



Hindi Chatbot

¹Pravin Bhosale, ²Taj Mohammad Ansari, ³Mohammad Ayaj Ansari, ⁴Smt. Sneha G. Farkade

^{1,2,3} BE Computer Engineering, Government College Of Engg and Research, Avasari Kh Pune

⁴Assistant Professor, Department of Computer, Engineering, Government College Of Engg and Research Avasari Kh Pune

¹pravin.bhosale122020@gcoeara.ac.in, ²tajmohammad.ansari122020@gcoeara.ac.in, ³mohammadayaj.ansari122020@gcoeara.ac.in,

⁴sgfarkade.comp@gcoeara.ac.in

ABSTRACT-

This research paper presents a Hindi Chatbot. With the rapid advancement of artificial intelligence and NLP, chatbots have emerged as versatile tools for human-computer interaction. In this Report, the development of a Hindi chatbot holds immense potential for improving user experiences, accessibility, and engagement for Hindi-speaking populations. This report presents the design, development, and evaluation of a Hindi chatbot that employs state-of-the-art language processing techniques to facilitate effective and intuitive communication in the Hindi language.

Keywords- Hindi Chatbot, NLP Information Retrieval, Human-Computer Interaction, Language Understanding, Data Preprocessing.

I. INTRODUCTION

In today's world, chatbots are everywhere, making it easier for businesses to talk to their customers. But there's a problem: most chatbots only speak English. This is a big issue for people who speak Hindi, one of the world's most common languages. So, we need better chatbots that can understand and speak Hindi just as well as English. This is important because

There are millions of Hindi speakers in India and around the world. We want everyone to have access to the same great digital experiences, no matter what language they speak.

To fix this, researchers are starting to work on Hindi chatbots. These chatbots are designed specifically for Hindi speakers. They can help in many areas, like Hotel ordering, healthcare, education, and customer service.

For example, a Hindi healthcare chatbot could give medical advice to people who speak Hindi. This is important because it means everyone can get the help they need, no matter what language they speak [1].

Some studies have already shown the potential of Hindi chatbots. For instance, researchers have made chatbots to help cancer patients and handle ticket bookings. These chatbots can make life easier for Hindi speakers by providing support in their language. But there are still challenges, like making sure the chatbots understand Hindi well and respect Hindi culture. So, more work is needed to make these chatbots even better [2].

Our initial step involves delving into numerous research papers to gather insights and understand the existing methodologies. By studying some research papers, including the all referenced below, we aim to gain a comprehensive understanding of Natural Language Processing (NLP) and its applications in chatbot development.

We research NLP and we say that NLP serves as the backbone of our chatbot's intelligence, enabling it to comprehend and generate responses in Hindi. Through NLP, our chatbot learns to

understand the nuances of human language, such as grammar, context, and intent. This understanding is crucial for ensuring that the chatbot can interact effectively with users, providing accurate and relevant information or assistance.

By leveraging insights from research papers like those referenced, we aim to enhance our understanding of NLP techniques and algorithms. These techniques include Named Entity Recognition (NER), Trigrams 'n' Tags (TnT) tagging, and others mentioned in the literature. Each of these algorithms plays a pivotal role in enabling the chatbot to understand user queries, identify important entities, and generate contextually appropriate responses.

Through our investigation and analysis of existing research, we seek to identify areas for improvement and optimization in our chatbot development process. This may involve experimenting with different algorithms, fine-tuning parameters, or exploring novel approaches to enhance the chatbot's performance and user experience.

To create an effective chatbot for hotel food ordering, understanding the needs and requirements of users is paramount. Our chatbot aims to streamline the process of ordering food from hotels, providing convenience and efficiency to users.

Firstly, we study all other chatbots. One thing is to observe the complex structure of the chatbot for this overcome by understanding that the chatbot needs to have a user-friendly interface that allows seamless interaction. Users should be able to easily navigate through the chatbot's functionalities, browse menu options, and place orders effortlessly. Clear prompts and intuitive commands will be essential to guide users through the ordering process.

Additionally, the chatbot must be equipped with a comprehensive menu database, including various cuisines, dishes, and customization options. Users should have access to detailed descriptions of each menu item, including ingredients, portion sizes, and pricing information. This ensures that users can make informed decisions and tailor their orders according to their preferences and dietary requirements. Overall, the primary goal of our hotel food ordering chatbot is to simplify the food ordering process for users, offering convenience, choice, and personalized service. By understanding and addressing the needs

of users, we aim to create a chatbot that enhances the dining experience and fosters customer satisfaction. Handling long documents or capturing nuanced information accurately.

