

TF-IDF EXTRACTIVE TEXT SUMMARIZER IN JAVA

Problem Statement

In today's information-rich environment, readers, ranging from students and researchers to business professionals, are inundated with lengthy documents, articles, and reports. Manually skimming through large volumes of text to extract key ideas is time-consuming and error-prone.

Objectives

- Accurately extract the most informative sentences
- Offer flexible output formats for different reading styles
- Maintain a modular and extensible architecture

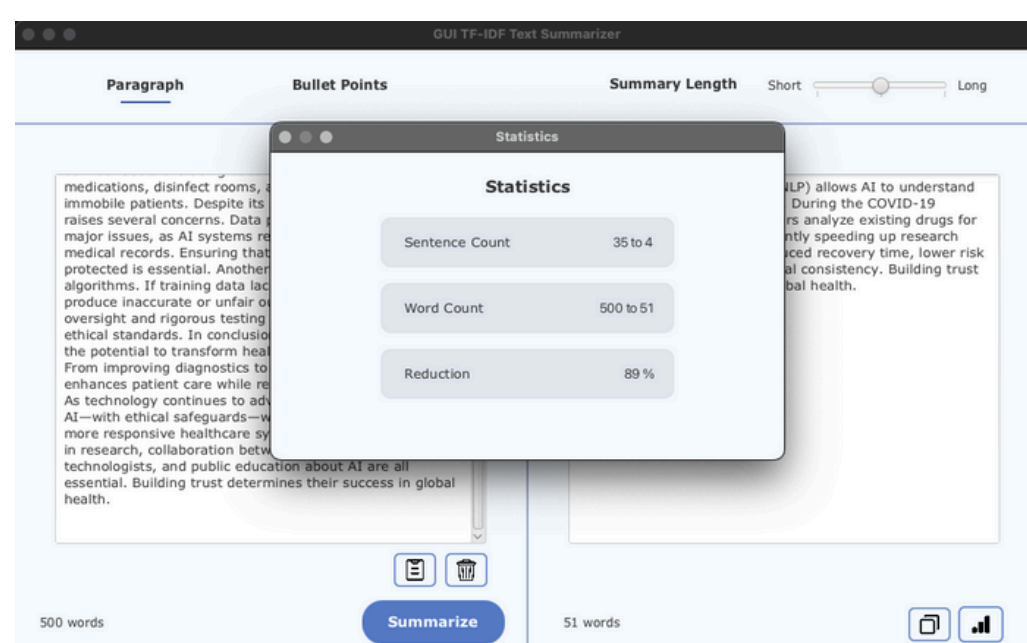
System

Input → Sentence Splitter → Text Processor → TF-IDF Scorer → Rank & Select → Formatter → Output

Algorithm

1. Split sentences from input by handling honorifics, punctuation, and quotes.
2. Preprocess split sentences through stop-word removal and stemming.
3. Compute TF-IDF score and rank sentences based on this.
4. Output top N sentences based on the length requested by the user.

GUI Program



COMP6699001 - Object Oriented Programming

Juan Felix Kusnadi - 2802536386
Jonathan Mulyono - 2802537054
Iglesias Sidharta Handjo - 2802530621