

Abstrak

Mencari teks Arab dalam Alquran yang memiliki banyak ayat dan perbedaan bahasa dengan Bahasa Indonesia menimbulkan kesulitan tersendiri untuk pencarian ayat oleh masyarakat muslim. Dibutuhkan sistem pencarian ayat Alquran berbasis fonetis yang dapat memudahkan pengguna dalam mencari ayat menggunakan tulisan Latin berhuruf alfabet yang merepresentasikan bunyi pengucapan pengguna. Sebagai contoh, jika dilakukan pencarian kata

\< الْحَمْدُ لِلَّهِ >

maka sistem akan menampilkan seluruh ayat yang memiliki kemiripan bunyi dengan kata kunci. Untuk saat ini, sudah ada sistem pencarian ayat Alquran dengan menggunakan phonetic string matching, namun terbatas hanya dapat menemukan ayat berdasarkan query yang tidak lintas ayat. Kemudian jika dilakukan pencarian query lintas ayat

\< يَوْمَ الدِّينِ 4 (إِيَّاكَ) >

dengan pencocokanstring dalam database, maka sistem tidak dapat memberikan hasil pencarian dua ayat sekaligus. Pencarian lintas ayat sangat dibutuhkan karena pengguna belum tentu tahu apakah query yang digunakan itu dalam satu ayat atau lebih. Oleh karena itu, dibangun suatu sistem pencarian ayat Alquran berdasarkan kemiripan bunyi (fonetis) yang dapat melintasi ayat. Algoritma N-gram berupa trigram digunakan untuk menemukan ayat-ayat yang memiliki kemiripan bunyi (fonetis) karena memiliki Mean Average Precision (MAP) yang tinggi untuk kata kunci panjang. Untuk mencari lintas ayat, lima buah trigram ayat selanjutnya ditambahkan ke ujung trigram ayat sebelumnya. Kemudian diperoleh nilai MAP 0,9 dan Recall 0,93.