

Abstrak

Natural Language Processing (NLP) atau pemrosesan bahasa alami merupakan sebuah disiplin ilmu yang khusus mengolah teks yang ditulis langsung oleh manusia dan mempunyai sifat tidak terstruktur. Pengukuran *semantic similarity* antar kata merupakan salah satu tugas penerapan dari *Natural Language Processing* yang intinya adalah mencari skor *semantic similarity* antar kata. Skor tersebut menunjukkan seberapa erat tingkat kesamaan antar dua kata. Salah satu metode untuk menghitung *semantic similarity* adalah PMI (Pointwise Mutual Information).

Keterkaitan semantik merupakan salah satu jenis pengukuran yang ada pada *text mining* untuk menggambarkan bagaimana hubungan antara kata. Tujuan dari pengukuran keterkaitan semantik ini adalah untuk memperoleh nilai atau skor yang merepresentasikan seberapa besar keterkaitan antar dua kata tersebut. *Pointwise Mutual Information* (PMI) adalah salah satu pengukuran secara statistik untuk keterkaitan semantik yang telah banyak digunakan. *Pointwise Mutual Information* (PMI) merupakan varian yang digunakan dalam tugas akhir ini untuk menghitung keterkaitan semantik. Perhitungan nilai PMI dilakukan pada dataset yang didapat dari *wikipedia* Bahasa Indonesia. Nilai PMI yang didapatkan oleh aplikasi dihitung korelasinya. Hasil dari penelitian pada tugas akhir ini merupakan nilai korelasi antara skor yang dihasilkan sistem dengan *gold standard SimLex-999*, *WordSim-353* dan *Miller and Charles* yang menghasilkan nilai korelasi yang akan menunjukkan seberapa baik metode pengukuran PMI. Nilai korelasi tertinggi yaitu 0,256 dengan menggunakan *window size* = 25.

Kata Kunci: Semantik, *Text Mining*, *Pointwise Mutual Information*.