
Abstrak

Untuk mengukur korelasi semantik antar kata ada banyak metode yang dapat digunakan, salah satunya adalah Word Alignment. Word alignment adalah metode yang melakukan penyejajaran kata-kata yang memiliki korelasi secara huruf atau korelasi secara makna di antara dua kalimat. Penelitian ini berfokus menggunakan Word Alignment pada terjemahan ayat Al-Quran. Metode ini dikembangkan untuk melakukan alignment terhadap data pasangan kalimat tetapi dapat digunakan untuk mengukur korelasi antar ayat. Dengan menggunakan algoritma Back to Basic Word Alignment yang dikembangkan oleh Sultan et al. peneliti mengembangkan lagi untuk penelitian alignment antar ayat dalam Al-Quran, untuk mengetahui pengaruhnya apabila digunakan dalam translasi Al-Quran sebagai datasetnya. Dataset Al-Quran yang digunakan akan diubah kedalam format dataset MSR-RTE[1] oleh peneliti, dengan tujuan memberikan hasil penelitian yang baru dalam konteks word alignment Al-Quran. Dalam Back to Basic Word Alignment ada yang disebut alignment pipeline yang berisi penggunaan urutan feature-feature, feature yang digunakan dalam penelitian ini yaitu align identical word, align PPDB, align word sequences, align named entities, align content words (dependency), align content words using surrounding words (text neighbor), align stop words, dan align PPDB Extended. features tersebut akan dikombinasikan untuk menentukan nilai korelasi antar dua ayat Al-Quran(F1 score). Nilai korelasi terbaik antar ayat yang dapat dihasilkan dalam penelitian adalah 51,02% dibandingkan dengan baseline penelitian oleh Sultan et al. yaitu 91,7%. Nilai korelasi antar ayat pada penelitian ini dapat disimpulkan sebagai nilai yang cukup, dan masih dapat ditingkatkan lagi melalui penambahan feature, knowledge base, maupun menggunakan kombinasi penerjemah Al-Quran yang berbeda.

Kata kunci : Back to Basic Word Alignment, Monolingual Alignment, MSR-RTE, Al-Quran, Text mining.
