

ABSTRAK

Hadits adalah perkataan (sabda), perbuatan, ketetapan dan persetujuan dari Nabi Muhammad SAW yang dijadikan landasan hukum syariat Islam setelah al-Qur'an. Saat ini telah banyak website yang menyediakan informasi mengenai hadits untuk memudahkan pengguna dalam proses pembelajaran hadits, seperti website Lidwa Pustaka yang digunakan untuk pencarian hadits terjemahan Bahasa Indonesia atau yang biasa kita kenal sebagai Information Retrieval (IR).

Pada dasarnya, IR menyediakan kotak pencarian bagi pengguna untuk memasukkan query yang mencerminkan kebutuhan informasi pengguna. Query yang dimasukan oleh pengguna dicocokkan (*match*) dengan indeks koleksi dokumen hadits untuk menemukan hadits yang mengandung *query* tersebut, yang kemudian diurutkan (*ranking*) berdasarkan berbagai metode atau model. Namun demikian, penulisan sinonim atau *string variant* yang berbeda-beda dalam hadits terjemahan bahasa Indonesia, seperti “خَمْرُال” (al-khomru) ditulis menjadi “khamar”, “khamer”, “khamr” atau “minuman keras”, menjadikan proses pencarian hadits kurang tepat.

Oleh karena itu, penelitian ini bertujuan untuk memperbaiki sistem pencarian hadits terjemahan Bahasa Indonesia tersebut, menggunakan pendekatan *query expansion* dan *smoothing probability model*, yaitu Jelinek-Mercer Smoothing, Dirichlet Smoothing dan Absolute Discounting Smoothing. Metode *query expansion* tersebut dapat menangani proses pencarian secara semantik (sinonim dan *string variant*), sedangkan metode *smoothing* dapat melakukan pencarian hadits terhadap *query* pengguna secara *partial matching* dan mengestimasi relevansi dokumen hadits dengan menghitung nilai probabilitas kata yang tidak terlihat (*unseen word*) pada masing-masing dokumen hadits terhadap *query*, sehingga menghindari probabilitas nol untuk proses retrieval dokumen. Penggunaan *query expansion* dan Dirichlet smoothing pada penelitian ini, menghasilkan performansi yang paling optimal dibandingkan metode LSI, Cosine Similarity dan probabilitas biasa (tanpa *smoothing*), yaitu sebesar 62.54% untuk MAP ALL, 57.39% untuk MAP@30 dan 76.59% untuk recall, hal ini dikarenakan *query expansion* dapat melakukan pencarian secara semantik. Sedangkan untuk pengaturan parameter *smoothing* yang paling optimal adalah pada μ : 500 untuk metode Dirichlet Smoothing.