

LINGO: A MULTI-LANGUAGE TRANSLATING APP

Project Reference No: 49S_BE_3426

College : BGS College of Engineering and Technology, Bengaluru
Branch : Department of Artificial Intelligence and Machine Learning
Guide(s) : Dr. Ravikumar. G. K
Dr. Madhura Gangaiah
Student(s) : Ms. Rhea Fatima Estibeiro
Ms. Sinchan Dechakka M K
Mr. Ansh Ajit Salunke
Mr. Aditya Suresh

Keywords:

Multilingual Translation, Real-Time Chat System, Slang Normalization, Code-Mixed Language Processing, Transformer Models, NLP, Context-Aware Translation.

Introduction:

In today's globally connected digital world, communication across different languages remains a significant challenge. Existing translation systems often fail to accurately interpret informal language, slang, emojis, and code-mixed text commonly used in real-time conversations. This leads to misunderstandings and ineffective communication.

To address these challenges, the proposed project "Lingo: A Multi-Language Translating App" aims to develop a real-time multilingual chat system that enables seamless communication between users speaking different languages. The system integrates advanced Natural Language Processing (NLP) techniques and transformer-based models such as mBART or T5 to provide context-aware and accurate translations.

Additionally, the project incorporates a preprocessing pipeline to normalize slang, remove noise, and handle mixed-language inputs, ensuring improved translation quality. The application is designed to work efficiently even in low-network conditions, making it practical for real-world usage.

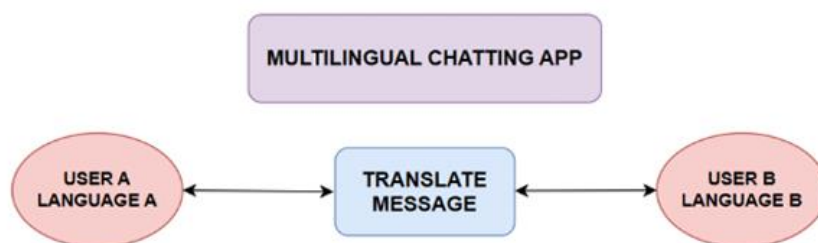


Figure 1: Workflow for Multilingual Chatting Application

Objectives:

1. To develop a real-time multilingual chatting system for seamless communication across languages.
2. To implement automatic language detection and personalized translation.
3. To improve translation accuracy using slang normalization and preprocessing techniques.

4. To support code-mixed and informal conversations with context-aware translation.
5. To ensure secure, low-latency communication using efficient APIs and messaging systems.

Methodology:

The project follows an Agile-based modular development approach:

1. Requirement analysis to understand multilingual communication challenges.
2. Design of system architecture including preprocessing, translation, and messaging layers.
3. Development of a preprocessing pipeline for noise removal, slang normalization, and code-mix detection.
4. Integration of transformer-based models such as mBART/T5 or Sarvam model for translation.
5. Implementation of real-time messaging using WebSockets and APIs.
6. Continuous testing and optimization based on accuracy, latency, and user experience.

This approach ensures scalability, flexibility, and efficient system performance.

Result and Conclusion:

The proposed system successfully enables real-time multilingual communication with improved accuracy compared to traditional translation systems. By incorporating slang normalization and context-aware processing, the system effectively handles informal and code-mixed conversations.

The application provides a robust, scalable, and user-friendly solution for overcoming language barriers, making it suitable for real-world applications such as education, customer support, and global collaboration.

In conclusion, the project demonstrates how AI-driven translation systems can transform communication into a more inclusive and seamless experience.

References:

1. Vaswani et al., "Attention is All You Need," in Proceedings of the IEEE Conference on Neural Information Processing Systems (NeurIPS), 2017.
2. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of NAACL-HLT, 2019.
3. Y. Liu et al., "Multilingual Denoising Pre-training for Neural Machine Translation (mBART)," in Transactions of the Association for Computational Linguistics, vol. 8, pp. 726–742, 2020.
4. S. Ramesh et al., "IndicTrans: Multilingual Neural Machine Translation for Indian Languages," in Proceedings of ACL Workshop on Asian Translation, 2021.

Future Scope:

The future scope of this project includes:

1. Integration of voice-based multilingual communication.
2. Emotion-aware translation using sentiment analysis.
3. Expansion to support more regional and low-resource languages.
4. Integration with enterprise systems such as customer support and social platforms.