

Perbandingan Performa Algoritma TextRank dengan Unweighted dan Weighted Word Embedding pada Peringkasan Dataset IndoSum = Performance Comparison of TextRank Algorithm with Unweighted and Weighted Word Embedding on IndoSum Dataset Summarization

Piawai Said Umbara, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920551418&lokasi=lokal>

Abstrak

Penelitian ini bertujuan untuk mengevaluasi efektivitas kombinasi metode TextRank dengan word embedding pada dataset IndoSum. Dua skenario eksperimen diterapkan: unweighted dan weighted. Hasil eksperimen menunjukkan bahwa penggunaan model word embedding pada algoritma TextRank terbukti meningkatkan performa sistem peringkasan. Skenario unweighted menguji perbandingan performa antara sistem baseline (kombinasi dari TextRank dengan Word2Vec dan pembobotan TF-IDF yang dilatih menggunakan dataset Liputan6) dengan beberapa sistem lainnya yang memadukan algoritma TextRank dengan model word embedding tanpa pembobotan TF-IDF. Skenario weighted menguji perbandingan sistem yang memadukan algoritma TextRank dengan model word embedding tanpa pembobotan TF-IDF dengan counterpart-nya yang memadukan algoritma TextRank dengan model word embedding ditambah pembobotan TF-IDF. Pada skenario unweighted, kombinasi TextRank dan fastText yang dilatih dengan dataset IndoSum menunjukkan peningkatan performa tertinggi untuk skor ROUGE-1 dan ROUGE-2 masing-masing sebesar 7,51% dan 4,24%. Sedangkan pada skenario weighted, pembobotan TF-IDF pada word embedding menunjukkan peningkatan performa sebagian besar sistem peringkasan dibandingkan penggunaan word embedding tanpa pembobotan TF-IDF, dengan rata-rata peningkatan performa sistem sebesar 5,55% pada ROUGE-1 dan 9,95% pada ROUGE-2.

.....This research evaluates the effectiveness of combining the TextRank method with word embedding on the IndoSum dataset. Two experimental scenarios were applied: unweighted and weighted. The experimental results show that using word embedding models in the TextRank algorithm significantly improves the summarization system's performance. The unweighted scenario tested the performance comparison between the baseline system (a combination of TextRank with Word2Vec and TF-IDF weighting trained using the Liputan6 dataset) and several other systems combining the TextRank algorithm with word embedding models without TF-IDF weighting. The weighted scenario compared systems that combined the TextRank algorithm with word embedding models without TF-IDF weighting to their counterparts that included TF-IDF weighting. In the unweighted scenario, the combination of TextRank and fastText trained with the IndoSum dataset demonstrated the highest performance improvement, with ROUGE-1 and ROUGE-2 scores increasing by 7.51% and 4.24%, respectively. Meanwhile, in the weighted scenario, applying TF-IDF weighting to word embedding showed performance improvement in most summarization systems compared to using word embedding without TF-IDF weighting, with an average performance increase of 5.55% in ROUGE-1 and 9.95% in ROUGE-2.