

ABSTRAK

Dalam pemrosesan bahasa alami (PBA), mengukur kesamaan semantik memiliki peranan penting. Hasil pengukuran tersebut seringkali dijadikan dasar dalam melakukan tugas-tugas pemrosesan bahasa alami lebih lanjut. Metode-metode yang digunakan dalam mengukur kesamaan semantik antara lain: metode berbasis vektor, metode berbasis *path*, metode berbasis konten informasi, metode berbasis *gloss*, dan metode *hybrid*. Pada tugas akhir ini, diimplementasikan metode berbasis konten informasi dalam mengukur kesamaan semantik untuk mengetahui keakuratan kesamaan antar kata. Konten informasi yang digunakan dalam implementasi berasal dari WordNet. Hasil implementasi dibandingkan dengan *dataset* Gold Standard untuk diukur nilai korelasinya. *Dataset* yang digunakan untuk pengujian yakni *dataset* WordSim-353 *Similarity*, WordSim-353 *Relatedness*, dan SimLex-999. Berdasarkan pengujian yang dilakukan, Korelasi Pearson dan Spearman metode Lin untuk *gold standard* kata benda dan kata kerja SimLex-999 berturut-turut sebesar 0,532 dan 0,543. Metode konten informasi intrinsik yang diwakili oleh metode Lingling Meng menghasilkan Korelasi Pearson, dan Korelasi Spearman berturut-turut sebesar 0,475 dan 0,551. Metode konten informasi dasar yang menggunakan kombinasi korpus dengan struktur WordNet, dan metode konten informasi intrinsik menghasilkan performa positif kuat.

Kata kunci: Pemrosesan Bahasa Alami, Kesamaan Semantik, Metode Berbasis Konten Informasi, WordNet, *Gold Standard*