Our chatbot must be adept in comprehending and responding to user queries in Hindi, ensuring accessibility and inclusivity for Hindi-speaking individuals. By incorporating advanced Natural Language Processing (NLP) techniques tailored to Hindi, such as sentiment analysis and language understanding models trained on Hindi datasets, we aim to enhance the chatbot's linguistic capabilities and conversational fluency.

Our study has a few main goals. To develop a chatbot that is capable of understanding and responding to user queries in Hindi. To create an engaging and interactive conversational experience. To provide a robust customer support. To utilize cutting-edge NLP techniques to improve the chatbot's working. To build a chatbot that serves multi-functionalities such as placing new orders and tracking previous orders.

II PROPOSED WORK

1. Natural Language Processing (NLP):

Natural Language Processing (NLP) forms the foundation of our chatbot's ability to understand and generate responses in Hindi. In this phase of the proposed work, we will conduct a detailed study of NLP techniques and algorithms tailored to the Hindi language. This includes exploring methods for Hindi text preprocessing, tokenization, part-of-speech tagging, Named Entity Recognition (NER), and language understanding. By leveraging existing NLP libraries and tools, and adapting them to Hindi, we aim to enhance the chatbot's linguistic capabilities and improve its ability to interpret user queries accurately.

2. Rasa Library:

As a pivotal component of our chatbot development, we will integrate the Rasa library, an open-source conversational AI framework, into our system architecture. The Rasa library provides a comprehensive suite of tools and functionalities for building AI-powered chatbots, including natural language understanding, dialogue management, and integration capabilities. By leveraging

Rasa, we aim to streamline the development process and empower our chatbot with advanced conversational capabilities.

3. CRF (Conditional Random Fields):

Conditional Random Fields (CRF) is a statistical modeling technique used for structured prediction tasks, particularly in natural language processing and sequence labeling tasks. In the context of our

Chatbot development, we plan to utilize CRF as part of the Named Entity Recognition (NER) process.

CRF is a type of probabilistic graphical model that models the conditional probability distribution of a sequence of labels given a sequence of observations. In simpler terms, it helps in predicting the most likely sequence of labels (such as named entities) for a given sequence of input tokens (words or characters).

In the case of NER, CRF algorithms are trained on labeled datasets where each token is in a response ranking, leading to improved user satisfaction. A summarized text sequence is associated with a label indicating its entity type (such as person names, locations, organizations, etc.).

4. Ranking Algorithm:

In our research, we recognize the pivotal role of ranking algorithms within chatbot systems, particularly in determining the relevance and order of responses to user queries. Ranking algorithms serve as critical components, influencing the overall user experience by ensuring that the most pertinent and helpful responses are presented promptly. Within this context, our investigation delves into various types of ranking algorithms, each offering distinct methodologies and approaches tailored to the specific needs and characteristics of chatbot applications.

Among the most common ranking algorithms are TF-IDF (Term Frequency-Inverse Document

Frequency) and its extension, BM25 (Best Matching 25), both leverage statistical measures to assess the importance of terms within documents or responses relative to a corpus. These algorithms are effective for ranking responses based on the frequency and relevance of keywords or terms in user queries and responses.

Learning to Rank (LTR) algorithms, a supervised machine learning approach, stand out for their ability to optimize ranking functions based on labeled training data or relevance judgments. By learning from human annotators or implicit feedback signals, LTR algorithms enhance the precision and accuracy. However, these methods cannot generate new sentences on their own. Therefore, the template-based algorithm was introduced, where a specific template needs to be predefined for a particular summarization task.

5. Trigrams 'n' Tags (TnT) Tagging

In our research paper, we explore the utilization of Trigrams 'n' Tags (TnT) tagging as a vital component in the development of a sophisticated chatbot tailored for hotel food ordering. TnT tagging represents a statistical approach to part-of-speech tagging, which plays a pivotal role in understanding the grammatical structure and semantic context of user queries and responses. TnT tagging operates on the principle of trigrams, wherein it considers the probability of a tag (part-of-speech label) occurring given the previous two tags in a sequence. By analyzing patterns and sequences of words within sentences, TnT tagging assigns the most probable part-of-speech labels to each word, thereby facilitating accurate and contextually relevant analysis of user input.

