News Translation and Summarizing

# Introduction:

The **News Translation and Summarizing** project aims to develop a Natural Language Processing model that can translate and summarize the input of news in text format from Telugu, Hindi, and Tamil languages to English. In today's globalized world, individuals and organizations often need to interact with people from different linguistic backgrounds. This can pose significant challenges, particularly when it comes to understanding and conveying complex information. Machine translation and summarization systems have the potential to overcome these barriers by enabling individuals to understand and get information of their own language to English.

# Objectives:

The primary objective of this project is to develop a machine learning model that takes news/articles in the form of text from Telugu, Hindi, Tamil languages and summarize it in English. To achieve this objective, we will use a combination of machine translation and summarization techniques. The following are the key objectives of this project:

* Develop a machine translation model: We will develop a machine translation model that can translate input text from Telugu, Hindi, Tamil languages to English.
* Develop a summarization model: We will develop a summarization model that can summarize input text in English.
* We will combine the machine translation and summarization models to develop a system that can translate and summarize input text in any language.

# Motivation:

The problem that we want to address through this project is the challenge of understanding and conveying complex information across languages. In today's globalized world, people and organizations frequently need to get information about their own places with individuals who speak different languages. This can create significant communication barriers, hindering effective information exchange and impeding business and get information regarding local actions happenings addressed in their native language newspapers.

Moreover, there is an abundance of information available in different languages that is inaccessible to individuals who do not speak those languages, limiting their access to knowledge and opportunities, this includes students with English as their primary language and wanted to learn the resources that only available in foreign language that they can’t understand.

The Translate and Summarize project plans to address these problems by developing a machine learning model that can translate and summarize input text from Telugu, Hindi, Tamil language to English. By doing so, the project aims to improve knowledge about their native places and information access, enabling individuals, organizations, and students to access information more effectively and easily. This will mainly help people who know only English language and need to get information of their native places and get knowledge of their surroundings.

# Features:

The following will be the features of our project:

* As there are two modules working one after other, we also wanted to use this modulus individually or combined. I.e. One should be able to use only translation part if they want or can only use summarization part or they can use the whole.
* We will try to implement a way to select the extent of summarization that the user wants. I.e. We will try to implement at least two levels of summarization.
* We will also try to design the system to produce accurate results.

# Significance:

Our project has a significant impact on student’s career and common man’s life too.

* With the ability to translate and summarize text from various languages, people around the globe can access a wealth of information and resources that would have otherwise been inaccessible to them quickly and accurately.
* The summarization feature can also help students to effectively manage the large volumes of information that they encounter, by extracting the most salient information and presenting it in a concise manner.
* By providing multilingual support and accurate summarization, our project can enable users to access and understand information from different sources in a timely manner, improving their overall efficiency and productivity.
  + For instance: Anyone can use the system to summarize news articles quickly and accurately from different countries, enabling them to stay informed about global events.
* Our project has the potential to enhance the quality of life for the common man by improving access to information and promoting cross-cultural communication.

# Workflow:

There are mainly three modules in our project which are.

* Data processing:
  + In this stage we will remove the stop words, noise, and other non-related information.
* Translation:
  + In this part of the project, we will convert the cleaned text that’s in different language to English.
* Summarization:
* In this section we will summarize the converted text that’s already in English language by using nlp techniques.
* Firstly, we will take the input text from user in any language.
* In next step we perform data processing on that data to remove any unwanted stop words or unnecessary information from that text.
* Later, we will perform the translation operation on the text which we get from the previous stage, which converts the text to English language.
* In the last stage we will perform summarization on the translated text, to simplify the text obtained from the translation module. Which results in the output we want (simplified English text of given language by the user).

# Conclusion:

In conclusion, our project aims to develop a system that can translate and summarize input news articles from Telugu, Tamil and Hindi languages to English using NLP techniques. The system will be developed using a combination of machine translation and summarization techniques. The successful completion of this project will contribute to the field of NLP and have practical applications in areas such as language learning, information retrieval, and document summarization.

# References:

[1] S. Vijay, V. Rai, S. Gupta, A. Vijayvargia and D. M. Sharma, "Extractive text summarisation in hindi," 2017 International Conference on Asian Language Processing (IALP), Singapore, 2017, pp. 318-321, doi: 10.1109/IALP.2017.8300607.

[2] L. Wan, "Extraction Algorithm of English Text Summarization for English Teaching," 2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Xiamen, China, 2018, pp. 307-310, doi: 10.1109/ICITBS.2018.00085.

[3] K. M. Chaman Kumar, S. Aswale, P. Shetgaonkar, V. Pawar, D. Kale and S. Kamat, "A Survey of Machine Translation Approaches for Konkani to English," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-6, doi: 10.1109/ic-ETITE47903.2020.110.

[4] V. A. Gandhi, V. B. Gandhi, D. V. Gala and P. Tawde, "A Study of Machine Translation Approaches for Gujarati to English Translation," 2021 Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 2021, pp. 1-5, doi: 10.1109/STCR51658.2021.9588859.

## Group members:

**Sai Aditya Teja Thogiti - 11555662**

**Prathyusha Manpur - 11550655**

**Saiteja Karanam – 11593704**

**Sindhu Marneni - 11555388**