

## ABSTRAK

Abdul Baquee Muhammad [1] telah membangun Corpus yang berisi domain AlQur'an, WordNet dan kamus. Dia telah melakukan inisialisasi penelitian dalam pengembangan pengetahuan tentang AlQur'an dan pengetahuan tentang keterkaitan antar teks di AlQur'an. Metode pengukuran berbasis jalur yang diusulkan oleh Liu, Zhou dan Zheng [3] belum pernah digunakan dalam domain AlQur'an. Dengan menggunakan dataset terjemahan AlQur'an dalam penelitian ini, metode pengukuran berbasis jalur yang diusulkan oleh Liu, Zhou dan Zheng [3] akan digunakan untuk menguji metode ini di domain AlQur'an untuk mendapatkan nilai kesamaan dan untuk mengukur nilai korelasinya.

Dalam penelitian yang dilakukan oleh Liu, Zhou dan Zheng [3] pada kesamaan semantik dengan menggunakan metode berbasis jalur berhasil mendapatkan nilai korelasi sebesar 92.6%. Nilai korelasi masih memiliki kesempatan untuk ditingkatkan karena pada penelitian oleh Liu, Zhou dan Zheng [3] hanya mengoptimalkan hubungan semantik hypernym dan hyponym. Pada hubungan semantik terdapat banyak hubungan semantik selain hypernym dan hyponym seperti sinonym, antonym, meronym dan holonym. Dengan memanfaatkan semua hubungan semantik tersebut diharapkan dapat meningkatkan nilai korelasi yang telah dicapai sebelumnya.

Untuk mendapatkan nilai korelasi yang lebih baik, nilai derajat diusulkan untuk digunakan dalam memodifikasi metode berbasis jalur yang diusulkan oleh Liu, Zhou dan Zheng [3]. Nilai Derajat adalah jumlah link yang dimiliki oleh node lcs (lower common subsumer) pada taksonomi. Link yang dimiliki oleh node pada taksonomi mewakili hubungan semantik yang dimiliki node dalam taksonomi. Dengan menggunakan nilai derajat untuk memodifikasi metode berbasis jalur yang diusulkan oleh Liu, Zhou dan Zheng [3] diharapkan nilai korelasi yang diperoleh dapat meningkat.

Setelah menjalankan beberapa percobaan dengan menggunakan metode yang diusulkan, pengukuran korelasi memperoleh nilai korelasi yang rendah untuk semua POS karena banyak kosakata yang ada di SimLex-999 tidak terdapat dalam kosa kata WordNet sehingga banyak pasangan kata dari SimLex-999 yang nilai kemiripannya adalah Nol dan hal itu membuat nilai korelasinya menjadi rendah.

**Kata kunci** : data mining, semantik, keterkaitan teks semantik, kesamaan semantik, persamaan berbasis jalur, panjang jalan terpendek, kedalaman subsumer, nilai derajat.