Abstrak

Tekxtual Similarity/Relatedness bisa jadi merupakan landasan yang penting dalam melakukan keterhubungan antar teks dalam ilmu-ilmu teks mining lainnya, seperti sentimen analisis, teks kategorisasi, dll. Keterhubungan teks tersebut dapat berupa keterkaitan yang ditentukan apakah teks tersebut memiliki keterkaitan secara *lexical* atau *semantic*. Sehingga penelitian ini dilatar belakangi dengan adanya masalah belum tersedianya komputer yang dapat menyamai persepsi manusia terkait penilaian keterkaitan antar kata untuk mempermudah dalam penelitian-penelitian teks lainnya. Keterkaitan dari inputan teks tersebut dapat dihitung dengan menggunakan metode *Explicit Semantic Analysis (ESA).* Metode ini menghitung keterkaitan antar dua kata yang akan dirubah menjadi bentuk skor dengan rentang antara 0-1 sesuai tingkat keterkaitannya. Database yang digunakan dalam perhitungan adalah artikel *wikipedia* dengan total sebanyak 2000 judul, digunakan karena dianggap merupakan sumber *knowledge* terbesar di internet. Pengujian dilakukan dengan membandingkan perhitungan *tf.idf ESA* secara biner dan nonbiner. Serta untuk mengetahui pengaruh jumlah judul artikel terhadap skor, pengujian juga dibedakan berdasarkan jumlah artikelnya sebanyak 500, 1000 dan 2000 judul. Setelah masing-masing skor keterkaitan dari setiap pengujian didapatkan, kemudian skor tersebut dibandingkan dan dihitung nilai korelasinya dengan *gold standard wordsim353 relatedness* dan *gold standard MEN*. Korelasi yang dihasilkan untuk pengujian 2000 judul artikel pada perhitungan biner sebesar 0,2789 dan nonbiner sebesar 0,3958 untuk *gold standard wordsim353 relatedness*. Sedangkan untuk *gold standard MEN,* korelasi yang dihasilkan untuk pengujian 2000 judul artikel pada perhitungan biner sebesar 0,4249 dan nonbiner sebesar 0,4998.

Kata kunci : *Tekxtual Similarity/Relatedness*, *semantic*, *Explicit Semantic Analysis (ESA)*, skor, *tf.idf*, kata, *wikipedia, gold standard*, korelasi, *MEN, wordsim353 relatedness*.