

ABSTRAK

Dalam penelitian ini bertujuan membangun aplikasi Automatic Text Summarization berbasis website dengan menggunakan teknik web scraping pada berita dari news.detik.com. Metode ekstraktif dengan algoritma TF-IDF digunakan untuk merangkum teks secara efisien, dengan fokus pada pembobotan kalimat dan seleksi berdasarkan nilai tertinggi. Hasilnya diukur menggunakan metode evaluasi Recall-Oriented Understudy for gisting Evaluation (ROUGE) dengan empat varian: ROUGE-1, ROUGE-2, ROUGE-3, dan ROUGE-L. Dalam tahapan penelitian, ekstraksi teks melalui web scraping diikuti oleh text preprocessing, text processing, dan penyusunan kalimat terpilih berdasarkan urutan aslinya. Rata-rata hasil evaluasi ROUGE terhadap studi kasus berita yang diberikan dan di evaluasi dengan sepuluh referensi *human gold standart* menunjukkan kinerja yang baik pada informasi unigram (ROUGE-1), dengan nilai recall 40,1%, precision 50,6%, dan F-Score 42,9%. Namun, kinerja menurun pada informasi bigram (ROUGE-2) dengan nilai recall 21,8%, precision 32,1%, trigram (ROUGE-3) dengan nilai recall 16,6%, precision 27,5%, f-score 20,2%. ROUGE-L menunjukkan keseimbangan antara recall dan precision, dengan nilai recall 30,8%, precision 40,8%, dan F-Score 33,6%, menunjukkan kemampuan sistem dalam menangkap informasi keseluruhan. Di sisi lain. Kesimpulannya, meskipun sistem efektif dalam menangkap informasi unigram, kinerjanya menurun pada bigram dan n-gram yang lebih panjang.

Kata Kunci : Peringkas teks, pembobotan kata, *website* berita, tf-idf

ABSTRACT

This research aims to build a website-based Automatic Text Summarization application by using web scraping techniques on news from news.detik.com. Extractive method with TF-IDF algorithm is used to summarize the text efficiently, focusing on sentence weighting and selection based on the highest value. The results are measured using the Recall-Oriented Understudy for gisting Evaluation (ROUGE) evaluation method with four variants: ROUGE-1, ROUGE-2, ROUGE-3, and ROUGE-L. In the research phase, text extraction through web scraping is followed by text preprocessing, text processing, and the arrangement of selected sentences based on their original order. The average ROUGE evaluation results on the given news case study and evaluated with ten human gold standard references show good performance on unigram information (ROUGE-1), with a recall value of 40.1%, precision of 50.6%, and F-Score of 42.9%. However, the performance decreases on bigram information (ROUGE-2) with a recall value of 21.8%, precision 32.1%, trigram (ROUGE-3) with a recall value of 16.6%, precision 27.5%, f-score 20.2%. ROUGE-L shows a balance between recall and precision, with a recall value of 30.8%, precision of 40.8%, and F-Score of 33.6%, indicating the system's ability to capture overall information. On the other hand. In conclusion, although the system is effective in capturing unigram information, its performance degrades on longer bigrams and n-grams.

Keywords: Text summarization, word weighting, news website, tf-idf