

# BAB I

## PENDAHULUAN

### 1. Pendahuluan

#### 1.1 Latar Belakang

Al-Qur'an merupakan kitab suci yang dijadikan sebagai pedoman hidup umat Islam untuk dipahami maknanya dan diamalkan dalam kehidupan. Al-Qur'an terdiri dari 114 surat, 6236 ayat, dan 77.845 kata [1]. Kitab ini diturunkan kepada Nabi Muhammad dalam bahasa Arab. Sekarang ini terdapat Al-Qur'an yang dilengkapi dengan transliterasi dari aksara Arab ke aksara Latin sebagai alternatif cara pembacaannya.

Banyaknya jumlah surat, ayat, dan kata dalam Al-Qur'an akan menyulitkan pencarian kata secara manual. Sistem pencarian ayat Al-Qur'an sebenarnya sudah tersedia, hanya saja kata kunci yang dimasukan adalah kata kunci dalam aksara Arab. Hal ini menyulitkan pengguna yang kurang mahir dalam berbahasa maupun menulis aksara Arab. Solusi untuk menangani masalah tersebut adalah dengan dibuatnya sistem pencarian ayat Al-Qur'an yang memanfaatkan transliterasi latin sebagai kata kunci pencarian. Akan menjadi kendala jika sistem pencarian terbatas hanya dapat menangani kata kunci sesuai dengan pedoman penulisan transliterasi yang digunakan. Sebagai contoh, (1) jika pengguna salah mengeja potongan ayat dan tidak sesuai dengan penulisan transliterasi, maka hasil pencarian tidak ditemukan. (2) Jika pengguna memasukkan potongan ayat berdasarkan kemiripan pengucapan, maka hasil pencarian juga tidak ditemukan. IslamiCity [2] merupakan contoh penerapan sistem pencarian berdasarkan pengucapan dengan *query* berupa transliterasi aksara Arab-Latin. Namun, masih terdapat beberapa variasi penulisan *query* berdasarkan pengucapan yang belum bisa tertangani. Contohnya pada *query* "angfusihina" yang tidak dapat ditemukan hasilnya. Oleh karena itu dalam penelitian ini dilakukan pengembangan terhadap sistem pencarian ayat Al-Qur'an yang dapat menangani keberagaman variasi kata kunci dengan memanfaatkan kemiripan fonetis atau pengucapannya.

Pencocokkan fonetis (*Phonetic String Matching*) merupakan salah satu cara yang dapat digunakan untuk pencarian *string* berdasarkan kemiripan ucapan [3]. Terdapat beberapa algoritma yang dapat digunakan pada metode *Phonetic String Matching*, yaitu algoritma Soundex, Metaphone, Caverphone, dan Double Metaphone. Pada penelitian ini, sistem dibangun dengan menggunakan algoritma Double Metaphone. Algoritma Double Metaphone merupakan algoritma yang mendukung banyak ejaan bahasa yang berbeda serta dapat dimodifikasi untuk menyesuaikan dengan ejaan bahasa yang digunakan [4]. Beberapa contohnya adalah penggunaan Double Metaphone untuk pengkodean nama dalam bahasa Bangladesh [5]. Serta penggunaan Double Metaphone untuk pengkodean dalam bahasa Ethiopia [6]. Karena algoritma Double Metaphone yang ada saat ini masih menggunakan ejaan dalam bahasa Inggris, maka diperlukan beberapa modifikasi terhadap algoritma tersebut agar dapat menangani ejaan dalam bahasa Arab. Algoritma Double Metaphone digunakan untuk melakukan pengkodean fonetis terhadap *query* inputan yang berupa transliterasi latin potongan ayat Al-Qur'an serta pada dataset transliterasi latin Al-Qur'an yang digunakan. Algoritma ini akan menghasilkan *string* yang memiliki kemiripan secara fonetis atau ucapan.

Selain itu, diperlukan pula suatu proses untuk melakukan pencocokkan antara kode fonetis *query* dengan kode fonetis dataset yang telah dihasilkan. Terdapat beberapa algoritma pencocokkan string dalam metode *approximate string matching*, diantaranya yaitu algoritma Jaro Winkler, Levenshtein, Hamming, dan Damerau Levenshtein. Dalam penelitian ini, digunakan algoritma Jaro Winkler dalam proses pencocokkan kode fonetis serta pengukuran nilai kesamaan *string query* dan dataset. Jaro Winkler merupakan modifikasi dari algoritma Jaro dan merupakan algoritma pencocokkan *string* pendek dengan hasil yang baik [7]. Untuk mengetahui seberapa baik akurasi sistem yang telah dibangun, dilakukan pengukuran nilai *mean average precision*, *recall*, dan korelasi terhadap hasil pencarian sistem. Hasil pengukuran nilai *mean average precision* dan *recall* sistem kemudian dibandingkan dengan hasil dari beberapa metode lainnya.

Oleh karena itu pada penilitain tugas akhir ini dilakukan pengembangan sistem pencarian ayat Al-Qur'an berdasarkan kemiripan fonetis dengan menggunakan algoritma Double Metaphone dan Jaro winkler.

#### 1.2 Perumusan Masalah

1. Bagaimana mengaplikasikan dan memodifikasi algoritma Double Metaphone dalam pengkodean fonetis serta algoritma Jaro Winkler dalam pencocokkan kode dan pengukuran kesamaan *string* untuk membangun sistem pencarian ayat Al-Qur'an berdasarkan kemiripan fonetis?
2. Bagaimana pengaruh variasi *query* terhadap hasil pencarian dan hasil akurasi sistem berdasarkan nilai *mean average precision*, *recall* dan korelasi?

#### 1.3 Batasan Masalah

1. Dataset yang digunakan merupakan transliterasi aksara Arab-Latin Indonesia yang berpedoman pada SKB Menteri Agama dan Menteri P&K RI No. 158 Tahun 1987 dan No. 0543b/U/1987.

2. Dataset yang digunakan adalah transliterasi Al-Qur'an Juz 1, 2, 3, 4, 5, dan 30.
3. Pengkodean fonetis dilakukan per kata pada dataset dan *query* karena pencocokkan dilakukan per kata.
4. *Query* masukan berupa *string* latin berdasarkan variasi pengucapan yang mirip.

#### **1.4 Tujuan**

Tujuan penelitian ini adalah mengimplementasikan algoritma Double Metaphone termodifikasi untuk melakukan pengkodean fonetis terhadap *query* dan dataset, serta penerapan algoritma Jaro Winkler dalam pencocokkan kode fonetis dan pengukuran kesamaan *string*. Tujuan lainnya adalah untuk mengukur seberapa baik sistem dalam menangani variasi penulisan *query* berdasarkan pengucapan terhadap hasil pencarian ayat Al-Qur'an. Pengukuran yang digunakan untuk menguji kinerja sistem ialah dengan mengukur nilai *average precision* (AP), *mean average precision* (MAP), *recall*, dan korelasi sistem.

#### **1.5 Sistematika Penulisan Tugas Akhir**

Urutan penulisan laporan ini adalah sebagai berikut : Bagian 2 menunjukkan penelitian-penelitian yang terkait dengan tugas akhir ini. Sistem yang akan diajukan untuk pencarian ayat Al-Qur'an berdasarkan kemiripan fonetis dengan menggunakan algoritma Double Metaphone dan Jaro Winkler akan dijelaskan pada bagian 3. Bagian 4 berisi diskusi hasil pengujian dan analisis hasil pengujian serta evaluasi sistem. Bagian 5 berisi kesimpulan serta saran untuk penelitian selanjutnya.