

ABSTRAK

Abstract in Bahasa Indonesia.

Memahami makna teks secara keseluruhan atau sebagian yang ditulis dalam dokumen merupakan kegiatan yang sangat penting dalam proses memahami dokumen secara keseluruhan. Kesalahan memahami makna teks yang merupakan bagian dari dokumen, terutama teks menjadi kata yang penting akan membuat salah dalam membuat persepsi keseluruhan dokumen. Kitab Suci Al-Quran adalah dokumen penting dalam kehidupan umat Islam di dunia. Al-Quran terdiri dari 114 surat, yang kemudian dibagi menjadi ayat. Karena tidak berbahasa Arab, Muslim Indonesia praktis sangat bergantung pada penerjemahan dokumen Quran dalam Bahasa Indonesia sebagai sarana untuk memahami makna dari isi dokumen dalam Quran itu sendiri.

Dalam konteks Al-Quran dalam Bahasa Arab, diketahui bahwa satu kata dalam Bahasa Arab dapat diterjemahkan menjadi dua kata atau lebih di Indonesia. Dalam hal ini kita harus mengurai kata-kata dalam dokumen dan menggabungkan beberapa kata menjadi satu kata awalnya dalam Bahasa Arab. Proses dekomposisi dan penentuan kata-kata di segmen ini disebut segmentasi frase.

Riset kami mengusulkan untuk melakukan proses segmentasi berbasis phrase untuk dokumen terjemahan quran Bahasa Indonesia, dalam hal ini surah Yasin, dengan pendekatan metoda monolingual alignment dan parsing, dimana dilakukan proses tokenization, word similarity dengan referensi quran phrase database, pos tagging dan parsing sehingga menghasilkan tingkat kesesuaian yang cukup apabila dibandingkan dengan gold standard.

Kata kunci: segmentasi berbasis frase, monolingual alignment