

Abstrak

Pencarian informasi sudah menjadi bagian dari kebutuhan manusia, terutama pencarian informasi menggunakan bahasa sehari – hari. Salah satu contohnya adalah Bahasa Indonesia. Dalam melakukan pencarian informasi yang efektif, diperlukan kecerdasan yang sama antara komputer dan manusia dalam mengolah informasi. Manusia terbantu dalam pencarian informasi karena manusia dapat mengolah kata yang digunakan dalam pencarian informasi. Manusia memiliki pengetahuan tentang hubungan satu kata dengan kata lainnya, sedangkan komputer tidak dapat mengetahuinya karena komputer tidak mengetahui sense dari satu kata tersebut. Agar komputer memiliki kecerdasan yang sama, dibutuhkan pencarian nilai ke-samaan semantik (semantic similarity) antar kata. Berdasarkan ide tersebut, metode similarity yang dipilih untuk mencari nilai similarity antar kata Bahasa Indonesia adalah metode PMI_{max} yang merupakan turunan dari metode PMI. Metode PMI_{max} dipilih karena metode ini dapat menghasilkan nilai similarity berdasarkan kemuculan suatu kata di dalam suatu korpus. Metode ini juga menghasilkan nilai similarity yang baik saat diterapkan dalam Bahasa Inggris. Sehingga penelitian ini menguji apakah metode PMI_{max} dapat diterapkan dalam pencarian nilai similarity dalam Bahasa Indonesia, dan seberapa baik metode ini saat diterapkan. Dengan menggunakan korelasi Pearson hasil penelitian ini menunjukkan bahwa, metode PMI_{max} cukup baik diterapkan dalam mencari nilai similarity dalam kata – kata Bahasa Indonesia dibandingkan dengan metode PMI dan Word2Vec. Nilai korelasi yang dihasilkan, 0,26 pada Miller and Charles, 0,33 pada Simlex-999 dan 0,52 pada WordSim-353 Similarity.

Kata kunci : PMI_{max}, PMI, Kesamaan Semantik, Kesamaan Semantik Antar Kata