

1. Pendahuluan

Latar Belakang

Al-Qur'an memiliki kedudukan yang penting serta memiliki cakupan pembahasan yang luas sebagai penjelas segala sesuatu dan petunjuk bagi seluruh umat Islam. Al-Qur'an pertama kali diwahyukan kepada nabi Muhammad SAW dalam bahasa Arab, sehingga masih banyak orang yang kurang memahami makna keseluruhan dari Al-Qur'an tersebut. Perkembangan zaman dan teknologi membuat banyak pakar agama di dunia membuat dokumen terjemahan Al-Qur'an ke dalam berbagai bahasa dan mempermudah seluruh umat Islam untuk memahami isi Al-Qur'an. Namun jika melakukan pencarian suatu kata atau makna dari dokumen terjemahan Al-Qur'an yang sudah ada secara manual ini akan sulit dan membutuhkan waktu yang lama [1]. Salah satu cara untuk mengatasi permasalahan pencarian kata dalam terjemahan Al-Qur'an yaitu dengan menggunakan metode konkordansi.

Konkordansi merupakan salah satu metode yang digunakan untuk menampilkan daftar kata-kata (keyword) yang diambil dari sejumlah teks atau dokumen. Pencarian dimulai dengan memasukkan keyword kemudian sistem konkordansi akan menampilkannya dalam bentuk list tanpa menghilangkan keseluruhan kalimat –ayat Al-Qur'an.

Hasil pencarian menggunakan metode konkordansi akan menghasilkan lebih banyak kata dibandingkan dengan pencarian yang biasa [2]. Sejauh ini, aplikasi yang sudah ada hanya melakukan pencarian dari kata yang telah disediakan sistem dan menelusuri pencarian pada terjemahan perkata dalam suatu ayat Al-Quran secara tidak menyeluruh. Oleh karena itu sistem pencarian dikombinasikan dengan Algoritma Longest Common Subsequence (LCS) untuk menghitung similarity diantara query dan terjemahan ayat secara menyeluruh dan mengurutkan ayat-ayat yang mirip dengan query. Konkordansi yang diterapkan pada pencarian kata dan frasa terjemahan Bahasa Indonesia ini terdiri dari konkordansi berdasarkan lema (kata dasar), kata/frasa, dan konkordansi tambahan sinonim. Berdasarkan hasil pengujian sistem pencarian menggunakan algoritma Longest Common Subsequence ini dapat memberikan hasil yang cukup relevan.

Longest Common Subsequence (LCS) merupakan salah satu algoritma yang biasa digunakan untuk mengukur kemiripan kata (similarity) pada rangkaian DNA dan file management [3]. Pada kasus ini Algoritma LCS diterapkan untuk menghitung nilai similarity diantara query yang ada dengan terjemahan ayat Al-Quran berbahasa Indonesia. Sistem pencarian memproses query yang kemudian akan menampilkan hasil pencarian yang telah diurutkan berdasarkan nilai similarity-nya. Pengembangan aplikasi serupa sudah dilakukan oleh Kementerian Agama (<https://quran.kemenag.go.id>). Akan tetapi aplikasi tersebut tidak dapat melakukan pencarian pada terjemahan ayat Al-Quran secara menyeluruh, melainkan hanya dapat melakukan pencarian berdasarkan surat dan ayat dengan format yang sudah ditentukan. Pada aplikasi tersebut kata-kata yang akan dicari telah disediakan sistem, sehingga pengguna tidak dapat memasukan query pencarian secara bebas [4].

Konkordansi pencarian kata dan frasa terjemahan ayat Al-Quran bahasa Indonesia pada sistem yang dibuat dibagi menjadi tiga bagian, yaitu berdasarkan kata dasar (lema), kata/frasa dan konkordansi tambahan sinonim. Hasil pencarian yang di tampilkan berupa daftar potongan ayat dan terjemahan Al-Quran yang relevan dengan query yang diurutkan.

