

Abstrak

Pengukuran *semantic similarity* menggunakan *monolingual alignment* memiliki kendala pada pengimplementasian sistem yaitu masih mejadi penelitian bagaimana memodifikasi *feature extraction* yang digunakan untuk melakukan *align* pada kata dari kalimat. Sistem yang telah ada meneliti *feature extraction* untuk diterapkan pada *alignment*. *Semantic similarity* adalah salah satu pengukuran dalam *text mining* yang mencari representasi hubungan antar kata dalam kalimat.

Pada penelitian ini metode yang digunakan adalah *monolingual alignment* dengan cara memodifikasi *feature extraction* yang berasal dari beberapa penelitian sebelumnya. Modifikasi *feature extraction* yang difokuskan pada penelitian ini adalah *identical word*, *named entities recognition*, *paraphrase database*, *part-of-speech*, *synonym*, *hyponym*, *hyponym* dan *positional*. Sistem yang dibangun pada penelitian ini menggunakan teknik *supervised* dan *unsupervised data mining*. Pengujian dilakukan dengan cara *extrinsic evaluation* dengan membandingkan skor dari sistem dengan skor yang telah dibuat oleh para ahli. Teknik membandingkan skor dari sistem dengan skor *gold standard* menggunakan korelasi pearson. Hasil pengujian penelitian ini dengan menggunakan teknik *unsupervised* mendapat korelasi 0.7077 sedangkan *supervised* mendapatkan korelasi 0.7228.

Kata kunci : *semantic similarity, monolingual alignment, extrinsic evaluation, identical word, named entities recognition, positional.*