

Abstrak

Salah satu cara untuk memahami Alquran adalah dengan melakukan penafsiran yang benar dan tidak menyimpang, yaitu dengan memperhatikan redaksi kata yang digunakan dalam merangkai ayat-ayat Alquran. Dengan cara tersebut, maka dapat ditemukan ayat lainnya yang memiliki rangkaian kata yang menyerupai. Salah satu cara untuk mencari ayat yang sama berdasarkan rangkaian kata tersebut adalah dengan menggunakan pendekatan *Longest Common Subsequence* (LCS) yang dapat mencari rangkaian kata terpanjang bersama dari suatu pasangan teks. Hasil dari pencarian ayat yang sama ini kemudian dikumpulkan hingga menjadi korpus yang diharapkan dapat membantu umat manusia dalam menafsirkan Alquran. Penelitian ini menghasilkan suatu sistem yang dapat mencari ayat yang sama menggunakan pendekatan LCS, kemudian hasilnya akan dikumpulkan menjadi suatu korpus berdasarkan hasil LCS. Hasil rata-rata yang diperoleh dari beberapa pengujian yang telah dilakukan adalah Data Arab mendapatkan nilai *precision* adalah 46.84%, nilai *recall* adalah 96.13%, dan nilai *f1-score* adalah 62.96%. Sedangkan untuk Data Indonesia mendapatkan nilai *precision* adalah 40.57%, nilai *recall* adalah 97.56%, dan nilai *f1-score* adalah 57.04%.

Kata Kunci: Alquran, Korpus, *Longest Common Subsequence*, Rangkaian Kata.