

# Building Generic Code-Mixed Chatbot for Banking System

Khasha Pavan Kalyan Raju<sup>1</sup>, Mancha Vinay<sup>2</sup>, Chintam Ravi Chandra<sup>3</sup>, Sanjay Chatterji<sup>4</sup>

*Department of CSE, Indian Institute of Information Technology Kalyani  
Nadia, Kolkata, India.*

{<sup>1</sup>khasha,<sup>2</sup>vinay,<sup>3</sup>rchintam\_bt17,<sup>4</sup>sanjayc}@iiitkalyani.ac.in

**Abstract**—Banking is very important to lead a good economic life. Banking Issues faced by people such as waiting in queues for knowing simple queries like knowing their balance, how to block ATM card etc. can be solved using chatbots. The idea is to create a Banking chatbot that can take query given by the user and provide some basic details related to the query. This will help to reduce waiting time of the user and also improve quality of customer service. The chatbots are computer programs that use natural language to understand human queries and respond accordingly in a conversation. Chatbot answer the query with the help of data that is stored in the Database. Ranking and sentence similarity calculation is performed using cosine similarity. The score will be obtained for each sentence from the given input sentence and more similar sentences will be obtained for the query given.

**Keywords**— Chatbot, Banking, Code-Mixed, Cosine Similarity, Multi language.

## I. INTRODUCTION

Computers, they engage us and help us in a lot of manners by running different kinds of programs. A chatbot is one such program which is intended to simulate smart communication with the user either by text or speech. This paper focuses only on text based chatbots. These chatbots Use existing conversation data (if available) or data gathered from human assistance and web resources to understand the type of questions people ask. This application is elementary as we have existing data or knowledge base.

The system application uses the question-and-answer protocol in the form of a chatbot to answer user queries. The idea of the system is to reduce the waiting time of the users and provide quality service, as it is difficult for the users to visit the banks or any related services for basic queries.

The response to the query given by the user will be replied based on the user query and knowledge base. The significant keywords are fetched from the sentences, then find the most related answer to those sentences. If the most related answer is discovered, it 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 banking, as not much time will be wasted by the user for visiting a bank. 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 keyword ranking and sentence similarity are found using the cosine similarity. An app interface is built using the JAVA programming language and an API is created using the FLASK framework.

## II. RELATED WORK

Currently chat bots are being developed using variety of methods like rule based where rules are hard-coded in code, AI based bots, pattern-based which can handle only mentioned patterns for retrieving answer. There are frameworks available for developing chat bots but they also use either rule-based or pattern-based techniques.

Research has been done in the recent past on code-mixed data, and especially involving language tagging. Jhamtani et al. (2014) created an ensemble model by combining two classifiers to create a Hindi-English code-mixed LID. The first classifier used word frequency, modified edit distance, and character n-grams as features. The second classifier used the output from the former classifier for the

current word, along with language and POS tag of neighbouring words to give the final tag.

The recent work in the area of natural language processing has contributed valuable solutions to calculate the semantic similarity between words and sentences. This section reviews some related work to investigate the strengths and limitations of previous methods and to identify the particular difficulties in computing semantic similarity. Related works can roughly be classified into following major categories:

- Word co-occurrence methods
- Similarity based on a lexical database
- Method based on web search engine results

### III. ARCHITECTURE

A query is retrieved from the user interface and is sent to the backend API as shown in Figure 1. The query is then sent to the language identification system that helps us to differentiate a code-mixed text apart from a normal text. If it is a code-mixed query then it is converted to a monolingual text and is passed through pre-processing else the normal query is directly passed to the pre-processing stage.

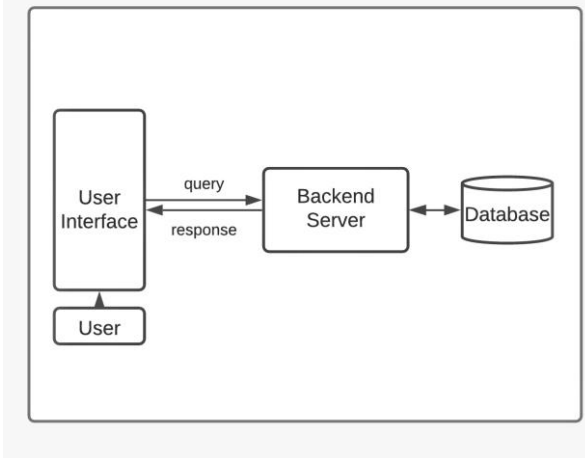


Figure.1. System Architecture

#### Code-Mixed Text Conversion:

Code-Mixed text are text which have mixed words of English, Telugu in English and Telugu in UTF format. The inputted query which is a code-mixed text is converted to a monolingual Telugu text by the following steps:

1. The English words are identified by using a bag of words (around 57000 English words) and are translated to their respective Telugu UTF format using Google translation.

2. The words which are not English are identified as Telugu words and are transliterated to their respective Telugu UTF format using Google transliterate API.



