International Journal of Research Publication and Reviews

IPR Languages Translation

Jivesh Kumar[1] ,Vishal Kumar[2], Dipak Lungba[3], Abba AM John Justin[4] , Dr. K Geetha[5]

Students, Professsor Department of Computer Science and Engineering,
Excel Engineering College, Komarapalayam, Tamil Nadu, India

ABSTRACT :

IPR Language Translation is an innovative web-based application aimed at bridging language barriers in promoting intellectual property rights (IPR) awareness, commercialization, and enforcement in India. Developed by the Cell for IPR Promotion and Management (CIPAM), the software translates IPR resource materials from English into multiple Indian regional languages, including Hindi, Marathi, Bengali, Gujarati, Tamil, and Telugu. Prioritizing clarity and accuracy, the software ensures complex legal and technical concepts are presented in an easily understandable manner. The methodology involves dataset collection, machine learning-based translation engine creation, user interface design, content analysis, translation quality assurance, simple text generation, performance optimization, and deployment with user support. This transformative solution aims to empower a broader audience, expanding the reach and impact of CIPAM's IPR education initiatives in India.

Key Words: Translation, Datasets, Recognition, Bilingual

INTRODUCTION

In the endeavor to advance awareness, commercialization, and enforcement of intellectual property rights (IPR) in India, the Cell for IPR Promotion and Management (CIPAM) acknowledges the pressing necessity to make its educational materials accessible to a diverse audience. Overcoming language barriers has consistently presented a formidable obstacle to this objective. Despite CIPAM's pioneering efforts in IPR advocacy and the development of a plethora of educational resources, their effectiveness is hampered by their confinement to the English language. To confront this challenge, CIPAM envisions the creation of an innovative software solution capable of seamlessly translating these resources into multiple Indian regional languages. This project aims to empower individuals throughout India by breaking down language barriers and providing access to IPR knowledge in languages such as Hindi, Marathi, Bengali, Gujarati, Tamil, and Telugu. The software will prioritize clarity and accuracy, ensuring that even intricate legal and technical concepts are communicated in a manner easily understandable to the general populace. This introduction lays the groundwork for a transformative solution poised to broaden the reach and influence of CIPAM's IPR education initiatives.

LITERATURE SURVEY

1. In the cited literature [1], the author predominantly discusses the architecture of a multi-engine web machine translation system, emphasizing engine allocation, development advantages, thesaurus upkeep, and service realization. Specifically, attention is given to a unique verb-driven algorithm and methodologies to enhance grammatical analysis efficiency. Through evaluation using automated metrics, the fusion of sci-tech journal translations surpasses single-rule and statistical systems, showcasing the effectiveness of the fusion approach in enhancing quality and performance for science-related journal translations.

2. In the referenced paper [2], the author introduces a Human-Computer Interaction Algorithm and designs an English machine translation system aimed at reducing translators' cognitive load and time consumption. By alleviating the need for exhaustive translation recognition, the system enhances translation efficiency.

3. In the literature [3], the implementation of machine translation for English to Bangla language processing and vice versa is proposed. Various approaches such as word-to-word, direct, transfer, interlingual, corpus-based, and statistical translation are explored for efficient natural language processing. The paper also focuses on performance analysis of different translation models.

4. In the scholarly work [4], the core concepts of natural language processing (NLP) and its global research progress are presented. The applications of NLP in machine translation are analyzed, along with identified challenges and emerging trends in the field.

5. Lastly, in the discussed study [5], a model for Marathi to Gujarati text translation using Text-to-Text Translation is outlined. Leveraging the adaptability of LSTMs and deep learning principles, the proposed system integrates multilingual features to develop a translation system with deep learning at its core, aiming for improved text translation outcomes.

PROBLEM STATEMENT

In its fervent promotion of IPR enforcement, commercialization, and awareness, CIPAM has diligently crafted an array of instructional resources and tools tailored for comprehension by a diverse audience including the public, businesses, law enforcement, courts, and customs officials. To cater to a wider demographic, there arises a crucial need for software capable of translating these materials from English into various regional Indian languages. This software should adeptly handle different formats such as Word documents, PDFs, and text within images. It is imperative that the software not only accurately translates the text but also captures its intended meaning. Translation options should include Hindi, Marathi, Bengali, Gujarati, Tamil, and Telugu. Furthermore, the translated content should maintain a formal tone, remain devoid of colloquial expressions, be presented in plain language, and be easily digestible for a general audience.

