

Abstrak

Alquran adalah pedoman hidup umat Islam yang sudah memiliki banyak tafsiran agar mudah untuk dipahami. Dari sekian banyak kalimat pada Alquran memungkinkan terdapat beberapa ayat yang memiliki kesamaan arti. Untuk mempermudah pencarian kesamaan ayat maka dilakukanlah penelitian pada teks terjemahan Alquran. Banyak metode dalam bidang *text mining* dan NLP yang dapat digunakan untuk mengukur kesamaan semantik antar kalimat, beberapa diantara yaitu menggunakan *alignment* dan vektor semantik. *Alignment* adalah salah satu metode yang dapat menghitung kesamaan semantik dengan melakukan penyejajaran kata-kata. Fitur *alignment* yang digunakan dalam penelitian ini yaitu *identical word*, *PPDB*, *word sequences*, dan *named entities*. Sementara vektor semantik adalah metode yang digunakan untuk menghitung kemiripan sebuah kata dari distribusi kata-kata di sekitarnya. Untuk penggunaan vektor semantik dibutuhkan *corpus text8* diambil dari Matt Mahoney. Agar memperoleh hasil yang lebih baik maka dilakukan kombinasi dari metode *alignment* dengan vektor semantik menggunakan *ridge regresi*. Selain itu, dilakukan juga pembuatan *gold standard* sebagai tolak ukur untuk mengetahui nilai korelasi yang menjadi kontribusi penulis untuk menambahkan data yang sudah ada pada penelitian tahun lalu. Kombinasi metode *alignment* dan vektor semantik untuk menghitung kesamaan semantik pada data terjemahan Alquran tahun 2016 menghasilkan nilai korelasi lebih tinggi 0,002029 dari Dwi Jayanti yaitu 0,90467, sementara untuk data terjemahan Alquran yang dikumpulkan pada tahun 2017 menghasilkan nilai korelasi 0,79756 lebih rendah 0,00593 dibandingkan dengan sistem yang dibangun oleh Md. Sultan.

Kata kunci: *text mining*, *nlp*, *word alignment*, *semantik*, *vektor semantik*, *quran*, *islam*, *regresi*.