# Chatbot for healthcare system using Artificial Intelligence

*Kavitha B. R.*
*kavithabr.sse17@rvce.edu.in*
*RV College of Engineering, Bengaluru, Karnataka*

*Dr. Chethana R. Murthy*
*chethanamurthy@rvce.edu.in*
*RV College of Engineering, Bengaluru, Karnataka*

## ABSTRACT

*The Chatbots are the computer programs that interact with the users using natural language. The chatbot stores the information in the database to identify the keywords from the sentences and make a decision for the query and answers the question. In this paper keyword, ranking and sentence similarity calculation are done using n-gram, TF-IDF and cosine similarity. From the given input sentence, the score will be obtained for each sentence and more similar sentences are obtained for the given query. The query posed to the bot which isn't comprehended or not present in the database is further processed by the third party, expert system.*

*Keywords— N-gram, TF-IDF, Cosine similarity*

## 1. INTRODUCTION

Computers give us information; they engage us and help us in a lot of manners. A chatbot is a program intended to counterfeit smart communication on a text or speech. Yet, this paper concentrates only on text.

These systems can learn themselves and restore their knowledge using human assistance or using web resources. This application is incredibly fundamental since knowledge is stored in advance. The system application uses the question and answer protocol in the form of a chatbot to answer user queries. This system is developed to reduce the healthcare cost and time of the users, as it is not possible for the users to visit the doctors or experts when immediately needed.

The response to the question will be replied based on the user query and knowledge base. The significant keywords are fetched from the sentence and answer to those sentences. If the match is discovered or the significant answer will be given or similar answers will be displayed.

The complex questions and answers present in the database are viewed and answered by an expert. Here the users can personally ask any questions regarding healthcare, as not much time will be wasted by the user for consulting a doctor. The input sentence of the chat pattern is stored in an RDBMS. The chatbot would coordinate the input sentence from the user question with the knowledge base. Each query is compared with the knowledge database of the chatbot. The important keywords are extracted from the given input sentence and the sentence similarity is found. The keyword ranking and sentence similarity are found using the N-gram, TF-IDF, and cosine .similarity. The interfaces are standalone built using the JAVA programming language.

## 2. LITERATURE SURVEY

Here the studies are based on to recognize emotions classification using AI methods. The studies train emotions classification models from a lot of labelled data based on RNN, deep learning, convolutional neural network. Linguistic interaction is most important in counselling using NLP and NLG to understand dialogues of users. Here the multi-modal approach is used of emotion-recognition. They have collected corpuses to learn semantic information of words and represent as vector using the word vector, synonym knowledge of lexical are collected. [1]

In this paper a voice recognition chat-bot is developed, if the questions are not understood asked to the bot is further processed using the third party expert-system. The web-bots are created as text-based web-friends, an entertainer for the user. Here they focused on the improved system if the system is not only text-based but also voice-based trained. Here the voice recognition requires a 2 part process of capturing and analysis of an input signal. Server response recognition data retrieval and information output. The server used here is SOAP based on black box approach. The use of expert system allows unlimited and autonomous intelligence improvements. [2]

This chatbot aims to make a conversation between human and machine. Here the system stores the knowledge database to identify the sentence and making a decision to answer the question. The input sentence will get the similarity score of input sentences using bigram. The chatbot knowledge is stored in RDBMS. [3]

The chatbot implemented using pattern comparison in which the order of the sentence is recognized and saved response pattern. Here the author describes the implementation of the chatbot Operating system, software, programming language, and database. How results input and output is stored. Here the input is taken using text () function and other punctuation is removed using trim () function and random () function is used

to choose a response from the database. The chatbot is used for an entertainment purpose. [4]

Here they use n-gram technique for extracting the words from the sentences. Here n-gram is used for comparison and deduction of the input with case data using Moro phonemes and phonemes as the deciding parameter. Probability analysis for the closest match is performed. The final expression is redirected through an expert system. [5]

The chatbot developed here for healthcare purposes for the android application. The user sends the text message or voice message using Google API. Here the user gets only related answer from the chatbot. SVM algorithm is used to classify the dataset. Here the Porter algorithm is used to discard unwanted words like suffixes or prefixes. [6]

The different documents served in web, the content is checked by tagging the dataset using n-gram based low dimensional demonstration, TF-IDF matrix that generates S, U, and V and finally multiplying the 3 matrices cosine similarity is calculated. [7]

Here the chatbot is created for the customer service that functions as public health service. The application uses N-gram, TF-IDF and cosine similarity. The knowledge base is created for storing the question and answer. The application clearly shows extracted the keyword from the question ad by using unigram, bigram, and trigram which helps in fast answering. [8]

