

1. Pendahuluan

1.1 Latar Belakang

Seiring dengan pertambahan jumlah umat islam yang meningkat tiap harinya [11], kebutuhan informasi hadits sebagai rujukan hukum ke dua setelah Al-Quran dalam agama islam menjadi sangat penting. Banyak beredar hadits palsu ataupun *bathil* yang akhirnya malah diyakini sebagai hadits *shahih* yang akhirnya dijadikan sebagai rujukan hukum. Seperti halnya hadits “*tuntutlah ilmu sampai ke negri Cina*” yang diyakini sebagian orang merupakan hadits *shahih*, tetapi nyatanya hanya berupa pribahas atau frasa arab saja [17].

Adapun hadits yang telah di *tashih* dan telah dibukukan masih dimiliki oleh kalangan terbatas dan membutuhkan keilmuan khusus dalam mempelajari atau menelaah kandungan daripadanya. Proses pencarian informasi didalamnya masih berbasis katalog buku.

Proses pencarian menggunakan katalog atau daftar isi berupa buku fisik memiliki banyak kekurangan, dan proses pencarian yang dilakukan bisa menghabiskan banyak waktu. Seiring dengan kemajuan teknologi *Information retrieval system* menjadi solusi akan permasalahan ini. *Information Retrieval System* adalah suatu sistem temu balik yang melakukan pencarian informasi dari kumpulan dokumen tertentu sesuai dengan *query* dari *user* [7]. *Query* itu sendiri berupa *keyword* (kata kunci) yang dimasukkan oleh *user* dan diyakini berhubungan dengan informasi yang dicari.

Aplikasi yang sudah ada memiliki beberapa kekurangan. Pada lidwa.com [6] proses pencarian di aplikasi ini menggunakan operator *or* saja dalam menangani *query user*. Hal ini mengakibatkan pencarian yang seharusnya semakin spesifik sesuai dengan *query user* menghasilkan nilai *recall* yang sangat tinggi.

Pembangunan sistem *information retrieval* pada penelitian ini menerapkan *vector space model* sebagai term *weighting* nya. Penggunaan metode ini didasarkan oleh keunggulannya dalam segi pencarian dibandingkan dengan metode sebelumnya [1]. Adapun metode ini memiliki beberapa kekurangan dimana sangat bergantung kepada kesamaan *query* dengan *indexing term* yang ada. Hal ini mengakibatkan sistem tidak mampu menampilkan dokumen yang relevan atau dikenal dengan “*false negative*” selain itu dokumen dengan konteks yang sama namun memiliki term yang berbeda juga tidak ditampilkan atau *retrieve*

mengakibatkan terjadinya “*false positive*” [18]. Untuk menutupi kekurangan ini diterapkan *thesaurus* dalam proses *query expansion* pada sistem yang dibangun. Dengan menerapkan *thesaurus* untuk *query expansion*, misal pencarian dengan *query* “kebersihan” akan dikembangkan menjadi beberapa *query* seperti *thaharah*, *tayamum*, *wudhu*, mandi dan *term* lainnya yang masih berhubungan berdasarkan dokumen *thesaurus* yang telah dibangun sebelumnya sehingga pencarian terhadap *query* bisa lebih meluas dan mengurangi nilai *false negative* dari hasil pencarian.

Pada penelitian ini merujuk pada jurnal acuan [13], dilakukan pengembangan pada proses pembangunan dokumen *thesaurus* dengan mengkombinasikan pembangunan secara otomatis menggunakan *analysis co-occurrence* dan pembangunan secara manual. Adapun dengan kombinasi ini pembangunan *thesaurus* dapat dilakukan lebih cepat dan terkontrol.

1.2 Perumusan masalah

Berdasarkan latar belakang yang telah dikemukakan, maka masalah yang dapat dirumuskan adalah sebagai berikut:

- a. Bagaimana membangun sebuah sistem *information retrieval* menggunakan metode VSM?
- b. Bagaimana membangun *thesaurus* untuk *query expansion* sistem temu balik hadits berbahasa Indonesia?
- c. Bagaimana pengaruh penerapan *thesaurus* untuk *query expansion* dalam sistem *information retrieval* yang dibangun terhadap nilai *precision*, *Mean Average Precision* (MAP), dan *recall* dari hasil pencarian *query user*?
- d. Bagaimana akurasi nilai balikan dari sistem *information retrieval* yang telah dibangun?

