DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Intelligent Information Aggregation and Condensation System

Under the guidance of:-
Mrs.B.Sushmitha
(Assistant professor)

Batch no- B10
JVR vinayak:-215U1A0574
K vamshi:-215U1A0575
K sai varshith:-215U1A0581
L lokesh:-215U1A05A7

Introduction

- Our system uses the latest transformer models to create clear and high-quality summaries from large amounts of text.
- It combines extractive and abstractive methods, ensuring that summaries are brief, meaningful, and easy to understand.
- This tool is useful for fields like news, research, healthcare, and law, helping users quickly grasp key information and make better decisions.

Objectives

- Use advanced transformer models to create accurate and high-quality text summaries.
- Ensure the system works smoothly with PDFs, Word documents, and plain text while being fast and scalable.
- Help users quickly process and understand large amounts of text, saving time and effort.

Literature review

S.NO	Author name and Year of publication	Title of the paper	Methods used	Parameters analysed	Limitations
1	Hassan Shakil, Ahmad Farooq, Jugal Kalita 2024	Abstractive text summarization	Sequence-to-Sequence (Seq2Seq) Model	Using NLP models	Using old models of NLP like seq2seq
2	Rupal Bhargavaa, Yashvardhan Sharmaa 2020	Deep Extractive Text Summarization	Deep Learning	Using Deep Learning For Extractive summarization	Relies Too Much on Sentence Ranking

Problem Statement

- There is a huge amount of digital content, and reading through all of it takes a lot of time and effort.
- Deep Extractive Text Summarization struggles to select key sentences while ensuring the summary remains clear and meaningful.
- Abstractive text summarization struggles with maintaining accuracy, coherence, and context while requiring high computational resources.

Disadvantages

•**High Processing Power Requirement** – The system relies on advanced transformer models, which need strong hardware to function efficiently. Running it on low-end devices can be slow or challenging.

•**Extractive Summarization Drawback** – This method picks sentences directly from the text without rewriting them, which can result in repetitive or less connected summaries.

•**Abstractive Summarization Drawback** – Since this method generates new sentences, it may sometimes misinterpret the meaning, leading to incorrect or unclear summaries, especially for complex topics.

Proposed Model

- Our proposed model uses advanced transformers to improve accuracy, readability, and context understanding while making the summarization process faster and more efficient.

Advantages

•**More Accurate and Easy to Read** – Using BERT, T5, and GPT-based transformers, the system ensures summaries keep important details while sounding natural and well-structured.

•**Less Repetition** – With extractive models like TextRank and abstractive models like T5, the system selects key sentences and rephrases them to avoid unnecessary repetition.

•**Better Understanding of Text** – Advanced transformers help the system grasp the true meaning of the text, creating summaries that are clear, meaningful, and contextually accurate.

Requirements

Hardware Requirements:

- **Processor:** Intel Core i5 or higher
- **RAM:** 8GB or more
- **Storage:** 50GB HDD/SSD
- **Operating System:** Windows 10/11

Software Requirements:

- **Languages:** Python, HTML, CSS, JavaScript
- **Framework:** Streamlit (for web deployment)
- **Tools:** PyCharm, Visual Studio Code, **Google Colab** (for cloud-based execution)

Libraries: Transformers (BERT, T5, GPT), NLTK, TextRank, NumPy, Pandas

THANK YOU