## 3. PROPOSED SYSTEM

The chatbots are conversational virtual assistants which automate interactions with the users. Chatbots are powered by artificial intelligence using machine learning techniques to understand natural language. The main motive of the paper is to help the users regarding minor health information. Initially when the user's visits the website first registers themselves and later can ask the bot, queries. The system uses an expert system to answer the queries if the answer is not present in the database. Here the domain experts also should register themselves by giving various details. The data of the chatbot stored in the database in the form of pattern-template. Here SQL is used for handling the database.
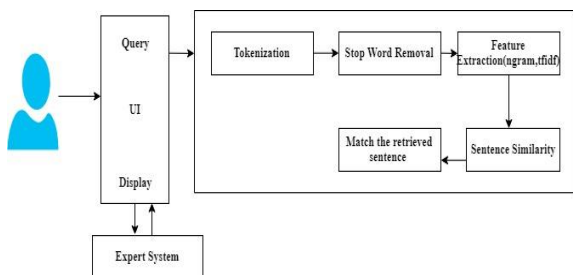
## 4. SYSTEM ARCHITECTURE



**Fig. 1: System architecture**

Figure 1 is the system Architecture outline of chatbot healthcare application. The client inputs the question in the UI as the text. The UI gets the user query and after that sends it to the chatbot application. In the chatbot application, the literary experiences pre-processing steps incorporate tokenization where the words are tokenized, at that point the stop words are removed and feature extraction depends on n-gram, TF-IDF, and cosine likeness. The question answers are stored in the knowledge database to recover the retrieve the answer.

### 4.1 Tokenization
The words or sentences separated word by word for increased processing. It separates text into words at whatever point it experiences one of the rundowns of indicated character. All the words are separated from sentences and the punctuation are disposed of. This implies the next steps.

### 4.2 Stop words removal
The stop words are removed from the sentences to extract important keyword. It is mainly employed to remove unnecessary things such as words occurring too frequently in sentences. It is also used to delete words that are not important or the words with no specific meanings such as an, a, or the. This step is applied to reduce processing time or computational complexity.

### 4.3 Feature extraction based on N-gram TFIDF
Feature extraction is a characteristic decrease process in the document; it ranks the attributes as per the document. By doing this step it upgrades the speed and adequacy of the document. It is used to extract the set of keywords and frequency of the keywords in the document.

**4.3.1 TF-IDF:** Term frequency and Inverse document frequency is used to calculate the weight of each term in the sentence.

The term frequency is used to check how many times the term as been occurred in a particular sentence using the formula below.

$$\mathbf{tf = tf_i} \qquad (1)$$

IDF used to compute the weight of uncommon words over all reports in the document. The words that appear in a while in the document have a high IDF score. It is given by the condition underneath

$$\mathbf{idf = log \frac{N}{df}} \qquad (2)$$

The tf and idf are combined to produce the weight of the term or word in the document. The tf and idf values are multiplied to obtain the weight of each term in the document.

$$\mathbf{W_i = tf_i * \ log \frac{N}{df}} \qquad (3)$$

**4.3.2 N-gram:** N-gram is an endeavour to expand N-gram models with variable length arrangements. A sequence can be a grouping of words, word class, grammatical feature or whatever a succession of something that the modeller thinks bearing significant language structure data. In this system, N-gram is used for text compression or reduce the data space in the document, to extract the relevant keywords from the database.

### 4.4 Sentence similarity
Cosine similarity is been used to check the similarity between two sentences. The similarity between the query and document is directly proportional to the number of query weights. The similarity calculation result of the two documents ranges from 0 to 1 since the term frequency cannot be negative. The formula for calculating cosine similarity is given below:

$$similarity = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

**4.5 Retrieve the matched sentence**

The answers for the query which are obtained from the above process are retrieved and displayed in the user interface.

