

ABSTRAK

Menghitung kesamaan kalimat Bahasa Indonesia menggunakan Latent Semantic Indexing berdasarkan KBBI (Kamus Besar Bahasa Indonesia)

Menghitung kemiripan semantik antara kalimat adalah masalah yang lama dibahas di bidang pengolahan bahasa. Bidang analisis semantik memiliki peran penting dalam penelitian yang berkaitan dengan analisis teks. Dalam penelitian ini, data yang digunakan adalah Definisi dari tiap kata yang memiliki makna yang sama berdasarkan Kamus Besar Bahasa Indonesia. Dalam penelitian ini pun, penulis menyajikan 2 metodologi dengan kemiripan semantik menggunakan algoritma TF-IDF tradisional dan metode LSI menggunakan TF-IDF dengan distribusi term dari definisi. Semua definisi dihitung dengan definisi lainnya, dengan menganalisa definisi yang mempunyai nilai similarity yang tinggi beserta akurasi. Setiap perhitungan bobot dan vektor yang dihasilkan oleh algoritma TF-IDF setiap metode, langkah perhitungan dengan mereduksi dimensi vektor sangat mempengaruhi hasil dari nilai kesamaan. Terbukti dari penelitian ini metode dengan menambah distribusi term, mendapatkan nilai bobot lebih besar dan mempengaruhi nilai kesamaan. Metode LSI pun memperkuat algoritma TF-IDF dengan distribusi untuk menentukan kesamaan kalimat yang lebih baik, dengan memperhatikan juga karakter dari KBBI. Hasil yang diperoleh pada penelitian ini, nilai akurasi sebesar 75.9 % untuk semantik similarity menggunakan TF-IDF tradisional dan 80% untuk metode LSI dan TF-IDF menggunakan bobot distribusi term.

Kata kunci: Sentence Similarity, TF-IDF, Semantic Similarity, Latent Semantic Indexing