

ABSTRAK

Jeremia Jordan Sihombing, NIM 4203250021 (2024), Implementasi Text Summarization pada Artikel Berbahasa Indonesia Menggunakan Algoritma TextRank dengan TF-IDF Berbasis Web

Peningkatan signifikan jumlah artikel ilmiah setiap tahun menciptakan tantangan dalam memproses informasi secara efisien. Hal ini menyebabkan kebutuhan akan teknologi *text summarization* yang dapat membantu peneliti dan pembaca menghemat waktu dalam memahami inti artikel. Penelitian ini menggunakan *extractive summarization* untuk mempertahankan kalimat asli dari teks sumber, menjaga integritas informasi, dan menghindari distorsi makna. Penelitian ini bertujuan mengembangkan sistem *text summarization* untuk artikel berbahasa Indonesia menggunakan algoritma *TextRank* dan TF-IDF berbasis web. Algoritma *TextRank* dipilih karena kemampuannya mengidentifikasi kalimat penting tanpa pelatihan, sementara TF-IDF digunakan untuk pembobotan kata berdasarkan frekuensi kemunculannya dalam dokumen. Data yang digunakan adalah 100 artikel berbahasa Indonesia dari Jurnal Kode Unimed tahun 2022-2024. *Text Summarization* dikembangkan menggunakan tahapan: *Text preprocessing*, pembobotan TF-IDF, *cosine similarity*, dan perangkingan kalimat. Validasi dilakukan oleh ahli bahasa dan pakar website menggunakan skala Likert untuk menilai kualitas ringkasan dan kegunaan sistem. Sistem berbasis website yang dibuat memungkinkan pengguna memasukkan teks artikel dan mendapatkan ringkasannya dengan mudah dan cepat. Hasil menunjukkan sistem mampu menghasilkan ringkasan yang mempertahankan informasi penting dari artikel asli. Evaluasi oleh ahli bahasa menunjukkan hasil ringkasan yang sangat baik, dengan performa terbaik pada *compression rate* 50% (88,533%), diikuti *compression rate* 40% (85,133%), dan *compression rate* 30% (81,26%). Evaluasi oleh pakar website juga menunjukkan hasil yang baik dengan tingkat kelayakan sistem sebesar 70,67%, mengonfirmasi kegunaan dan efektivitas sistem.

Kata Kunci : Bahasa Indonesia, *Text Summarization*, *TextRank*

ABSTRACT

Jeremia Jordan Sihombing, NIM 4203250021 (2024), Implementation of Text Summarization on Indonesian-Language Articles Using a Web-Based TextRank Algorithm with TF-IDF

The significant increase in the number of scientific articles each year creates challenges in efficiently processing information. This leads to the need for text summarization technology that can help researchers and readers save time in understanding the essence of articles. This study uses extractive summarization to retain the original sentences from the source text, maintain the integrity of the information, and avoid distortion of meaning. The research aims to develop a text summarization system for Indonesian-language articles using a web-based TextRank and TF-IDF algorithm. TextRank was chosen for its ability to identify important sentences without training, while TF-IDF is used for word weighting based on their frequency in the document. The data used consists of 100 Indonesian-language articles from the Unimed Code Journal from 2022-2024. The text summarization is developed through stages including text preprocessing, TF-IDF weighting, *cosine similarity*, and sentence ranking. Validation is conducted by language experts and website experts using a Likert scale to assess the quality of the summaries and the usefulness of the system. The web-based system allows users to input article text and quickly obtain its summary. Results show that the system can produce summaries that retain the important information from the original articles. Evaluations by language experts show excellent summary results, with the best performance at a 50% *compression rate* (88.533%), followed by a 40% *compression rate* (85.133%), and a 30% *compression rate* (81.26%). Evaluations by website experts also show good results with a system feasibility rate of 70.67%, confirming the system's usefulness and effectiveness.

Keywords: Indonesian language, Text Summarization, TextRank

