

Brief Motivation

We chose our minor project “Legal Terminology Adaptation in Machine Translation for Indian Legal Documents” because of deep interest in Natural Language Processing and personal legal challenges that highlighted the importance of making legal information more accessible.

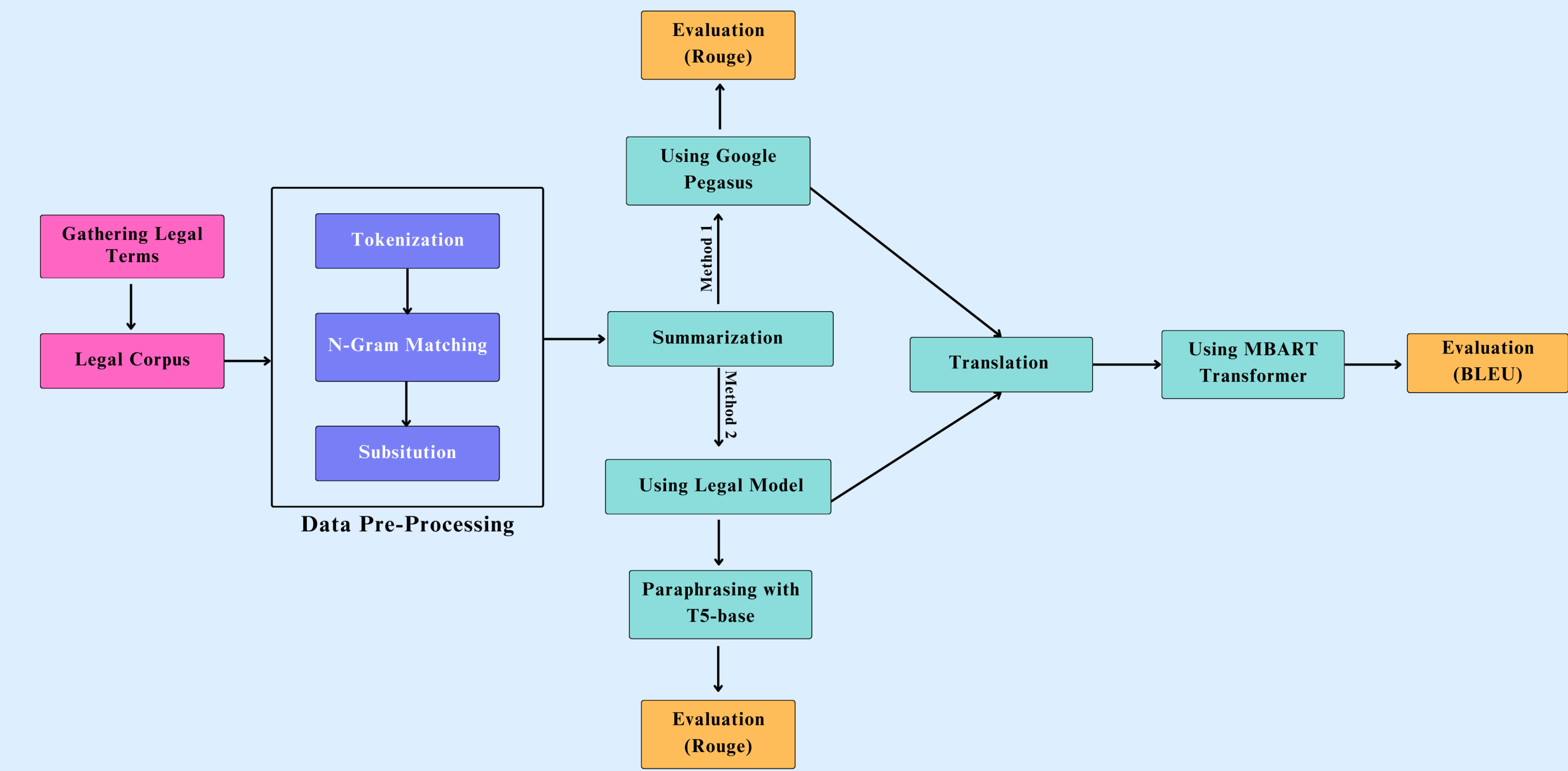
- **For Law Students and the Public:** Transforms complex legal language into clear, accessible terms, aiding students in exam preparation and empowering the public to understand laws and their rights more easily.
- **For Legal Professionals:** Accelerates case analysis by providing simplified summaries of legal documents, enabling judges and lawyers to quickly identify critical information.

Objective

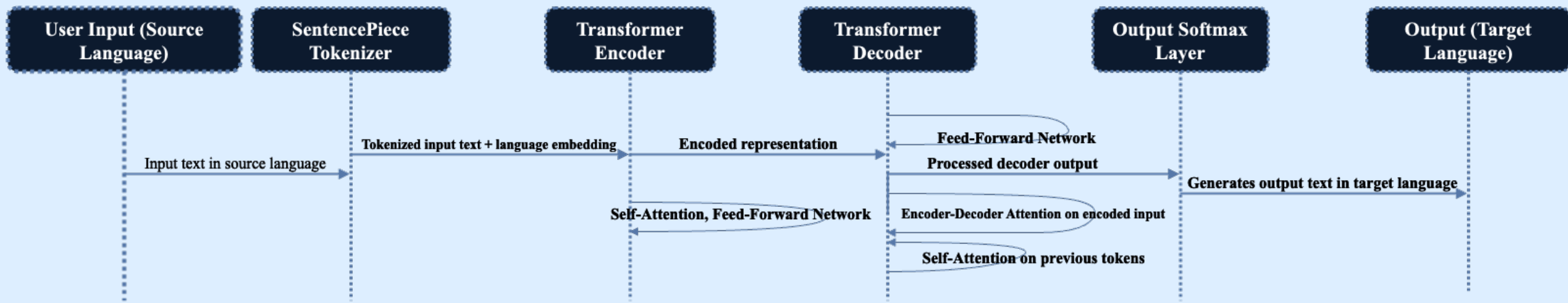
- Build a comprehensive corpus of complex legal terms and phrases from Indian legal documents.
- Enhance machine translation for Indian legal documents by simplifying and summarizing legal language.
- Translate the simplified summaries into Indian languages for better accessibility and understanding.