Figure.2. User Interface

#### Text Pre-Processing:

The query is then cleaned by removing punctuations, special characters etc. and then it is vectorized in which the text is mapped into words or phrases from vocabulary to a corresponding vector of real numbers which used to find text similarities.

#### Data:

We prepared a dataset of 30 most frequently asked question and answers related to banking from different banking web pages and converted them into two other languages which are Hindi and Telugu respectively, the data is stored similar to the example data shown in Figure 2. These collected data are stored in the API in a dictionary format

where question is the key and value are the respective answer.

- **English:** How to block a ATM card? >> Debit card can be blocked as below  
Through SMS:SMS syntax for blocking ATM/Debit Card:  
CARDBLOCK<SPACE>XXXX (XXXX Last four digits of A/c No.) to 56161
- **Telugu:** ఎటిఎం కార్డును ఎలా బ్లాక్ చేయాలి? >> డెబిట్ కార్డును ఈ క్రింది విధంగా బ్లాక్ చేయవచ్చు ఎస్ఎంఎస్ ద్వారా: ఎటిఎం / డెబిట్ కార్డును నిరోధించడానికి ఎస్ఎంఎస్ సింట్లాక్స్: కార్డ్బ్లాక్ <SPACE> XXXX (XXXX చివరి నాలుగు అంకాలు A / c నం) నుండి 56161 వరకు
- **Hindi:** एटीएम कार्ड को कैसे ब्लॉक किया जा सकता है? >> डेबिट कार्ड को एसएमएस के माध्यम से नीचे ब्लॉक किया जा सकता है: एटीएम / डेबिट कार्ड को ब्लॉक करने के लिए एसएमएस सिंटैक्स: कार्डब्लॉक <SPACE> XXXX (XXXX पिछले चार अंकों के ए / सी नंबर) से 56161

Figure.3. Structure of Data

### Sentence Similarity:

Cosine similarity is being used to check the similarity between two sentences. A probability is assigned to the query and each question in the database using Cosine Similarity. The most probable question is found among the questions from the database which have probability above a certain threshold.

$$\text{similarity}(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}$$

### Retrieve the matched sentence:

The respective answer related to the Most probable question is retrieved from the and displayed in the user interface. If there is no most probable query among the questions in the database that crossed the threshold then a default response is displayed to the user.

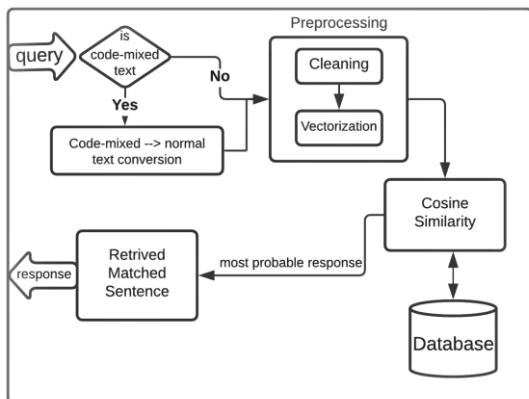


Figure.4. Backend Architecture

## IV. RESULTS AND ANALYSIS

We have tried different queries to validate the implementation of banking chatbot. The responses are expected/correct 78.9% of the total queries (19 queries) and the rest are Incorrect.

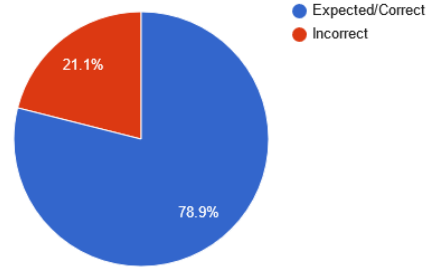


Figure.5. Statistics of Results

## V. CONCLUSIONS

A Chatbot is a great tool for conversation. Here the application we created helps to reduce waiting time of the user and also improve quality of customer service. It removes the burden from the bank service by directly delivering the answer to the user using an expert system and removes the burden for the user to visit the bank for every simple query. The user interacts within the chat interface which uses the cosine similarity to obtain the proper answer for the query. The App interface is developed for the users, to gain knowledge of banks within their reach.

## REFERENCES

- [1] Vinod Kumar Shukla, Lekha Athota, Ajay Rana, Nitin Pandey. Chatbot for Healthcare System Using Artificial Intelligence (2020).
- [2] Sainik Kumar Mahata, Soumil Mandal, Dipankar Das, Sivaji Bandyopadhyay. Code-Mixed to Monolingual Translation Framework.
- [3] Alfirna Rizqi Lahitani, Adhistya Erna Permanasari, Noor Akhmad Setiawan. Cosine similarity to determine

similarity measure: Study case in online essay assessment.

International Journal of Advanced Research in Computer Science and Software Engineering,2017

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

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

[5] C.P. Shabariram, V. Srinath, C.S. Indhuja, Vidhya (2017). Ratatta: Chatbot Application Using Expert System,