Sedangkan batasan masalah yang diberikan pada penelitian ini adalah sebagai berikut:

- a. Dokumen yang digunakan adalah dokumen Shahih Bukhari Muslim berbahasa Indonesia kitab ilmu, wudhu, tayamum, haid, sholat, zakat, dan thaharah (1161 dokumen hadits yang telah dipilih sebelumnya)
- b. Model sistem akan berjalan secara offline.

Pembangunan dan analisis sistem akan berfokus pada penggunaan *thesaurus* untuk *query expansion* dalam meningkatkan nilai *precision*, MAP serta *recall* dalam sistem *information retrieval* yang dibangun.

1.3 Tujuan

- a. Membangun sebuah sistem *information retrieval* dengan metode VSM.
- b. Membangun *thesaurus* untuk *query expansion* dari sistem temu balik hadits berbahasa Indonesia.
- c. Menganalisis bagaimana pengaruh *thesaurus* untuk *query expansion* terhadap sistem yang dibangun dengan melihat perbandingan nilai *precision*, *Mean Average Precision* (MAP) serta *recall* dari hasil balikan sistem yang dibangun.
- d. Menganalisis bagaimana akurasi hasil balikan dengan melihat parameter *Precision at 10* (P@10), *Mean Average Precision* (MAP), dan *recall*.

1.4 Metodologi penyelesaian masalah

Metodologi yang digunakan dalam penyusunan tugas akhir ini adalah sebagai berikut:

- a. Studi literatur.
 - Jurnal [13] dijadikan sebagai paper acuan dalam penggunaan *thesaurus* untuk *query expansion* dalam sistem *information retrieval* yang dibangun. Pada penelitian ini dilakukan pengembangan pada proses pembangunan dokumen *thesaurus* dengan mengkombinasikan pembangunan secara otomatis menggunakan *analysis co-occurrence* dan pembangunan secara manual.
 - Jurnal [4] sebagai paper pendukung untuk *query expansion* dalam sistem *information retrieval* yang dibangun.
 - *Course* [5] sebagai pendukung proses pembangunan *thesaurus* yang akan digunakan dalam proses *query expansion* nantinya.
- b. Menentukan requirement.

Sokumen hadits diambil dari “Hadits 9 Imam” dari lidwa.com [6]. Kemudian menentukan dokumen-dokumen hadits mana yang dijadikan sebagai dataset dalam pembangunan sistem.
- c. Membuat rancangan.

Dilakukan *preprocessing* terhadap dokumen dan *query user*. Meliputi proses *tokenizing*, *stemming*, *weighting* dan pembuatan *inverted index*. Setelah itu *query* dari *user* diproses dan diberikan nilai balikan hasil pencarian yang telah diranking terlebih dahulu. Hasil balikan merupakan hasil dari *query expansion* dengan *thesaurus* yang telah dibangun sebelumnya.
- d. Realisasi.

Model sistem dibangun dengan menggunakan PHP sebagai *interface* nya dan MySQL sebagai *database* nya.

e. Pengujian.

Dokumen hadits yang diuji sebanyak 1161 dokumen hadits berbahasa Indonesia dari Shahih Bukhari dan Shahih Muslim. Diambil dari kitab ilmu, *wudhu*, *tayamum*, haid, shalat, zakat, dan *thaharah*. *User* memasukkan *query* pada sistem yang dibangun kemudian sistem akan memberikan nilai balikan dokumen yang relevan hasil dari *query expansion*.

f. Analisis hasil.

Input akan berupa sebuah *query* dan output akan berupa kumpulan dokumen hasil pencarian dengan *query expansion*. Akurasi hasil yang akan diukur menggunakan parameter *precision at 10 (P@10)*, *Mean Average Precision (MAP)*, dan *recall*.