## 5. RESULTS AND ANALYSIS

The application uses a question and answers protocol where it consists of login page figure 2, where the user needs to give the details to register in the application if you are a new user, figure 3. Figure 4 shows the answer for the query is available in the database or displays similar answers for the query, figure 5. Figure 6 is the expert answering page where experts answer directly to the user's question. The application uses n-gram for text compression using bigram and trigram for faster execution of the query. The N-gram, TF-IDF, and cosine similarity to convey the answers to the users.
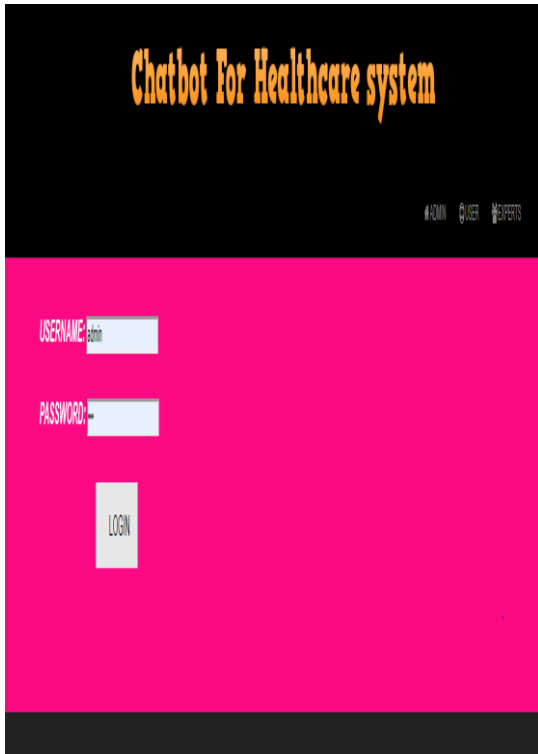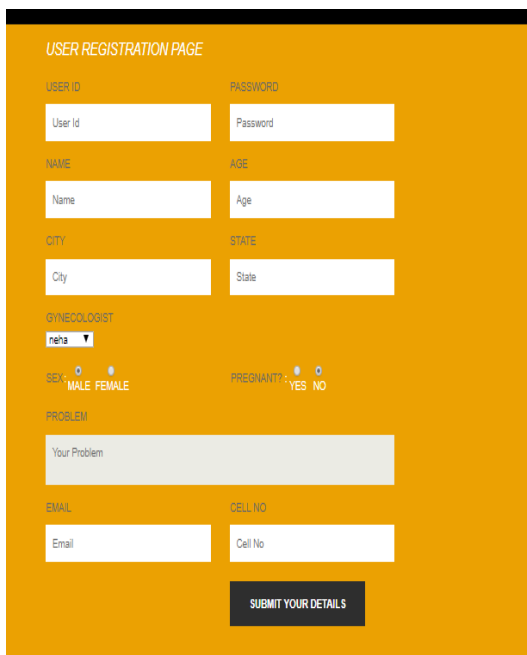


**Fig. 2: Login page**



**Fig. 3: Registartion page**



**Fig. 4: User inputs query**



**Fig. 5: Similar answers are displayed**



**Fig. 6: Expert Answering Page**

# 6. CONCLUSION

A chatbot is a great tool for conversation. Here the application is developed to provide quality of answers in a short period of time. It removes the burden from the answer provider by directly delivering the answer to the user using an expert system. The project is developed for the user to save the user their time in consulting the doctors or experts for the healthcare solution.

Here we developed the application using the N-gram, TF-IDF for extracting the keyword from the user query. Each keyword is weighed down to obtain the proper answer for the query. The Web-interface is developed for the users, to the input query. The application is improved with the security and effectiveness upgrades by ensuring user protection and characters and retrieving answers consequently for the questions.

# 7. REFERENCES

[1] K. Oh, D. Lee, B. Ko and H. Choi, "A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation," 2017 18th IEEE International Conference on Mobile Data Management (MDM), Daejeon, 2017, pp. 371-375. doi: 10.1109/MDM.2017.64

[2] Du Preez, S.J. & Lall, Manoj & Sinha, S. (2009). An intelligent web-based voice chat bot. 386 - 391.10.1109/EURCON.2009.5167660

[3] Bayu Setiaji, Ferry Wahyu Wibowo, "Chatbot Using a Knowledge in Database: Human-to-Machine Conversation Modeling", Intelligent Systems Modelling and Simulation (ISMS) 2016 7th International Conference on, pp. 72-77, 2016.

[4] Dahiya, Menal. (2017). A Tool of Conversation: Chatbot. INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING. 5. 158-161.2017.

[5] C.P. Shabariram, V. Srinath, C.S. Indhuja, Vidhya (2017). Ratatta: Chatbot Application Using Expert System, International Journal of Advanced Research in Computer Science and Software Engineering,2017

[6] Mrs Rashmi Dharwadkar1, Dr.Mrs. Neeta A. Deshpande, A Medical ChatBot, International Journal of Computer Trends and Technology (IJCTT) – Volume 60 Issue 1-June 2018

[7] Farheen Naaz, Farheen Siddiqui, modified n-gram based model for identifying and filtering near-duplicate documents detection, International Journal of Advanced Computational Engineering and Networking, ISSN: 2320-2106, Volume-5, Issue-10, Oct.-2017

[8] N-gram Accuracy Analysis in the Method of Chatbot Response, International Journal of Engineering & Technology. (2018)

[9] Aishwarya Hajare, Priyanka Bhosale, Rasika Nanaware, Guruswami Hiremath, Chatbot for education system ISSN: 2454-132X, Impact factor: 4.295, Volume 4, Issue 3, IJARIIT.

[10] Ardi, Hario Laskito. "Support Vector Machine Classifier for Sentiment Analysis of Feedback Marketplace with a Comparison Features at Aspect Level." (2017).

[11] Dimple veer Singh, Sumit Malhotra (2018). Intra News Category Classification using N-gram TFIDF Features and Decision Tree Classifier, IJSART.