:

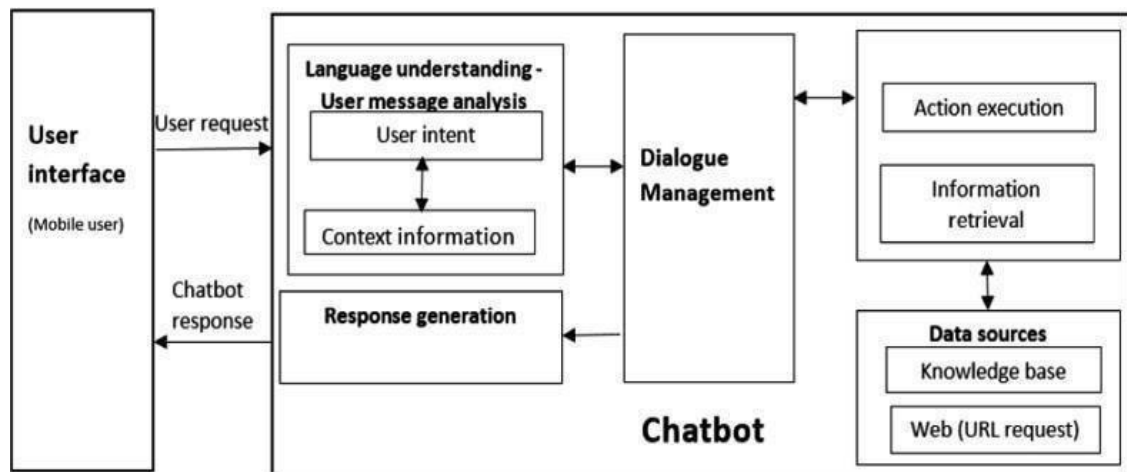
In this section, we describe a general architectural design and delve deep into the important parts of each component. A general chatbot architecture consists of five main components, Interface component, a Natural Language Understanding (NLU) component, a Dialogue Management (DM) component, a Backend component, and a Response Generation (RG) Component. **Design Principles**

There are two approaches that can be used to develop a chatbot depending on the algorithms and techniques adopted: rule-based approach and machine learning approach.

Machine Learning: The recent advancement in machine learning has made it possible to develop more intelligent chatbots. Chatbots that adopt machine learning approaches use machine learning algorithms to extract information and generate responses and are able to improve through previous conversations. An extensive training set is required for machine-learning-based chatbots. Two types of models can be used, retrieval or generative. Retrieval-based models involve choosing the optimal response from a set of responses, and generative models, on the other hand, use deep learning techniques to generate the response.