III. SYSTEM ARCHITECTURE

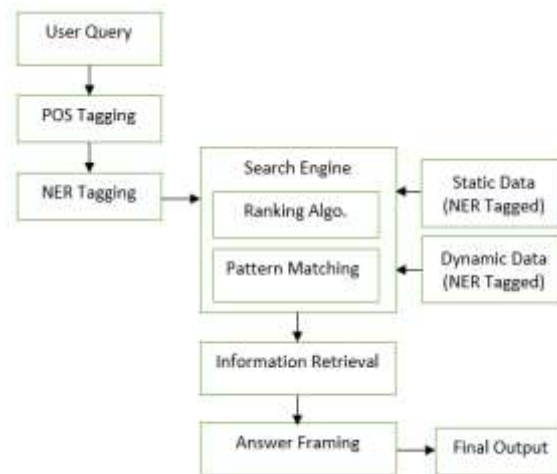


Figure 1: system architecture

IV. METHODOLOGY

Chat: The chat is created using a pattern that is known to the user and could be easy to understand. Chat dialog boxes show up to create conversation.

POS Tagging: It is the procedure of identifying correct tags like nouns, adjectives, verbs, adverbs, etc. to each word of the input sentence.

NER Tagging: Named entity recognition is the procedure of distinguishing named entities that are present in the database. NER is 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 the 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 performance and relevance. saved in the database and consistent response is given back.

Intent Recognition: Intent recognition, also known as intent detection is a technology used in natural language processing (NLP) that helps computers understand the purpose or goal behind a user's input.

V. CHALLENGES

1. Language Understanding: Hindi has a rich morphology and syntax, which can make natural language understanding (NLU) more challenging compared to languages with simpler structures. Building accurate language models for intent recognition and entity extraction requires extensive training data and linguistic expertise.

2. Data Availability: There may be limited publicly available data in Hindi for training machine learning models, including pre-trained word embeddings and language models. This can make it challenging to achieve high performance in NLU and dialogue management tasks.

3. Tokenization and Word Segmentation: Hindi text does not have explicit word boundaries like English, which makes tokenization and word segmentation more complex. Proper tokenization is crucial for accurate language processing, including part-of-speech tagging and named entity recognition.

4. Spelling Variations: Hindi words can have multiple spellings and variations due to transliteration and phonetic differences. Handling spelling variations and normalization is important for improving the robustness of the chatbot's language understanding capabilities.

5. Cultural Sensitivity: Building a chatbot for Hindi speakers requires consideration of cultural nuances, conventions, and sensitivities specific to the Hindi-speaking audience. Adapting the chatbot's responses and behavior to align with cultural norms is essential for building trust and engagement.

6. Evaluation and Testing: Evaluating the performance of a Hindi chatbot requires appropriate evaluation metrics and test data in Hindi. Developing comprehensive test suites and conducting user testing with native Hindi speakers are important steps in ensuring the effectiveness and usability of the chatbot.

7. Domain-specific Knowledge: Depending on the application domain, acquiring and integrating domain-specific knowledge and terminology in Hindi can be challenging. Building domain-specific language models and knowledge bases tailored to the target domain is essential for enhancing the chatbot's

8. Integration with External Systems: Integrating the chatbot with external systems, APIs, and services may require support for Hindi language input and output formats. Ensuring compatibility and seamless interaction with external systems in Hindi is crucial for providing accurate and relevant responses to user queries.

VI. CONCLUSION

This paper enlightens all the issues, challenges, and issues faced by the system to provide efficient help for students, colleges, teachers, and all other sub-parts. The advanced chatbot system in the Hindi language is used for a better user interface for users to get all data and answers in the Hindi language. The system makes a simple user interface so they easily interact with the system. All the schedule data is available on a single portal.

VII. FUTURE SCOPE

The proposed chatbot can answer only textually typed questions which can also be implemented for voice-based question answering systems and can also be implemented for other regional languages.

REFERENCES

- [1] Prasenjeet Sinha, Prafulla B. Bafna, Jatinder Kumar R. Saini, "Hindi Speech-Based Healthcare Chatbot", SN - 978-981-16-9704-3 Hindi Speech-Based Healthcare Chatbot, (2022).
- [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.Arfa, 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] A.M Rahman, Alma Islam, "Programming Challenges of Chatbot.", 2017 IEEE Region 10 Humanitarian Technology Conference, (21-23 Dec 2017).
- [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] Patrick 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).