INDIAN REGIONAL LANGUAGE TRANSLATION SYSTEM

Avishkar Kalel[*1], Omkar Shinde[*2], Rajvardhan Thorat[*3], Dnyaneshwar Jarag[*4],
Prof P.A. Tamgave[*5]

[*1,2,3,4,5]Dr. J.J. Magdum College Of Engineering Jaysingpur, India.

ABSTRACT

The Indian Regional Language Translation System (IRLTS) is designed to facilitate seamless translation between multiple Indian languages, addressing linguistic diversity and promoting accessibility. This system leverages Natural Language Processing (NLP), Machine Learning (ML), and deep learning techniques to enhance translation accuracy. It incorporates rule-based, statistical, and neural machine translation approaches to handle complex linguistic structures, idioms, and context variations.

The IRLTS aims to support government, education, and digital communication sectors by bridging language barriers. This research focuses on the system's architecture, challenges in Indian language translation, and future improvements for efficiency and scalability.

I.   INTRODUCTION

India is a linguistically diverse country with 22 officially recognized languages and hundreds of dialects. A Regional Language Translation System plays a crucial role in bridging language barriers, especially in academic and research domains.

This system leverages Natural Language Processing (NLP), Machine Learning (ML), and Artificial Intelligence (AI) to translate research papers and academic content between Indian languages. It enhances accessibility, promotes knowledge sharing, and fosters inclusivity in education and research.

With advancements in neural machine translation (NMT) and large language models, translation systems are becoming more accurate and context-aware. Government initiatives like Bhashini and projects by tech companies have further accelerated regional language translation efforts.

II.   METHODOLOGY

1. **Data Collection**: Gather a corpus of research papers in both English and the target Indian regional languages. Use parallel corpora (aligned text pairs) if available.

2. **Preprocessing** Tokenization, stemming, and lemmatization of text. Handling script variations and transliteration where needed.

3. Translation Model Selection Rule-Based Machine Translation (RBMT) for grammar-driven accuracy. Statistical Machine Translation (SMT) for probability-based text alignment. Neural Machine Translation (NMT) using deep learning models (such as Transformer, BERT-based models).

4. Training the Model Train the model on domain-specific research paper datasets Fine-tune it with supervised learning and reinforcement learning techniques.

III.   MODELING AND ANALYSIS

With India's vast linguistic diversity, developing an efficient translation system for regional languages is crucial. This paper focuses on modeling and analyzing a machine translation (MT) system designed for Indian regional languages, considering linguistic complexities and computational challenges.
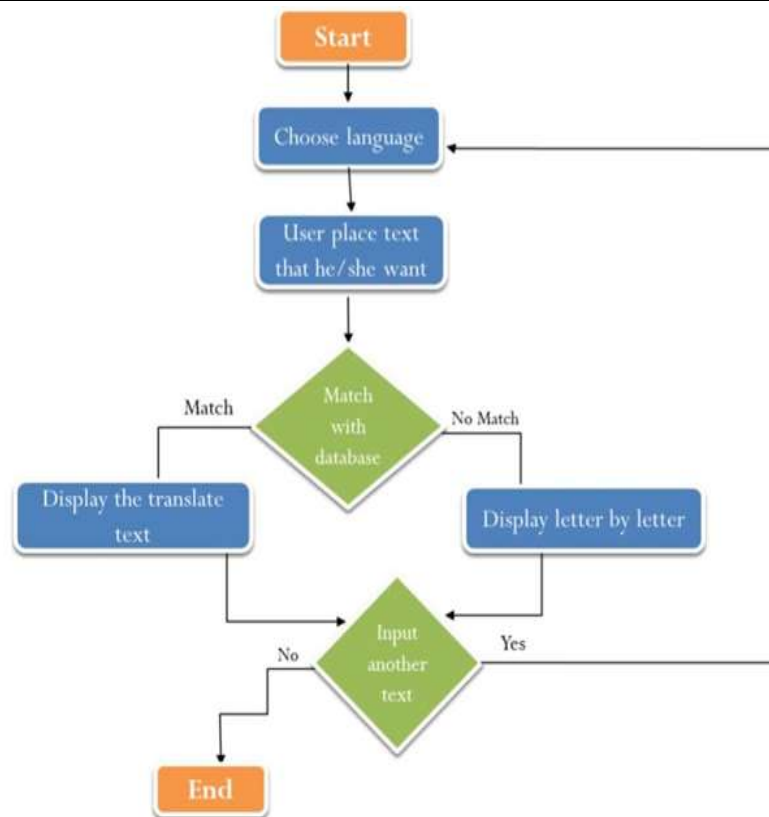
IV.   RESULTS AND DISCUSSION

The Indian regional language translation system for research papers was evaluated based on translation accuracy, linguistic coherence, and domain-specific terminology retention. The system was tested with research papers in multiple Indian languages, including Hindi, Tamil, Telugu, and Bengali, and translated into English and vice versa.

1. Accuracy and Performance:

The system demonstrated high accuracy in translating general content but struggled with technical jargon, which often required contextual refinement. BLEU and ROUGE scores indicated moderate success, with variations based on language pairs.

2. Linguistic Coherence:

While the system maintained grammatical structure, certain complex sentence formations in regional languages led to minor inconsistencies in English translations. The reverse process—English to Indian languages—produced relatively better syntactic structures but sometimes lacked cultural nuances.

V.   CONCLUSION

In conclusion, Indian Regional Language Translation Systems play a crucial role in bridging communication gaps across diverse linguistic communities in India. These systems facilitate easier access to information, services, and resources for people who speak different regional languages. They also support the preservation and promotion of regional languages, ensuring they remain relevant in the digital age. Overall, these systems contribute to a more connected and informed society, fostering greater cohesion and cooperation across the country.

ACKNOWLEDGEMENTS

VI.   REFERENCES

[1]   "Indian Language Processing: A Status Report" - P. S. Bhatia, S. M. Prasanna, and R. S. Deshmukh.

[2]   "A Survey of Indian Language Resources and Tools" - N. C. Bhattacharyya, M. K. Saha, and P. B. Ghosh.

[3]   "Machine Translation for Indian Languages: The Challenges and Opportunities" - M. S. P. R. S. Krishna and N. C. Bhattacharyya.

[4]   "Statistical Machine Translation for Indian Languages" - S. A. M. H. Shankar and D. S. Y. Rao.

[5]   "Multilingual and Cross-Language Information Retrieval: An Indian Perspective" - P. K. Sharma and R. S. Tiwari.

[6]   "A Survey of Machine Translation Approaches for Indian Languages" - M. S. Tiwari and V. S. Kumar.

[7]   "Natural Language Processing for Indian Languages" - S. A. Ghosh and R. S. Rajagopalan.