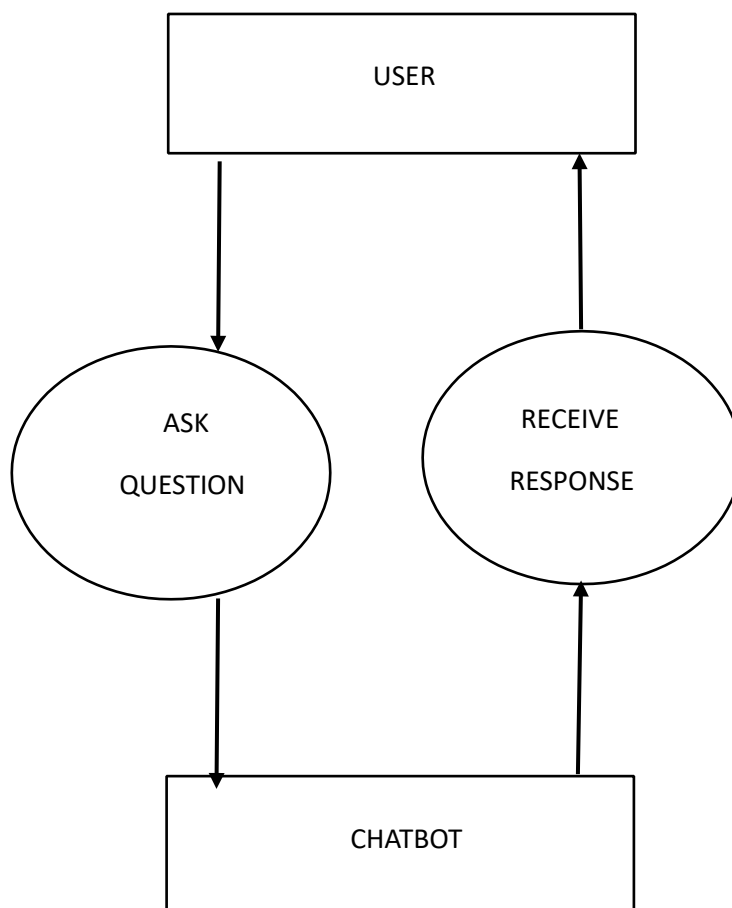
Rule-based: A rule-based chatbot processes information and provides responses based on a set of predefined rules with the use of pattern matching algorithms. Although the pattern matching techniques vary in complexity, the basic idea is the same. The user input is classified as a pattern, and the chatbot selects a predefined answer by matching the pattern with a set of stored responses. The pattern and response matching algorithms are handcrafted. Pattern matching is adopted by many chatbots and is especially popular among the early chatbots like ELIZA, PARRY, and ALICE. The advantage of the rule-based approach is its speed as it does not require any deep analysis of the input text. However, the responses are repeated and lack flexibility and originality as the knowledge is set by the developer in advance.

9. ARCHITECTURE OF CHATBOT:

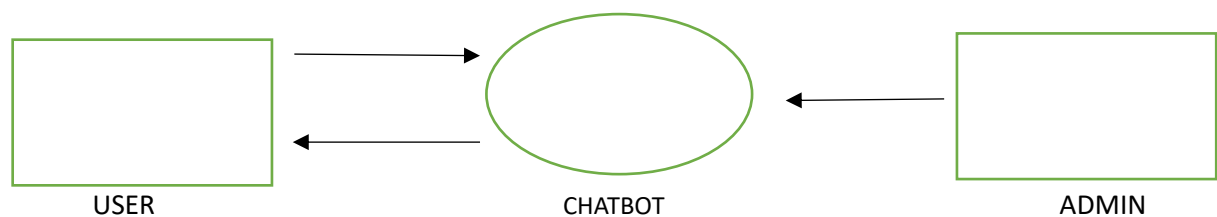


Architecture of chatbot

10. USE CASE DIAGRAM:



11.DFD:



User Interface

The user interface allows users to communicate and interact with a chatbot through messenger applications like Facebook Messenger, Cortana, or Slack. The operation of a chatbot begins with a user's request.

Natural Language Understanding

After the system receives the user's request, it uses the NLU component to extract information from the input and produce a representation of its meaning that can be used later on in the process. NLU generally deals with three tasks, dialogue act classification, intent classification, and slot filling.

Dialogue act classification deals with determining the function of the user's input, or more precisely, mapping the user's utterance to a dialogue act type. The utterance can be classified as a question, a statement, an offer, or some other type of dialogue act. Knowing the dialogue act being performed is critical to better comprehend the user's request and to decide an appropriate response.

Intent classification identifies the primary goal of the user. Intents are mainly domain-independent. For example, a request can be in the domain of food ordering, hotel reservations, weather forecasts, and so on. The intent of an agent within the hotel reservations domain can be to book, cancel or change a reservation, and similarly, the intent of an agent within the food ordering domain can be to place, query, or change an order.

Slot filling is the final step in NLU. The agent extracts other necessary details, which when combined with the dialogue act and the intent, allow it to fully understand the user's request.

Slot filling is the core of NLU and has received the most research attention. The main task of slot filling is to extract information from the input and to better understand its meaning. More specifically, slot filling aims to analyze the context of the input by separating the text into smaller units and then assigning the units tags or labels based on their functions and roles in the text. In this subsection, we will describe several different methods that are commonly used in slot filling.

