AI ENABLED REGIONAL LANGUAGE SEARCH ENGINE

Jayanthi N [1], Madhavi G[2] , Manisha kudchi[3] , Manisha Gajre[4], Mrs.Priyanka Chavan[5]

[2345]Student , Department   of Computer Science and Engineering , AMCEC , Bangalore-083,Karnataka

[1]Faculty , Department   of Computer Science and Engineering , AMCEC , Bangalore-083,Karnataka

Abstract

AI-enabled regional language search engines have gained significant importance in recent years due to the growing need for accessing information in local languages. These search engines leverage artificial intelligence techniques to enable users to search for and retrieve content in their regional languages. This abstract presents an overview of an AI-enabled regional language search engine and its key components. The objective of the AI-enabled regional language search engine is to bridge the language barrier and provide users with the ability to search for information in their preferred regional language. The search engine utilizes natural language processing (NLP) techniques to understand and process queries in regional languages. It employs machine learning algorithms to analyze and index regional language content, ensuring accurate and relevant search results. The proposed solution involves the development of a comprehensive search engine infrastructure that includes data collection, preprocessing, indexing, and retrieval mechanisms. It employs language-specific models and algorithms to handle the unique characteristics and complexities of regional languages. The search engine integrates advanced NLP techniques, such as part-of-speech tagging, entity recognition, and sentiment analysis, to enhance the search experience and provide more nuanced results.The architecture design of the AI-enabled regional language search engine includes components such as web crawling, indexing, query processing, and user interface. It leverages scalable and distributed computing frameworks to handle large volumes of regional language content and user queries efficiently. The search engine also incorporates user feedback mechanisms to continuously improve search results and user satisfaction.

Keywords: AI, regional language, search engine, natural language processing, machine learning, indexing, user interface

I. INTRODUCTION

Regional languages, also known as vernacular languages, are languages spoken in specific geographical regions or by particular communities within a country. These languages are distinct from official or national languages and often have deep cultural and historical significance.

THE REGIONAL OF KARNATAKA: KANNADA

Kannada is a Dravidian language primarily spoken by the people of Karnataka, a state in southern India. It is one of the oldest languages in India and has a rich literary tradition that dates back over a thousand years.

Here is an overview of Kannada:

1. Geographic Distribution: Kannada is the official language of Karnataka and is spoken by the majority of the state's population. It is also spoken by significant communities in neighboring states like Andhra Pradesh, Telangana, Tamil Nadu, and Maharashtra. There are also Kannada-speaking diaspora communities around the world.

2. Script: Kannada has its own script, known as Kannada script, which is derived from the ancient Brahmi script. It is a syllabic alphabet that consists of 49 characters, including vowels and consonants. The script is written from left to right.

3. Literature and Cultural Heritage: Kannada has a rich literary tradition that dates back to the 9th century. It has produced numerous renowned poets, writers, and playwrights. The earliest known Kannada literary work is the "Kavirajamarga," written by the Rashtrakuta king, Nripatunga Amoghavarsha. The works of celebrated Kannada poets like Pampa, Ranna, and Kuvempu have made significant contributions to Indian literature. The Vachana literary movement emerged in the 12th century as a protest against social inequality and religious orthodoxy. Vachanas are devotional poems composed by saints known as "Vachanakaras." These poems, written in a simple and direct style, express deep spiritual insights and advocate social equality. The Vachana literature is considered a unique contribution to Kannada literature.etc………

II. OBJECTIVE

The main objectives of this project are:

- ✓ To develop a user-friendly web application that provides AI-powered Q&A in the Kannada language.
- ✓ To support both text and audio input from users to enhance accessibility and user experience.
- ✓ To create a seamless communication experience between the user and the AI by leveraging real-time web technologies.
- ✓ This aims to explain the significance, features, and potential benefits of such a search engine for regional languages.
- ✓ It also aims to outline the technical components and considerations involved in developing and implementing an AI-enabled regional language search engine.
- ✓ This will highlight the importance of digital inclusion, linguistic diversity, and cultural preservation in promoting regional languages through advanced technology.
- ✓ Additionally, this will discuss the potential impact of an AI-enabled regional language search engine on users, content creators, businesses, and society as a whole

III. PROPOSED SYSTEM

This project aims to develop a web application that provides AI-powered Q&A assistance in Kannada, making information more accessible to speakers of the language.

Comprehensive Language Support: The search engine should prioritize regional language support, ensuring a wide range of regional language is included. This involves building language models and datasets specify the regional language, enabling accurate understanding and processing of search queries in the form of text and audio.

Rich Regional Language Content Indexing: Create a comprehensive index of regional language content available on the internet. Implement web crawling techniques that specifically target regional language websites, blogs, forums, and social media platforms. This will ensure a wide range of content is indexed and made accessible through the search engine.

IV. SYSREM ARCHITECTURE

The web application's architecture consists of three main components:

Front-end: The user interface (UI) is designed using HTML, CSS, and JavaScript to provide an intuitive and accessible experience. It includes options for text and audio input, as well as a display for the conversation list.