SOFTWARE REQUIREMENTS

i. *Opus MT Datasets:* Opus MT serves as a vital resource for machine translation research, offering parallel text corpora in various languages. These datasets enable training and testing of machine translation models across diverse domains and language pairs.

ii. *Hugging Face Transformers* **:** A popular NLP library, Hugging Face Transformers provides access to a wide array of pre-trained models for tasks like sentiment analysis and language translation. Its user-friendly interface and collaborative platform make it indispensable for language-related research and development.

iii. *Tensor Flow :* Google's Tensor Flow framework is essential for deep learning research and applications, offering flexibility and support for building and training intricate neural networks efficiently on both CPU and GPU.

iv. *OCR Tool :* Optical Character Recognition tools utilize pattern recognition algorithms to convert text from images and scanned documents into machine-readable formats, facilitating data extraction and document management tasks.

v. *Translation Evaluation Metrics :* BLEU metric assesses machine translation quality by comparing output to human translations, primarily focusing on n-gram precision. While widely used, it has limitations in capturing translation nuances like semantic accuracy.

vi. *Visual Studio :* Microsoft's Visual Studio IDE enables comprehensive software development, supporting various platforms and offering features like code completion, debugging, and GUI design tools, along with plugin compatibility for added functionality.

vii. *Web Browsers :* Software applications facilitating web exploration and interaction, web browsers provide user-friendly interfaces, rendering engines for displaying web content, security features, and support for extensions. Notable examples include Google Chrome, Mozilla Firefox, Microsoft Edge, and Apple Safari.
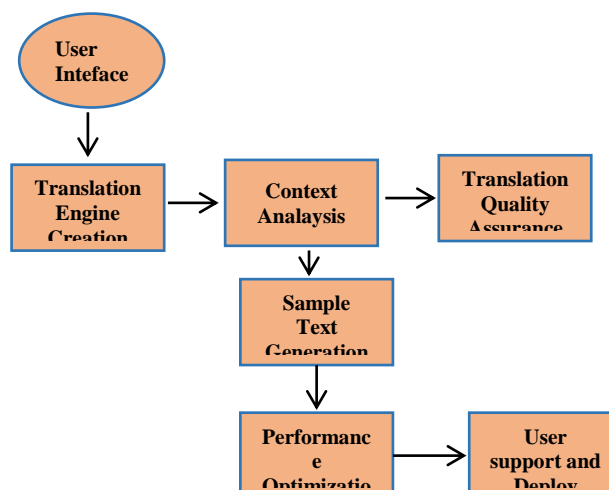
FLOW DIAGRAM

Fig 1: flow Daigram

METHODOLOGY

i. Data Collection:

Acquiring a diverse dataset is essential, comprising texts in English and their translations into target Indian regional languages. This dataset forms the foundation for our translation system, encompassing various resources and materials relevant to user preferences.

ii. Development of Translation Engine:

The translation engine, employing machine learning and NLP techniques, is the system's backbone. Training this engine with bilingual data enhances translation accuracy, ensuring proficiency in rendering text across languages.

iii. Interface Design:

Crafting an intuitive interface using HTML, CSS, and JavaScript facilitates user interaction. Users can input text or documents and select their preferred language for translation, with customization options tailored to individual preferences.

iv. Content Processing:

Utilizing algorithms and tools like OCR, document parsing libraries, language identification, and named entity recognition ensures versatility in handling diverse source formats. This adaptability is crucial for effective translation across various file types.

v. Quality Assurance:

Implementing a rigorous quality assurance process guarantees accurate and high-quality translations. Automated and manual assessments validate the translated content, ensuring adherence to desired standards.

vi. Text Validation:

Generating sample sentences in the target language enables validation of translation quality. Comparative analysis against human-made translations aids in refining and optimizing the software for enhanced accuracy.

vii. Performance Enhancement:

Optimizing software performance focuses on efficiency, speed, and resource utilization. Code optimizations, algorithm enhancements, and streamlined data processing mitigate computational overhead, ensuring seamless translation even under high request volumes.

viii. Deployment and Support:

Deploying the software on web servers for accessibility via popular browsers entails scalability measures for handling concurrent requests efficiently. Robust security protocols safeguard user data, while comprehensive documentation and user support resources facilitate seamless user experience and continual improvement based on feedback.

