

Abstrak

Implementasi perhitungan kesamaan semantik antar kata merupakan salah satu tugas yang dapat diselesaikan dalam bidang Natural Language Processing(NLP). Perhitungan kesamaan semantik antar kata dapat digunakan untuk membantu mesin dalam memahami bahasa manusia. Selain itu, perhitungan kesamaan semantik juga dapat digunakan sebagai dasar penelitian tahap selanjutnya pada bidang NLP. Penelitian ini dilatar belakangi oleh suatu masalah dimana pada saat ini pencarian sistem informasi banyak melibatkan teks atau dokumen, namun mesin belum dapat menyamakan persepsi manusia dengan baik sehingga mesin perlu dibantu untuk memahami teks atau dokumen tersebut. Sepasang kata dinyatakan mempunyai kesamaan semantik apabila memiliki kesamaan pada makna atau konsep. Pada penelitian ini, dilakukan implementasi perhitungan kesamaan semantik antar kata untuk bahasa Inggris. Korpus yang digunakan pada penelitian ini yaitu Brown Corpus, Berita Corpus, dan Harry Potter Corpus. Dokumen tersebut diubah kedalam bentuk vektor dengan Word2vec. Selanjutnya nilai kesamaan semantik yang dihasilkan dari vektor tersebut dibandingkan dengan dataset Gold Standard SimLex999 untuk mengukur nilai korelasinya. Hasil pengujian menunjukkan bahwa pengukuran Word2vec menghasilkan korelasi sebesar 0.192 dengan perhitungan korelasi Pearson.

Kata kunci : Kesamaan Semantik, Natural Language Processing, Word2vec