**Task-6**

**Introduction**The Multilingual Educational Content Summarizer is a Streamlit-based web application designed to help users summarize, translate, and analyze educational content in multiple languages. It provides key features such as text summarization, keyword extraction, readability analysis, and translation. The system utilizes state-of-the-art machine learning models like MarianMT for translation and BART for text summarization, with natural language processing (NLP) functionalities provided by libraries such as spaCy and textstat.**Key Libraries and Models**

* **Streamlit:** A framework for building web applications with a simple Python interface.
* **Transformers (by Hugging Face):**
* **MarianMT:** A model used for machine translation.
* **BART:** A sequence-to-sequence model for text generation and summarization.
* **Sumy:** A library for automatic text summarization using algorithms such as LSA.
* **Textstat:** A library that provides text readability scores such as the Flesch Reading Ease and Flesch-Kincaid Grade.
* **spaCy:** An NLP library for text processing, including keyword extraction.
* **SentencePiece**: A tokenizer library, particularly useful for models that use subword tokenization like MarianMT.

**System Features**The application offers multiple key features:

* **Text Summarization:** Automatically generates a concise summary of the input educational content using the BART model.
* **Translation:** Translates the summarized content into a target language using the MarianMT model. The default target language is Spanish ('es'), but more languages can be added.
* **Keyword Extraction:** Extracts significant keywords from the summarized text using the spaCy NLP pipeline.
* **Readability Analysis:** Computes readability scores using the Flesch Reading Ease and Flesch-Kincaid Grade to measure how easy or difficult the text is to read.

**Code Breakdown**

* **Model Initialization:**
* MarianMT Model and Tokenizer: This is initialized using a pre-trained translation model from English to Spanish. The MarianMT model is designed to handle multilingual translation tasks efficiently.
* BART Model and Tokenizer: The facebook/bart-large-cnn model is loaded to handle text summarization. It is designed to generate high-quality, concise summaries of long texts.
* **Text Summarization:** The summarize\_text function encodes the input text into tokens using the BART tokenizer, then generates a summary using the BART model. Parameters such as max\_length and num\_beams are adjusted to control the length and quality of the summary.
* **Translation:** The translate\_text function translates the summarized text from English into the selected target language. It utilizes the MarianMT model to perform the translation. The default language is Spanish, but other languages can be supported by selecting from the dropdown menu.
* **Keyword Extraction:** Keywords are extracted from the text using spaCy. It removes stopwords and punctuation, keeping only relevant tokens that may be of interest to users.
* **Readability Analysis:** This feature uses the textstat library to compute two popular readability metrics:
* **Flesch Reading Ease:** Scores how easy the text is to read.
* **Flesch-Kincaid Grade:** Indicates the U.S. school grade level necessary to understand the text.
* **Streamlit User Interface:** The Streamlit interface allows the user to input educational content, select a target language, and execute summarization, translation, and readability analysis through simple button clicks. Each result is displayed under relevant headers.

**System Workflow**

* **Text Input:** Users input educational content, which can be a passage or article.
* **Summarization:** The input text is summarized using the BART model.
* **Keyword Extraction:** The summarization result is passed through spaCy for keyword extraction.
* **Readability Analysis**: Readability scores are calculated to give insights into the complexity of the text.
* **Translation:** The summarized text is translated into the selected target language (e.g., Spanish or Chinese).
* **Results Display:** The original text, summary, extracted keywords, readability scores, and translation are displayed for the user.

**Strengths**

* **Multilingual Support:** The application supports multilingual translation, making it accessible for non-English speakers.
* **Easy-to-Use Interface:** Streamlit provides an intuitive interface, enabling users to perform complex NLP tasks without technical expertise.
* **Efficient Summarization:** BART's pre-trained model produces high-quality summaries, enhancing the understanding of large educational texts.
* **Comprehensive Readability Metrics:** Readability scores offer users a quick assessment of how challenging the text is for different reading levels.

**Conclusion**The Multilingual Educational Content Summarizer provides a robust tool for summarizing and translating educational content, making it easier for students and educators to access and comprehend complex information across languages. With further improvements and extensions, this application could become a valuable asset in educational environments, particularly in multilingual and multicultural contexts.