

A Software That Automatically Translate Resource Materials between English to Indian Regional Languages

Mustafa Masuldar

Computer Science and Engineering
N B Navale Sinhgad College of
Engineering, Solapur

Sakshi Deshmukh

Computer Science and Engineering
N B Navale Sinhgad College of
Engineering, Solapur

Nandini Madre

Computer Science and Engineering
N B Navale Sinhgad College of
Engineering, Solapur

Pooja Pawar

Computer Science and Engineering
N B Navale Sinhgad College of
Engineering, Solapur

ABSTRACT

CIPAM, dedicated to advancing Intellectual Property Rights (IPR) awareness, commercialization, and enforcement, has created extensive educational materials. The software aims to bridge linguistic gaps by translating these materials from English to key Indian regional languages: Hindi, Marathi, Bengali, Gujarati, Tamil, and Telugu.

- The software's core features include its ability to translate various formats, such as Word documents, PDFs, and text within images. Notably, it focuses on maintaining the contextual integrity and
- Professionalism of the original content, ensuring that translations are not only accurate but also accessible to the general public.
- This abstract outlines a comprehensive approach involving Natural Language Processing (NLP) techniques, machine learning algorithms, and collaboration with experts to ensure the software's accuracy and user-friendliness.

Keywords: Research Paper, Technical Writing, Translation Software, Natural Language Processing (NLP), Engineering and Technology

I. INTRODUCTION

In today's globalized world, protecting Intellectual Property Rights (IPR) is crucial for driving innovation, economic growth, and sustainable development. Organizations like the Cell for IPR Promotion and Management (CIPAM) in India understand the importance of raising awareness about IPR and making it accessible to everyone. To tackle this challenge, CIPAM has developed advanced software that can translate IPR educational materials from English into major Indian languages like Hindi, Marathi, Bengali, Gujarati, Tamil, and Telugu.

This software is designed to handle various types of content, including text-based documents, PDFs, and even text within images. But what sets it apart is its ability to maintain the original meaning and professionalism of the content it translates. This ensures that the translated materials are not only

accurate but also easy to understand for a wider audience, thus empowering more people with essential knowledge about IPR.

In this research paper, we'll take a closer look at CIPAM's multilingual translation software. We'll explore how it works, including the technical aspects like Natural Language Processing (NLP) techniques and machine learning algorithms. Additionally, we'll delve into real-world examples to understand how this software is making a difference in spreading awareness about IPR.

By examining the impact of CIPAM's innovative approach to translation, we hope to highlight the role of technology in promoting IPR awareness across diverse linguistic landscapes. Furthermore, we'll provide practical recommendations based on feedback

from users and stakeholders, aiming to improve similar initiatives and further advance the cause of IPR promotion in the digital age.

II. LITERATURE REVIEW

Adapted from the paper “*Jayanthi, N., Lakshmi, A., Raju, C. S. K., & Swathi, B. (2020). Dual Translation of International and Indian Regional Language using Recent Machine Translation. 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*”

Title: [Dual Translation of International and Indian Regional Language using Recent Machine Translation](#)

Neural Machine Translation (NMT) is a modern and powerful approach that has resulted in major improvements compared to traditional techniques of machine translation for translating one language into another. In the world, India is a very multicultural and multilingual country. People of India from various regions use their own regional communication languages, making India stand at the second position in the world to have maximum languages. In India, English is provided as the second extra official language. But the usage of English in India is very less forming a communication gap. To minimize this gap by translating one language into another language is almost impossible for humans. It can be achieved by a machine translation. This paper focuses on translating Indic languages using the translation technique of Neural technology. A Sequence to Sequence model with encoder-decoder attention mechanism of neural machine translation is proposed for Telugu language conversion into English and vice versa.

Results of the paper “*Nadeem Jadoon Khan, Waqas Anwar, Nadir Durrani, Machine Translation Approaches and Survey for Indian Languages*”

Title: [Machine Translation Approaches and Survey for Indian Languages](#)

In this study, we present an analysis regarding the performance of the state-of-art Phrase based Statistical Machine Translation (SMT) on multiple Indian languages. We report baseline

systems on several language pairs. The motivation of this study is to promote the development of SMT and linguistic resources for these language pairs, as the current state-of-the-art is quite bleak due to sparse data resources. The success of an SMT system is contingent on the availability of a large parallel corpus. Such data is necessary to reliably estimate translation probabilities. We report the performance of baseline systems translating from Indian languages (Bengali, Gujarati, Hindi, Malayalam, Punjabi, Tamil, Telugu and Urdu) into English with average 10% accurate results for all the language pairs.

From this paper “*Pulipaka, S. K., Kasaraneni, C. K., Sandeep Vemulapalli, V. N., & Mourya Kosaraju, S. S. (2019). Machine Translation of English Videos to Indian Regional Languages using Open Innovation. 2019 IEEE International Symposium on Technology and Society (ISTAS).*”

Title: [Machine Translation of English Videos to Indian Regional Languages using Open Innovation](#)

In spite of many languages being spoken in India, it is difficult for the people to understand foreign languages like English, Spanish, Italian, etc. The recognition and synthesis of speech are prominent emerging technologies in natural language processing and communication domains. This paper aims to leverage the open source applications of these technologies, machine translation, text-to-speech system (TTS), and speech-to text system (STT) to convert available online resources to Indian languages. This application takes an English language video as an input and separates the audio from video. It then divides the audio file into several smaller chunks based on the timestamps. These audio chunks are then individually converted into text using IBM Watson’s speech-to-text (STT) module

From this paper “*RAJANI S, Translation across Cultures: From the Regional to the Universal*”

Title: [Translation Across Cultures: From the Regional to the Universal](#)

India is a nation of many regional languages. The advent of the Europeans brought in a necessity to understand the regional languages through a