**AUTOMATED TEXT SUMMARIZATION SYSTEM**

**ABSTRACT**

Abstractive multi-document summarization is a type of automatic text summarization. It obtains information from multiple documents and generates a human-like summary from them.Text summarization is an essential task in Natural Language Processing (NLP) that involves condensing large volumes of text into shorter, meaningful summaries while retaining key information. This technique various NLP techniques for text summarization, focusing on both extractive and abstractive methods. Extractive summarization selects significant sentences or phrases directly from the original text to form a summary, while abstractive summarization generates new sentences that capture the essence of the input text.In this era of technology, everything around us is digitized. People tend to develop ideas that perform activities that only humans were able to do before the innovation of modern technology. Summarizing text documents is one such example. We have developed various NLP models to perform text summarization. While efficient models exist for native English, little attention is given to Indian languages. Through comparative analysis, this system highlights the strengths and limitations of each approach and outlines potential applications, including document summarization, news aggregation, and query-based summarization systems. The integration of NLP techniques into real-world applications showcases the potential of automatic text summarization to enhance information processing and decision-making across various domains.

**MODULES**

* **Dataset Collection**
* **Data Preprocessing**
* **Feature Extraction**
* **Summarization Model**
* **Model Training and Evaluation**
* **User Interface Module**

**LANGUAGES USED:**

**FRONT END: HTML, CSS, JAVASCRIPT**

**BACK END: PYTHON**

**FRAMEWORK: FLASK**

**MODULE DESCRIPTION**

**Dataset Collection:**

We have large volumes of text and need to generate summaries quickly, you can use pre-trained models to generate summaries. While this won’t replace human-written summaries, it can still be useful for training machine learning models.Generate summaries automatically for large datasets and then fine-tune your model based on the generated data.

**Data Preprocessing:**

This module handles the preprocessing of raw text data, which is crucial for improving the quality of the input text before feeding it into a machine learning or NLP model.Splitting text into smaller units like words or sentences.Eliminating common but irrelevant words (e.g., "the", "is", "in") and Removing noise like punctuation, numbers.

**Feature Extraction:**

Converts preprocessed text data into numerical features that can be used by machine learning algorithms.Represents text as word frequency counts.Measures the importance of words in a document relative to a corpus, uses pre-trained models.

**Summarization Model:**

Summarization module implements the machine learning algorithm that generates the summary of the text. There are two primary approaches to text summarization:

* **Extractive Summarization**: Extracts key sentences or phrases from the text based on their importance.
* **Abstractive Summarization**: Generates new sentences that summarize the text, similar to how humans write summaries.

**Model Training and Evaluation:**

This module is responsible for training the summarization model using labeled data (for supervised learning) or unsupervised learning methods. It also handles the evaluation of the model to ensure it is generating accurate summaries.Fine-tuning the summarization model on a dataset of articles and summaries.Measuring the performance of the model using metrics.

**User Interface Module:**

This module involves deploying the trained summarization model for real-time use. The model can be deployed as a web service.The front-end module allows users to interact with the text summarization system. Users can input text or upload documents, and the interface will display the generated summary.Allow users to upload documents for summarization.

**ALGORITHMS**

**NATURAL LANGUAGE PROCESSING**

The meaning of NLP is Natural Language Processing (NLP) which is a fascinating and rapidly evolving field that intersects computer science, artificial intelligence, and linguistics. NLP focuses on the interaction between computers and human language, enabling machines to understand, interpret, and generate human language in a way that is both meaningful and useful. With the increasing volume of text data generated every day, from social media posts to research articles, NLP has become an essential tool for extracting valuable insights and automating various tasks.NLP powers many applications that use language, such as text translation, voice recognition, text summarization, and Chabot’s. You may have used some of these applications yourself, such as voice-operated GPS systems, digital assistants, speech-to-text software, and customer service bots. NLP also helps businesses improve their efficiency, productivity, and performance by simplifying complex tasks that involve language.