Methodology



Methodology Flow Diagram for Simplifying and Translating Legal Text Using NLP Models



Transition Diagram for Facebook's mBART Large 50 Many-to-Many Multilingual Model Machine Translation



Expected Results

|                        |   |  |
|------------------------|---|--|
| Original Legal Text    | During the hearing of the case, the jury deliberated on charges of arson.<br>The arson led to extensive property damage. The defendant's legal team argued for an acquittal.  |  |
| Summarization Method 1 | The (the criminal act of deliberately setting fire to property.) led to extensive (the name for anything which can be owned.) damage .<n>The defendant 's legal team argued for an (the court's decision that a person is innocent of the crime they were charged with.   | <pre>{'rouge1': Score(precision=0.5176470588235295, recall=1.0, fmeasure=0.682170542635659), 'rouge2': Score(precision=0.5, recall=0.9767441860465116, fmeasure=0.6614173228346456), 'rougeL': Score(precision=0.5176470588235295, recall=1.0, fmeasure=0.682170542635659)}</pre>  |
| Translation            |  Enter a sentence in English to translate to Hindi: The (the criminal act of deliberately setting fire to property.) led to extensive (the name for anything whi<br>Translation in Hindi: प्रतिवादी के कानूनी दल ने एक (न्यायालय के निर्णय के लिए तर्क दिया कि कोई व्यक्ति उन अपराधों से निर्दोष है जिन पर वह आरोपित किया गया था) के लिए तर्क दिया।<br>Python Translated Reference: (जानबूझकर संपत्ति में आग लगाने का आपराधिक कृत्य) से व्यापक क्षति हुई (किसी भी चीज़ का नाम जिसका स्वामित्व हो सकता है।) प्रतिवादी की कानूनी टीम ने अदालत<br>BLEU Score: 0.11059204097669541 |  |
| Summarization Method 2 | The jury deliberated on charges of the criminal act of deliberately setting fire on property . led to extensive the name for anything which can be owned . the defendant 's legal team argued for an the court's decision that a person is innocent of the crime they were charged with .   | <pre>{'rouge1': Score(precision=0.5411764705882353, recall=0.9787234042553191, fmeasure=0.696969696969697), 'rouge2': Score(precision=0.5238095238095238, recall=0.9565217391304348, fmeasure=0.6769230769230768), 'rougeL': Score(precision=0.5411764705882353, recall=0.9787234042553191, fmeasure=0.696969696969697)}</pre> |
| Translation            |  Enter a sentence in English to translate to Hindi: the jury deliberated on charges of the criminal act of deliberately setting fire to property . led to extens<br>Translation in Hindi: जुरी ने जानबूझकर संपत्ति पर आग लगाने के अपराध के आरोपों पर विचार-विमर्श किया. किसी भी संपत्ति के लिए विस्तृत नाम बनाने के लिए प्रेरित किया. '<br>Python Translated Reference: जुरी ने जानबूझकर संपत्ति में आग लगाने के आपराधिक कृत्य के आरोपों पर विचार-विमर्श किया।<br>BLEU Score: 0.14887309680192354  |  |

Outcomes

The project succeeded in creating simplified, accessible versions of complex legal documents.

- Summarization: Google Pegasus and a Legal-specific model produced concise summaries, achieving high ROUGE scores and retaining key details. Pegasus outperformed T5-base, enhancing readability.
- Translation: Using MBART, high BLEU scores confirmed clear, contextually accurate translations. MBART excelled in maintaining meaning and clarity.
- Legal Terms Identification: Complex legal terms were identified and adapted, aiding simplification.
- Accessibility: NLP techniques significantly improved the readability of legal information, simplifying complex content for broader audiences.

Overall, the project showed NLP’s potential to make legal documents more accessible while preserving essential information.

References

Luong, T., Pham, H., & Manning, C. D. (2015). Effective Approaches to Attention-based Neural Machine Translation. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Gehring, J., Auli, M., Grangier, D., & Dauphin, Y. N. (2017). A Convolutional Encoder Model for Neural Machine Translation. Proceedings of the 34th International Conference on Machine Learning, 70, 1243-1252.

Elnaggar, A., Gebendorfer, C., Glaser, I., & Matthes, F. (2018). Multi-Task Deep Learning for Legal Document Translation, Summarization, and Multi-Label Classification. arXiv preprint arXiv:1810.07513.

Premjith, B., Kumar, M. A., & Soman, K. P. (2019). Neural Machine Translation System for English to Indian Language Translation Using MTIL Parallel Corpus. Journal of Intelligent Systems, 28(3), 387–398.

Pathak, A., & Pakray, P. (2019). Neural Machine Translation for Indian Languages. Journal of Intelligent Systems, 28(3), 465–477.

Alves, A., Miranda, P., Mello, R., & Nascimento, A. (2023). Automatic Simplification of Legal Texts in Portuguese Using Machine Learning. Legal Knowledge and Information Systems.