



Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Google sebagai perusahaan teknologi terbesar telah mengembangkan sebuah mesin pencarian berbasis *web* dan berhasil menerapkan metode IR (*information retrieval*) dalam penemuan kembali dokumen yang tersimpan di internet. Seiring dengan berkembangnya teknologi dan informasi, IR (*Information Retrieval*) atau Sistem Temu Kembali Informasi mulai diterapkan pada pencarian dalam halaman sebuah perpustakaan digital atau *digital library*.

Perpustakaan Digital sendiri merupakan terobosan baru dalam berbagai informasi terutama di bidang ilmu pengetahuan dan bidang pendidikan di mana para penggunanya bisa berbagi dan mengakses dokumen dengan mudah dan cepat. Salah satu bentuk perpustakaan digital yang telah banyak digunakan oleh lembaga pendidikan seperti universitas disebut dengan *repository*. Banyak universitas terkenal di Indonesia sudah menggunakan Perpustakaan Digital atau biasa disebut juga dengan *repository* dalam menyimpan dan berbagai informasi dokumen digital yang bisa diakses dan diunduh oleh siapa saja. Isi atau dokumen dari sebuah perpustakaan digital atau *repository* sebuah universitas bisa berupa jurnal, tesis, *ebook*, bahan skripsi, dan lain-lain yang kegunaannya sebagai referensi mahasiswa dalam menunjang materi perkuliahan dan penelitiannya.

Banyaknya data dokumen digital yang disimpan dengan banyak judul serta kategori menyebabkan waktu pencarian lama dan tidak efektif. Penerapan Sistem Temu Kembali Informasi dengan teknik *Hierarchical Agglomerative Clustering* ke dalam pencarian halaman sebuah perpustakaan digital atau *repository* ini bertujuan untuk mengelompokkan dokumen-dokumen dengan topik yang sama ke dalam satu *cluster* sehingga memberikan kemudahan dalam pencarian dokumen. Dengan adanya pengelompokkan dokumen, waktu yang

dibutuhkan untuk pencarian akan lebih cepat. Dokumen skripsi atau jurnal yang telah dikelompokkan akan tersusun dengan terstruktur dan rapi sesuai dengan kemiripan dokumen tersebut.

Menurut Rijbergen (1979), *clustering* dokumen telah lama diterapkan untuk meningkatkan efektivitas temu kembali informasi. Penerapan *clustering* ini bersandar pada suatu hipotesis (*cluster-hypothesis*) bahwa dokumen yang relevan akan cenderung berada pada *cluster* yang sama jika pada koleksi dokumen dilakukan *clustering*. Beberapa penelitian untuk dokumen berbahasa inggris menerapkan *clustering* dokumen untuk memperbaiki kinerja dalam proses *searching* (Voorhess,1986; Tombros,2002). Untuk dokumen berbahasa Indonesia penelitian bidang IR adalah oleh Vega (2001) dan Tala (2004) yang meneliti efek *stemming* pada hasil pencarian.

Penelitian tentang penerapan *clustering* untuk perbaikan kinerja perolehan informasi untuk dokumen berbahasa Indonesia belum pernah dilakukan. Hal ini mengingat secara umum penelitian tentang komputasi bahasa untuk dokumen Bahasa Indonesia juga masih sangat minim (Nazief, 2000), bahkan *test-bed* yang dapat digunakan secara standar untuk penelitian IR belum ada (Asian, 2004).

Berdasarkan permasalahan di atas, maka penulis memutuskan melakukan penelitian mengenai seberapa efektif kombinasi penerapan metode Model Ruang Vektor dan Teknik *Hierarchical Clustering* terhadap temu kembali informasi dokumen berbahasa Indonesia pada halaman pencarian perpustakaan digital atau *repository* UIN SUSKA RIAU Fakultas Sains dan Teknologi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan di atas, maka dapat dirumuskan masalah dalam penelitian ini adalah bagaimana merancang bangun sebuah perpustakaan digital yang memiliki mesin pencarian dokumen efektif yang relevan dengan *query* dan dapat menyimpan hasil keluaran dokumen yang

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

©Hak cipta milik UIN Suska Riau
Sta...
of S...
Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

sudah *tercluster* menggunakan Kombinasi Metode Model Ruang Vektor dan *Hierarchical Agglomerative Clustering*.

1.3 Batasan Masalah

Agar cakupan masalah tidak terlalu luas, maka diperlukan pembatasan masalah. Adapun batasan masalah dalam penelitian ini yaitu:

1. Sample pengujian adalah 40 koleksi dokumen abstrak tugas akhir dan jurnal berbahasa Indonesia dengan format *docx*.
2. Algoritma Stemmer yang dipakai adalah Algoritma Porter Stemmer Bahasa Indonesia.

1.4 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dari penelitian ini adalah membangun sistem temu kembali informasi yang dapat menyimpan hasil keluaran dokumen yang sudah *tercluster* dan relevan dengan *query* sehingga dapat memaksimalkan waktu dan hasil pencarian menggunakan Kombinasi Metode Model Ruang Vektor dan *Hierarchical Agglomerative Clustering*.

1.5 Sistematika Penulisan

Sistematika penulisan laporan tugas akhir ini dibagi menjadi enam bab, seperti dijelaskan sebagai berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan dasar-dasar penulisan tugas akhir yang terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan penelitian dan sistematika penulisan tugas akhir.

BAB II LANDASAN TEORI

Bab ini membahas teori-teori yang berhubungan dengan spesifikasi pembahasan penelitian yang akan diangkat.

BAB III METODOLOGI PENELITIAN

Bab ini membahas langkah-langkah yang dilaksanakan dalam proses penelitian yaitu, perumusan masalah, studi pustaka, pengumpulan data dan informasi, serta alur penelitian.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB IV ANALISA

Bab ini berisikan tentang analisa dari penelitian yang dilakukan dalam tugas akhir ini.

BAB V IMPLEMENTASI

Bab ini berisikan tentang langkah-langkah implementasi berdasarkan analisa yang telah dilakukan pada bab sebelumnya.

BAB VI KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang dihasilkan dari pembahasan penelitian dan beberapa saran sebagai hasil akhir dari penelitian yang telah dilakukan dan pengembangan lebih lanjut

