

**IMPLEMENTASI METODE EKSTRAKTIF UNTUK SISTEM
RINGKASAN OTOMATIS ARTIKEL BERITA BERBAHASA INDONESIA
MENGUNAKAN *TERM FREQUENCY-INVERS DOCUMENT
FREQUENCY* (TF-IDF)**

Andhanu Prakoso Aji

ABSTRAK

Pesatnya penyebaran informasi melalui portal berita daring mendorong kebutuhan akan sistem peringkasan teks otomatis untuk membantu pembaca memahami inti berita dengan lebih cepat dan efisien. Penelitian ini mengembangkan sistem peringkasan otomatis berbasis ekstraktif menggunakan algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk artikel berita berbahasa Indonesia dari portal Kompas.com. Proses pengumpulan data dilakukan melalui teknik *web scraping* menggunakan pustaka BeautifulSoup, diikuti dengan tahapan prapemrosesan data yang meliputi pembersihan *noise*, pemotongan kalimat (*sentence splitting*), tokenisasi, penghapusan *stopword*, serta *stemming* menggunakan Sastrawi. Sistem ini diimplementasikan dalam bentuk aplikasi web menggunakan framework Streamlit. Evaluasi performa sistem dilakukan terhadap 50 artikel berita dengan membandingkan hasil ringkasan sistem terhadap ringkasan manual menggunakan metrik *ROUGE Score* (ROUGE-1, ROUGE-2, dan ROUGE-L). Hasil pengujian menunjukkan nilai rata-rata akurasi ROUGE-1 sebesar 57,47%, ROUGE-2 sebesar 44,16%, dan ROUGE-L sebesar 46,23%. Temuan ini membuktikan bahwa algoritma TF-IDF efektif dalam mengidentifikasi unit informasi utama yang relevan dari teks asli meskipun dipengaruhi oleh subjektivitas manusia dan batasan teknis pada proses akuisisi data. Sistem ini diharapkan dapat mempermudah akses informasi bagi pembaca berita serta menjadi landasan pengembangan teknologi pemrosesan bahasa alami di masa mendatang.

Kata Kunci: Ekstraktif, Kompas.com, Peringkasan Teks Otomatis, *ROUGE Score*, Streamlit, TF-IDF, *Web Scraping*.

IMPLEMENTATION OF EXTRACTIVE METHOD FOR INDONESIAN NEWS ARTICLE AUTOMATIC SUMMARIZATION SYSTEM USING TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF)

Andhanu Prakoso Aji

ABSTRACT

The rapid dissemination of information through online news portals has driven the need for automatic text summarization systems to help readers understand the core of news more quickly and efficiently. This study develops an extractive-based automatic summarization system using the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm for Indonesian news articles from the Kompas.com portal. The data collection process was carried out through web scraping techniques using the BeautifulSoup library, followed by data preprocessing stages including noise removal, sentence splitting, tokenization, stopword removal, and stemming using Sastrawi. The system is implemented as a web application using the Streamlit framework. System performance evaluation was conducted on 50 news articles by comparing the system's summary results to manual summaries using ROUGE Score metrics (ROUGE-1, ROUGE-2, and ROUGE-L). The test results showed average accuracy values of 57.47% for ROUGE-1, 44.16% for ROUGE-2, and 46.23% for ROUGE-L. These findings demonstrate that the TF-IDF algorithm is effective in identifying the main relevant information units from the original text, despite being influenced by human subjectivity and technical limitations in the data acquisition process. This system is expected to facilitate information access for news readers and serve as a foundation for future natural language processing technology development.

Keywords: Automatic Text Summarization, Extractive, Kompas.com, ROUGE Score, Streamlit, TF-IDF, Web Scraping.