RESULT AND DISCUSSION

Language Selection

Fig 2: User Interface
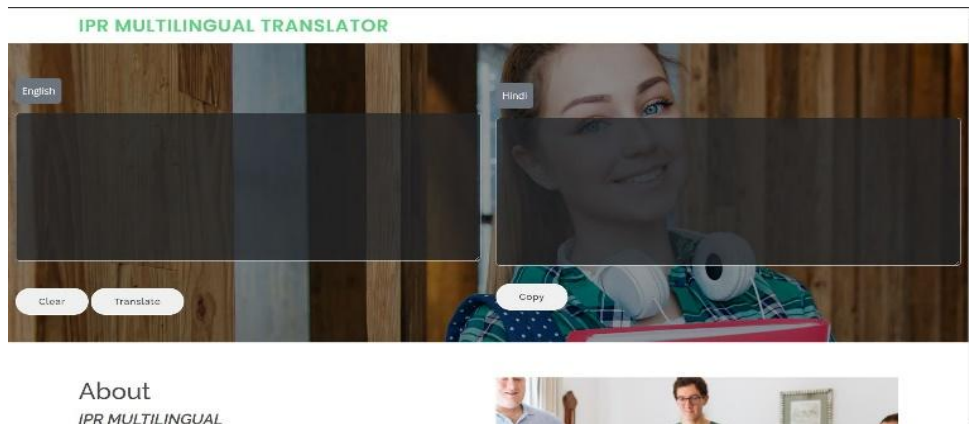
Text to Text Translation

Fig 3: English text translate to Hindi

Fig 4: English text to Tamil

CONCLUSION

The IPR Language Translation initiative marks a substantial advancement in tackling the language obstacles that have long impeded the advancement

of intellectual property rights (IPR) awareness, commercialization, and enforcement in India. Recognizing the urgent necessity to make educational materials accessible to a diverse array of stakeholders, including students, industries, law enforcement agencies, the judiciary, customs officials, and the general populace, has been identified as a pivotal priority.

This endeavor aims to bridge this divide by creating a state-of-the-art web-based application software adept at translating IPR resources from English into numerous Indian regional languages such as Hindi, Marathi, Bengali, Gujarati, Tamil, and Telugu. The software places significant emphasis on ensuring clarity and formal precision, ensuring that even intricate legal and technical concepts are presented in an easily understandable manner.

The approach adopted encompasses a wide array of activities including dataset procurement, the creation of a machine learning-driven translation engine, designing user-friendly interfaces, conducting content analysis, assuring translation quality, generating sample texts for verification, optimizing performance, and deploying the solution with comprehensive user assistance. This groundbreaking initiative not only seeks to surmount language barriers but also aims to empower a broader audience by expanding the outreach and influence of CIPAM's IPR education endeavors across India.

Utilizing advanced NLP methodologies, machine learning techniques, and incorporating versatile tools such as Optical Character Recognition (OCR) and translation evaluation metrics, alongside leveraging crucial resources like Opus MT datasets and NLP libraries, underscores the commitment of the project to deliver accurate translations of superior quality in an accessible manner.

This project represents not just a technological innovation but a significant stride toward democratizing knowledge and championing intellectual property rights on a wider scale. The overarching objective is to cultivate a more knowledgeable and empowered society capable of fully realizing the potential of IPR in India.

References

1. Aarati H. Patil, Snehal S. Patil, Shubham M. Patil, Tatwadarshi P. Nagarhalli, "Real-Time Machine Translation System between Indian Languages" 978-1-6654-8328-5, 2022 IEEE
2. XiaoXue Fu, "English Machine Translation System based on Human-Computer Algorithm" 2022 IEEE
3. Ohidujjaman, Fahim Faysal, Shams Sumon, Mohammad Nural Huda, "Automatic Machine Translation for Bangla and English Resolving Ambiguities" 978-1-6654-1576-7, 2021 IEEE
4. Bala Harshithaa B, "Language Translator App"
5. M Vaishnavi, HR Dhanush Datta, Varsha Vemuri, L Jahnavi, "Language Translator Application" IJRASET45484, 2022
6. Jing Zhang, Fang Sun, "Research on the Application of Computer Artificial Intelligence Machine Translation System in Sci-tech Journals" 978-1-6654-0902-5, 2022 IEEE
7. Kai Jiang, Xi Lu, "Natural Language Processing and its Application in Machine Translation" 978-1-7281-7738-0, 2020 IEEE
8. Zhenhua Wei, "The Development Prospect of English Translation Software Based on Artificial Intelligence Technology"