

1. Pendahuluan

Latar Belakang

Al-Quran merupakan kitab suci bagi 200 juta lebih penduduk muslim di Indonesia [1]. Kitab suci ini ditulis dalam Bahasa Arab. Dan secara statistik memiliki 77845 kata [2]. Hal tersebut membuat pencarian kata pada Al-Quran sulit untuk dilakukan secara manual. Oleh karena itu dengan bantuan komputer dapat membantu melakukan pencarian tersebut.

Dalam melakukan pencarian suatu kata tentu yang diharapkan adalah hasil yang relevan. Namun pada kenyataannya terkadang sistem pencarian tidak memberikan hasil apapun. Tentu untuk menangani masalah tersebut dapat dilakukan dengan beberapa cara, salah satunya dengan memberikan *nearest result*. Dimana ketika sistem tidak menemukan hasil sesuai *query*, sistem akan memberikan hasil yang mirip dengan *query*. Sebagai contoh ketika user memasukkan *query* **trus** pada sistem pencarian kamus Bahasa Inggris tidak memberikan hasil apapun karena memang kata tersebut tidak terdaftar dalam kamus. Maka sistem perlu untuk memberikan hasil kata **trust** karena mirip dengan *query*.

Dengan banyaknya jumlah penduduk muslim yang ada di Indonesia dapat mengakibatkan kasus *spelling error* menjadi besar pada sistem pencarian. Hal tersebut dapat terjadi jika berkaca pada kasus *web search engine*. Dimana tingkat kesalahan penulisan yang terjadi dapat mencapai 10 sampai 12 persen [3]. Angka tersebut cukup besar dari jumlah 200 juta penduduk. Sehingga perlu untuk menangani kasus kesalahan penulisan dengan *Tolerant Retrieval*.

Terdapat beberapa kasus *misspelling* yang terjadi. Kesalahan secara pengucapan yang mirip, contohnya pada kata رَسُول, dimana terdapat beberapa variasi *query search* untuk kata tersebut seperti pada Tabel 1.

Tabel 1. Variasi Input Untuk Kasus Variasi Penulisan Pada Kata رَسُول

Arab	Transliterasi Standar	Variasi Penulisan
رَسُول	rasul	rasul
		rosul
		rosuul
		rasuwl

Kemudian keteledoran user ketika memasukkan *query* juga menimbulkan masalah *misspelling* karena kedekatan huruf pada keyboard. Contohnya ketika hendak mencari kata رَسُول malah memasukkan *query* rasuk. *query* tersebut terjadi karena huruf L dekat dengan huruf K. Dan masih terdapat variasi lainnya seperti pada Tabel 2.

Tabel 2. Variasi Input Untuk Kasus Kesalahan Penulisan Pada Kata رَسُول

Arab	Transliterasi Standar	Variasi Karena Typo
رَسُول	rasul	tasul
		rasil
		rasuk
		rasyl

Dari masalah-masalah yang telah disebutkan, maka akan dibangun aplikasi pencarian kamus Al-Quran yang toleran terhadap bunyi dan kesalahan penulisan *suggestion result* ketika pencarian tidak ditemukan. Pengguna memasukkan *query* berupa huruf latin dan sistem akan memberikan kata yang relevan berupa aksara arab.

Topik dan Batasannya

Topik yang diangkat pada tugas akhir ini yaitu bagaimana membuat sistem pencarian kata pada ensiklopedia Al-Quran berdasarkan kemiripan fonetis agar mampu menangani pencarian dengan variasi penulisan. Contohnya ketika mencari kata رَسُول maka terdapat beberapa variasi penulisan seperti **rasul** atau **rosul**. ‘*suggestion* yang mendekati kata yang dicari. Contohnya ketika mencari kata رَسُول tetapi *query* yang ditulis malah **tasul** maka sistem harus memberikan *suggestion* **rasul**.

Adapun batasan pada tugas akhir ini adalah

1. Hanya menggunakan pemadanan aksara Arab-Latin sesuai dengan pelafalan orang Indonesia.
2. Hanya menangani Ensiklopedia Al-Quran.

Tujuan

Tujuan dari tugas akhir ini adalah untuk mengembangkan sistem pencarian pada kamus Al-Quran yang mampu untuk menangani variasi penulisan dan kesalahan penulisan. Menghitung nilai MAP (Mean Average Precision) dan Recall dari aplikasi yang dihasilkan.

Organisasi Tulisan

Organisasi tulisan pada jurnal tugas akhir ini terdiri dari pendahuluan, studi terkait, sistem yang dibangun, evaluasi dan kesimpulan. Pendahuluan berisi latar belakang, topik dan batasan, tujuan dan organisasi penulisan. Studi terkait berisi penjelasan singkat mengenai studi yang terkait dan beberapa metode terkait seperti dataset, phonetic string matching, n-gram dan levensthein distance. Sistem yang dibangun berisi gambaran sistem, dataset, tahapan kerja sistem pencarian fonetis dan sistem *suggestion*. Evaluasi berisi penjelasan teknik evaluasi yang digunakan, test set yang digunakan, hasil pengujian dan analisis pengujian. Kesimpulan berisi tentang kesimpulan yang dapat diambil dari evaluasi serta berisi saran yang dapat dijadikan sebagai dasar penelitian selanjutnya.