

Abstrak

Implementasi pengukuran kesamaan semantik memiliki peran yang sangat penting dalam beberapa bidang *Natural Language Processing* (NLP), dimana hasilnya seringkali dijadikan dasar dalam melakukan *task* NLP yang lebih lanjut. Salah satu penerapannya yaitu dengan melakukan pengukuran kesamaan semantik multibahasa antar kata. Pengukuran ini dilatarbelakangi oleh suatu masalah dimana saat ini banyak sistem pencarian informasi yang harus berurusan dengan teks atau dokumen multibahasa. Saat ini, sejauh yang penulis ketahui, masih kurangnya sistem komputer yang dapat mengukur nilai kesamaan semantik sepasang kata dalam lintas bahasa. Sepasang kata dinyatakan memiliki kesamaan semantik jika pasangan kata tersebut memiliki kesamaan dari sisi makna atau konsep. Terdapat banyak metode yang dapat digunakan dalam mengukur kesamaan semantik antar kata, salah satunya adalah pengukuran secara statistik menggunakan metode *Pointwise Mutual Information* (PMI). Pada penelitian ini, diimplementasikan perhitungan kesamaan semantik antar kata pada bahasa yang berbeda yaitu bahasa Inggris dan bahasa Spanyol. Korpus yang digunakan pada penelitian ini yakni *Europarl Parallel Corpus* pada bahasa Inggris dan bahasa Spanyol. Konteks kata bersumber dari *Swadesh list*, serta hasil dari kesamaan semantiknya dibandingkan dengan dataset *Gold Standard SemEval 2017 Crosslingual* untuk diukur nilai korelasinya. Hasil pengujian yang didapat terlihat bahwa pengukuran metode PMI mampu menghasilkan korelasi sebesar 0.577 untuk rata-rata harmonik antara korelasi Pearson dan korelasi Spearman. Nilai korelasi sistem tersebut lebih tinggi dibandingkan dengan nilai korelasi yang dihasilkan oleh tim Rufino yang juga menggunakan metode PMI untuk pengukuran *cross-lingual semantic similarity* pada dataset *semEval 2017*. Nilai korelasi yang dihasilkan oleh tim Rufino yaitu 0.340.

Kata Kunci: Kesamaan Semantik, *Crosslingual Semantic Similarity*, *Pointwise Mutual Information*.