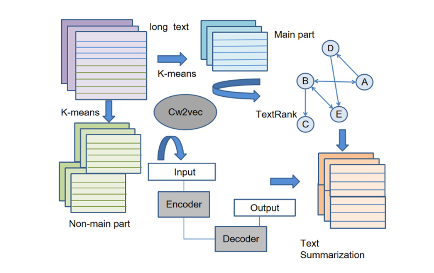
**TEXT SUMMARIZATION METHODS**

**INTRODUCTION:**

Text summarization is a Natural Language Processing (NLP) technique which condenses long documents into shorter, coherent summaries while retaining the most important information. As technology is growing fast these days, digital content is growing as well. Sectors like business, healthcare, news etc. have a lot of digital content. As digital content growing methods like Text summarization are useful to efficient information retrieval and decision-making. This literature survey explores the evolution of text summarization methods, focusing on **extractive** and **abstractive** approaches, and highlights recent advancements in combining these methods for improved performance. The survey also examines the role of **language-specific models** (e.g., Cw2vec for Chinese) and **structured text processing** in enhancing summarization quality. The goal is to provide a comprehensive overview of existing techniques and their applicability to real-world tasks, such as summarizing emails, news articles, or reports.

**BODY:**

Text summarization uses the K- Means algorithm to divide the long text into main-part and non-main part. The Summarization task into two sub-task modes, use the extractive method to locate and extract the key content, and then use the extractive method to rewrite the content.



**Extractive Summarization:**

Extractive summarization involves selecting the main part of text from original text and concatenating them to form a summary. It does not generate new text but rather extracts existing content.

**Techniques**

**Graph-Based Algorithm:**

**TextRank** is a graph-based algorithm used for extractive summarization. It ranks sentences or phrases based on their importance within the text, similar to how Google's PageRank algorithm ranks web pages. The text is represented as a graph, where each sentence is a node, and edges represent the similarity between sentences. algorithm iteratively computes the importance of each sentence based on the number and importance of sentences that link to it. Sentences with higher scores are considered more important.

Summary Generation: The top-ranked sentences are selected and concatenated to form the summary.

**Relation to my project:**

Extractive methods can be used to pull out key sentences from emails, such as meeting details, deadlines, or action items. This ensures that the most critical information is retained in the summary.

**Abstractive Summarization:**

Abstractive summarization generates new sentences that capture the essence of the original text, often using paraphrasing or rewriting techniques.

Techniques:

Sequence-to-Sequence Models: Seq2Seq models encode the input text and decode it into a shorter version. These models are often based on LSTM or Transformer architectures. LSTM is a type of recurrent neural network (RNN) designed to handle sequential data, such as text. It is particularly effective for tasks that require understanding long-term dependencies in the data. LSTMs use memory cells to store and retrieve information over long sequences. These cells can maintain information for extended periods, making LSTMs suitable for tasks like text generation and summarization.

LSTMs have three gates (input, forget, and output) that control the flow of information. These gates decide what information to keep, discard, or output at each step. Sequence Processing: For text summarization, the input text is processed sequentially, and the LSTM generates a summary by predicting the next word or sentence based on the context.

**Relation to my project:**

Abstractive methods can be used to **rewrite** or **condense** longer explanations or discussions in emails, making the summaries more concise and readable. This is particularly useful for summarizing complex or lengthy emails.

RESEARCH PAPER: A Text Structure-based Extractive and Abstractive Summarization Method.

Link: https://ieeexplore-ieee-org.rlib.pace.edu/document/9778497.