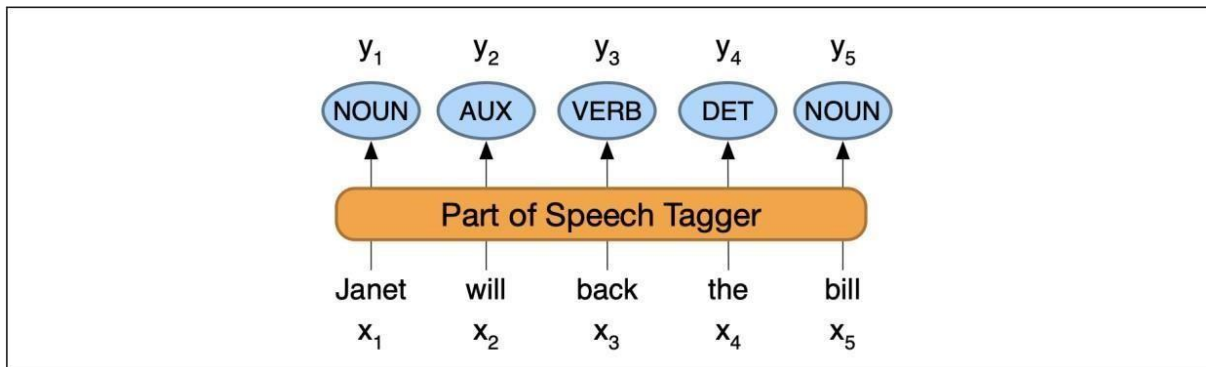
Regular expression is a tool for describing text patterns. Each pattern is specified by a sequence of characters and each character is either a regular character or a metacharacter that has a special meaning.

Tokenization involves breaking up a text into units by words, punctuation marks, or numbers. Generally, English words are separated by white space, so tokenizing an English text should be fairly straightforward. However, there are some problematic cases where the boundaries of words or sentences are ambiguous. Contracted items (e.g. isn't), phrases (e.g. San Francisco), abbreviations (e.g. PhD.), and acronyms (e.g. AT&T) are examples of the special cases. One commonly used tokenization standard is known as the Penn Treebank tokenization standard. According to the Penn Treebank tokenization standard, contractions and punctuation between tokens are separated out and hyphenated words are kept together.

Text normalization refers to converting the text into a standard form. Word normalization is the task of identifying words that have the same meaning but spelled differently (e.g. U.S.A. and USA). Case folding is the task of transforming everything to lower cases. Lemmatization is the process of identifying the ~~roots~~ roots of the words and mapping the morphological variants into their base form (e.g. produce, product, produces, production produce).

Bag of words, also known as the vector space model, is one of the simplest approaches to analyze the input. We count the occurrences of each word but ignore the syntactic information or word order information. This can be done by performing text normalization and eliminating stop words.

Part-of-speech (POS) tagging is the process of marking up each word in the text with a tag that indicates its syntactic role, for example, whether it is a verb, noun, pronoun, conjunction, and so on. POS tagging is especially useful in cases where we encounter words that have different meanings and their use in the sentence may be ambiguous.



Name entity recognition (NER) involves extracting items such as persons, dates, and organizations. POS tagging and NER are disambiguation tasks and provide useful information about sentence structure and meaning. NER is often more complicated than POS tagging because it involves determining the appropriate segmentation .

12.CODING:

13. CONCLUSION:

In this paper, we reviewed the history of chatbots and saw how they have evolved since their inception. We then examined the design principles and general architecture in detail. Next, we presented how we can build specific applications and use cases. Finally, we discussed the future direction of research as well as ethical and social considerations during the design and development stages. Enhancing language comprehension and generation ability is a critical step in future development.

14. BIBLIOGRAPHY:

- *. <https://github.com/Abhinav-26/AI-chatbot>
 - *. <https://builtin.com/>
 - *. <https://www.chatbot.com/>
 - *. <https://www.ibm.com/>
 - *. <https://www.ncbi.nlm.nih.gov/>
-