Abstrak

Natural Language Processing atau premrosesan bahasa alami merupakan sebuah disiplin ilmu yang khusus mengolah teks yang ditulis langsung oleh manusia yang bersifat tidak terstruktur. Pengukuran *semantic similarity* antar kata merupakan salah satu tugas penerapan dari *Natural Language Proccessing* 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_{max} (Pointwise Mutual $Information_{max}$). PMI_{max} mengestimasi korelasi maksimum antara dua kata dan korelasi antara makna terdekat kedua kata tersebut karena sebuah kata seringkali memiliki banyak makna atau bisa disebut dengan kata Polisemi.

Pada tugas akhir ini, diimplementasikan penghitungan *semantic similarity* antar kata menggunakan PMI_{max} dengan menggunakan estimasi dari kata polisemi. konteks kata bersumber dari dataset Brown Corpus dan dataset Gutenberg. Hasil dari keterkaitannya dibandingkan dengan dataset *Gold Standard* WordSim-353 *semantic relatedness, semantic similarity,* Miller Charles dan Simlex-999.

Hasil penelitian yang didapat terlihat bahwa dengan menggunakan PMI_{max} didapatkan korelasi terbaik yaitu 66,5% dengan dataset *gold standard* WordSim-353 *semantic similarity* menggunakan korelasi Pearson dan dengan menggunakan nilai sense hasil analisis variabel p dan q. Nilai *semantic similarity* setiap pasang kata sangat dipengaruhi oleh nilai *Co-Occurence* sepasang kata teresebut, semakin tinggi nilai *Co-Occurence* suatu pasangan maka akan menghasilkan skor *semantic similarity* yang tinggi.

Kata Kunci: Kesamaan semantik, Pointwise Mutual Information, kata polisemi.