

Abstrak

Kesamaan semantik adalah metrik kesamaan antar kata, kalimat atau dokumen yang berbagi dalam elemen makna. Perhitungan terkaitan semantik memiliki peranan penting dalam *data mining*, pengambilan informasi, dan bahkan *natural language processing*. Pada bahasa Indonesia, perhitungan kesamaan semantik mendapat peran penting karena banyak dimanfaatkan untuk aplikasi lain, seperti klasifikasi teks. Pengukuran kesamaan semantik dapat dilakukan dengan pendekatan berbasis korpus dan pendekatan berbasis kamus. Pada Tugas Akhir ini dilakukan pembangunan model kesamaan semantik berbasis korpus yang direpresentasikan dengan distribusional *semantic vector*. Model kemudian diujikan pada beberapa pasang kata dengan derajat kesamaan semantik bervariasi. Model kesamaan semantik dibangun berdasar korpus Wikipedia Bahasa Indonesia, dengan metode *word2vec*. Hasil pengujian pada dataset uji yang juga digunakan pada penelitian sebelumnya berdasar pada referensi *SimLex999* dan *Rubenstein-goodenough* menunjukkan nilai korelasi yang diperoleh 0.2753. Walaupun nilai korelasi tersebut lebih kecil dibanding nilai pada penelitian sebelumnya dengan pendekatan korpus, terdapat beberapa kasus di mana model semantik berbasis korpus mampu menangkap korelasi semantik lebih baik.

Kata kunci : kesamaan semantik, bahasa Indonesia, persamaan kosinus.