Back-end: The back-end is built using Node.js and Express.js, which handle user input and communication with the AI. Web sockets are used for real-time communication between the UI and back-end, enabling a seamless user experience.

API integrations: The application relies on three main APIs to function. The Google Language Translation API is used for translating text between Kannada and English, while the ChatGPT API is responsible for generating answers to users' questions. Finally, the Google Text-to-Speech API converts Kannada text to audio for playback in the web audio player.
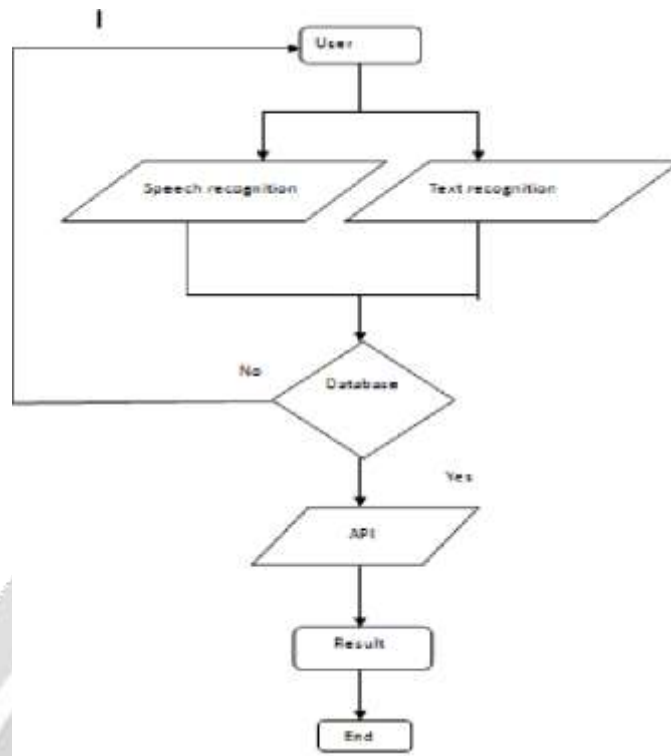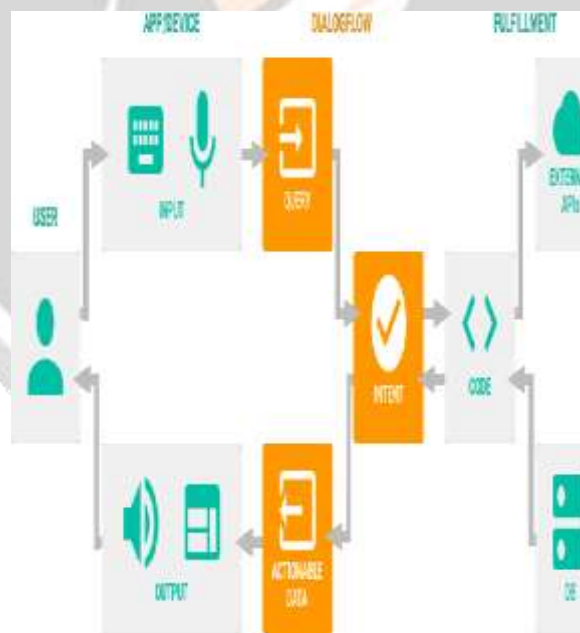
FIG:1 SYSTEM ARCHITECTURE

FIG: 2 WORKFLOW OF TARGET CREATION

V. CONCLUSION

This project aimed to develop a web application that provides AI-powered Q&A assistance in the Kannada language. The application supports both text and audio input, improving accessibility and user experience. A combination of tools and technologies, such as Node.js, Express.js, and various APIs, were utilized to create a seamless communication experience between users and the AI.

VI. REFERENCES

Here are some references on AI regional language search engines:

1. Sarkar, A., et al. "Language Technology for Developing World Languages." Language Resources and Evaluation, vol. 47, no. 1, 2013, pp. 121-147.

2. Ganesh, S., et al. "Building Low-resource Machine Translation Systems for Indian Languages." Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 2015, pp. 171-176.

3. Kavya, S., and Ramakanth, P. "Development of Kannada Stemmer for Information Retrieval." International Journal of Computer Applications, vol. 139, no. 11, 2016, pp. 22-25.

4. Garg, R., et al. "Automatic Text Classification for Regional Language Documents." 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 869-874.

5. Rajesh, R., and Balasubramanian, R. "Design and Implementation of a Tamil Search Engine." International Journal of Computer Science Issues, vol. 9, no. 1, 2012, pp. 165-171.

6. Kumar, T. S., and Ramalingam, V. "Multilingual Information Retrieval for Indian Languages." International Journal of Applied Engineering Research, vol. 13, no. 9, 2018, pp. 7065-7072.

7. Sivakumar, P., and Sujatha, S. "Integrating WordNet and Wikipedia for Information Retrieval in Indian Languages." Proceedings of the 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2015, pp. 1-6.