

Text Summarization

Abstract:

In this new era, where tremendous information is available on the internet, it is most important to provide an improved mechanism to extract the information quickly and most efficiently. It is very difficult for human beings to manually extract the summary of a large document of text. There is plenty of text material available on the internet. So there is a problem of searching for relevant documents from the number of documents available, and absorbing relevant information from it. In order to solve the above two problems, the automatic text summarization is very much necessary. Text summarization is the process of identifying the most important meaningful information in a document or set of related documents and compressing them into a shorter version preserving its overall meanings.

Existing system:

Text summarization is an important NLP task, which has several applications. The two broad categories of approaches to text summarization are *extraction* and *abstraction*. Extractive methods select a subset of existing words, phrases, or sentences in the original text to form a summary. In contrast, abstractive methods first build an internal semantic representation and then use natural language generation techniques to create a summary. Such a summary might contain words that are not explicitly present in the original document. Most text summarization systems are based on some form of extractive summarization.

Thus, the proposed system would solve these kind of texts.

Proposed System:

This process can be divided into two steps: Pre Processing step and Processing step. Pre Processing is a structured representation of the original text. It usually includes: a) Sentences boundary identification. In English, the sentence boundary is identified with the presence of a dot at the end of a sentence. b) Stop-Word Elimination—Common words with no semantics c) Stemming—The purpose of stemming is to obtain the stem or radix of each word, which emphasises its semantics. In the Processing step, features influencing the relevance of sentences are decided and calculated and then weights are assigned to these features using a weight learning method. Final score of each sentence is determined using the Feature-weight equation. Top ranked sentences are selected for final summary. Summary evaluation is a very important aspect for text summarization. Generally, summaries can be evaluated using intrinsic or extrinsic measures. While intrinsic methods attempt to measure summary quality using human evaluation and extrinsic methods measure the same through a task-based performance measure such as the information retrieval-oriented task.

Software Tools:

1. VS Code
2. Jupyter Notebook
3. TensorFlow
4. Keras
5. Anaconda
6. Scikit-Learn
7. Matplotlib

Hardware Tools:

1. Laptop
2. Operating System: Windows 11
3. RAM: 16GB
4. ROM: 4GB
5. GPU
6. Fast Internet Connectivity