Topik dan Batasannya

Konkordansi bertujuan untuk menampilkan konteks yang saling terkait dari suatu kata yang terdapat di dalam korpus, sehingga pengguna dapat memahami makna yang terdapat dalam konteks kata tersebut. Tampilan konkordansi menunjukkan setiap kemunculan *query* tertentu, bersama dengan beberapa konteks dari *query* tersebut. Namun, hasil pencarian yang ditampilkan sama persis dengan *query* yang dicari.

Agar menampilkan hasil pencarian yang mirip dari *string* yang digunakan, LCS perlu diterapkan bersamaan dengan konkordansi untuk menghitung nilai *similarity* diantara teks. LCS membentuk suatu *string* dari beberapa *string* dengan menghapus beberapa karakter dengan memperhatikan urutan karakternya [5]. Dalam terjemahan Al-Quran, banyak kasus dimana terdapat kata atau frasa yang mirip dari *string* yang digunakan dan maknanya pun mirip. LCS dapat membantu sistem pencarian dalam menemukan terjemahan ayat yang memiliki kemiripan, sehingga hasil pencarian yang ditampilkan tidak 100% sama dengan *query*, namun masih memiliki kemiripan *string*.

Adapun batasan masalah dalam Tugas Akhir ini adalah sebagai berikut.

1. Algoritma LCS merupakan salah satu algoritma untuk menghitung *string similarity*, sehingga penerapannya masih sederhana dan belum bisa menentukan apakah kedua *string* memiliki makna yang mirip.
2. Pencarian kata berdasarkan sinonim pada terjemahan sistem yang dibuat masih dalam tahap awal, sehingga fungsi pencarian kata berdasarkan sinonim belum dapat berfungsi sesuai yang diharapkan.

Tujuan

Pembangunan sistem pencarian kata dan frasa ini ditujukan agar sistem dapat menampilkan hasil pencarian yang relevan dengan query, serta makna yang mirip berdasarkan nilai similarity pada string yang digunakan. Sehingga akan diperoleh hasil terjemahan ayat Al-Quran yang telah diurutkan berdasarkan nilai similarity nya. Oleh karena itu diperlukan implementasi konkordansi dan algoritma Longest Common Subsequence pada sistem pencarian kata dan frasa.

Sistem pencarian kata dan frasa yang dibuat akan menampilkan hasil performansi terhadap pencarian yang dilakukan oleh sistem dengan menggunakan konkordansi dan algoritma Longest Common Subsequence yang diukur dengan precision, recall, dan F1-Measure. Selain itu perhitungan performansi sistem dilakukan proses perangkaian rangking yang dilakukan untuk menampilkan hasil pengurutan pencarian ayat berdasarkan nilai similarity yang dihitung dengan menggunakan average precision (AP) untuk tiap query, dan mean average precision (MAP) untuk kumpulan query.

Hal ini perlu dilakukan untuk mengetahui apakah performansi sistem konkordansi yang dikombinasikan dengan algoritma Longest Common Subsequence dalam menampilkan hasil pencarian sudah bagus dan sesuai dengan gold standar yang dibuat.

Untuk memaksimalkan nilai performansi yang dihasilkan, penerapan Range perlu dilakukan agar sistem hanya menampilkan hasil pencarian yang mirip. Range merupakan rentang pada nilai similarity yang di terima sistem untuk ditampilkan sebagai hasil pencarian.

Organisasi Tulisan

Organisasi penulisan Tugas Akhir ini dibagi menjadi beberapa bagian. Bagian 2 menjelaskan studi-studi yang terkait dengan Tugas Akhir ini, yaitu penjelasan konsep konkordansi, algoritma LCS, pra proses, dan perhitungan performansi yang digunakan. Bagian 3 menjelaskan bagaimana sistem pencarian kata atau frasa akan dikombinasikan dengan LCS untuk menampilkan hasil pencarian yang relevan dan bagaimana skema pengujian yang akan dilakukan. Bagian 4, evaluasi yang dilakukan pada sistem pencarian, yang terdiri dari hasil pengujian dan analisis hasil pengujian. Bagian 5 adalah kesimpulan dan saran pada Tugas Akhir ini.