

## I. PENDAHULUAN

**A**LQURAN adalah suatu kitab mukjizat melalui ungkapannya yang demikian indah memukau, terdiri dari huruf-huruf seperti *alif- lam- ra* dan berfungsi sebagai kitab suci yang diturunkan atas izin Allah SWT kepada Nabi Muhammad SAW untuk menjadi pedoman hidup bagi umat manusia yang mengeluarkan mereka dari gelapnya kekufuran menuju cahaya keislaman yang terang benderang [15] [10]. Untuk menjadikannya sebagai pedoman hidup, Alquran harus dipahami dengan melakukan penafsiran secara benar dan tidak menyimpang. Salah satu caranya adalah dengan memperhatikan redaksi kata, yaitu struktur kata yang digunakan dalam merangkai atau menyusun ayat-ayat pada Alquran [5]. Dengan melihat redaksi atau rangkaian kata tersebut maka dapat ditemukan ayat lainnya yang memiliki rangkaian kata yang menyerupai. Contohnya seperti pada Quran Surah *al-Fatihah* ayat 1 (QS:1:1) yang artinya *Dengan menyebut nama Allah yang Maha Pemurah lagi Maha Penyayang* dengan Quran Surah *al-Fatihah* ayat 3 (QS:1:3) yang artinya *Maha Pemurah lagi Maha Penyayang*.

Untuk membangun suatu korpus dari rangkaian kata yang berulang pada Alquran ini haruslah dilakukan pencarian ayat yang memiliki struktur kata yang sama pada ayat lainnya dan selanjutnya hasil tersebut dikumpulkan. Pencarian ayat pada Alquran telah banyak dilakukan oleh beberapa penelitian sebelumnya dengan menggunakan metode ataupun pendekatan yang beragam, dan memiliki kesamaan dengan penelitian kali ini [2] [9] [12]. Hasil yang diperoleh dari beberapa penelitian tersebut cukup baik namun hanya menyajikan hasil pencarian ayat berupa tema ayat, nomor ayat, dan nama surah berdasarkan masukan dari pengguna. Berbeda dari penelitian sebelumnya, penelitian kali ini adalah melakukan pencarian ayat yang sama dengan ayat lainnya sekaligus menghimpun atau mengelompokkan hasil pencarian tersebut menjadi sebuah korpus yang berisi himpunan ayat yang sama berdasarkan rangkaian kata terpanjang yang digunakan dalam menyusun ayat tersebut.

Berdasarkan permasalahan di atas, dibuatlah suatu sistem yang dapat mencari ayat yang memiliki redaksi kata yang sama dengan ayat lainnya menggunakan pendekatan *Longest Common Subsequence* (LCS), yaitu suatu metode untuk mencari rangkaian data terpanjang bersama dari suatu kumpulan data [7] [6], dalam hal ini data tersebut adalah ayat Alquran. Data yang digunakan pada penelitian ini adalah seluruh ayat pada Alquran dan terjemahan Alquran berbahasa Indonesia yang terlebih dahulu dilakukan *Text Preprocessing* berupa penggunaan bentuk teks *lowercase*, penghapusan *stopword*, dan penghapusan semua karakter yang tidak dibutuhkan. Dari hasil tersebut selanjutnya dikumpulkan menjadi suatu korpus berupa himpunan ayat berdasarkan rangkaian kata terpanjang bersama. Data akhir tersebut dievaluasi dengan cara dibandingkan dengan data *gold-standard* yang dibuat oleh penulis dengan ketentuan bahwa ayat yang diprediksi sama dengan ayat lainnya adalah ayat yang memiliki rangkaian kata terpanjang dengan jumlah rangkaian kata lebih dari atau sama dengan selisih antara jumlah kata pada ayat pembandingan dan jumlah rangkaian kata terpanjang tersebut. Hal tersebut dikarenakan penulis berasumsi bahwa jumlah rangkaian kata terpanjang yang baik adalah yang jumlah katanya sama dengan atau lebih dominan dibandingkan sisa kata dari teks tersebut. Evaluasi yang dilakukan adalah dengan menghitung *precision*, *recall*, dan *f1-score*. Korpus ini diharapkan dapat membantu umat manusia dalam menafsirkan Alquran dengan cara yang benar dan tidak menyimpang.