

Abstraksi

Penelitian ini mengkaji perhitungan kesamaan antara kata-kata dalam bahasa Inggris menggunakan teknik representasi kata. Word2Vec adalah model yang digunakan dalam makalah ini untuk mewakili kata-kata ke dalam bentuk vektor. Model dalam penelitian ini dibentuk menggunakan 320.000 artikel dari Wikipedia bahasa Inggris sebagai korpus dan kemudian metode perhitungan Cosine Similarity digunakan untuk menentukan nilai kesamaan semantik. Model ini kemudian diuji dengan test set gold standard WordSim-353 sebanyak 353 pasang kata dan SimLex-999 sebanyak 999 pasang kata, yang telah diberi label dengan nilai kesamaan menurut penilaian manusia. Pearson Correlation digunakan untuk mengetahui akurasi korelasi antara hasil sistem dan gold standard. Hasil korelasi dari penelitian ini adalah 0,665 untuk WordSim-353 dan 0,284 untuk SimLex-999 menggunakan konfigurasi Windows Size 9 dan 300 dimensi vektor.