

ABSTRAK

Dalam Al-Qur'an, sangat banyak kata yang dapat digunakan untuk diteliti khususnya dalam hal *Natural Language Processing (NLP)*, salah satunya yaitu Model Semantik Distribusi untuk kesamaan kata dan keterkaitan seperti pada penelitian ini. Penelitian sebelumnya, yang menggunakan dataset Al-Qur'an hanya dilakukan kesamaan semantik untuk surah tertentu, evaluasi dataset kesamaan dan keterkaitan semantik namun penerapan pada bahasa Turki, dan penerapan evaluasi dataset untuk kata kerja saja. Tujuan dari penelitian ini adalah untuk memberikan wawasan kepada peneliti model semantik dengan mengevaluasi model dalam beberapa atribut. Menyeimbangkan pasang kata dataset kosakata Al-Qur'an untuk kelas kata benda dan kata kerja dengan frekuensi manusia untuk mengevaluasi kekokohan model semantik mengenai masalah kemunculan kata langka. Pasangan kata dipilih untuk mengaktifkan evaluasi distribusi model semantik dengan beberapa atribut kata dan hubungan pasang kata seperti frekuensi kemunculan kata, konkret dan jenis hubungan (misalnya, sinonim dan antonim). Dataset ini terdiri dari 500 pasang kata dan dibantu oleh 15 responden, yang setiap pasangan memiliki dua nilai yang berbeda untuk kesamaan dan keterkaitan. Metode yang digunakan adalah Sim-Rel vektor, kuesioner dan perhitungan *gold standard*, hingga hasil perhitungan performansi menggunakan korelasi *Spearman Rank* sebesar 0,909. Sumbu vektor Sim-Rel mendapatkan hasil dengan 4 daerah yaitu SU = 23 pasang kata, SR = 77, DU = 192, dan DR = 208.

Kata Kunci : *Natural Language Processing*, Model Semantik Distribusi, *Sim-Rel Vector*, *Spearman Rank*.