

STMIK GI MDP

Program Studi Teknik Informatika
Skripsi Sarjana Komputer
Semester Ganjil Tahun 2010/2011

PENERAPAN METODE *CLUSTERING* HIRARKI *AGGLOMERATIVE* UNTUK KATEGORISASI DOKUMEN PADA *WEBSITE* SMA NEGERI 2 PALEMBANG

Lili Andriani	2007250026
Kgs. M. Iqbal	2007250079

Abstrak

Dalam berbagai organisasi seperti kantor, sekolah, perguruan tinggi dan industri saat ini masih mempertahankan banyak informasi dalam bentuk dokumen, namun khusus pada organisasi sekolah maka dokumen tentang materi pelajaran, dokumen soal-soal ujian dan koleksi berbagai dokumen lainnya itu sangatlah penting karena dokumen tersebut berisi miliaran informasi. Salah satu cara untuk mengatasi dokumen yang luar biasa banyak adalah analisis *cluster* dan menyebabkan kebutuhan untuk mengatur seperangkat besar dokumen tersebut kedalam kategori melalui *clustering*. Hal ini digunakan untuk membagi dokumen yang besar ke dalam kelompok dokumen yang saling berkaitan erat atau memiliki ukuran kesamaan. Kami mengusulkan ukuran kesamaan baru untuk menghitung kesamaan dokumen berbasis teks berdasarkan *Term Frequency and Inverse Document Frequency (TF/IDF)* menggunakan *Vector Space Model*, dan menerapkan ukuran kesamaan baru itu ke dalam *Clustering Hierarchical Agglomerative*. Model ini akan menyediakan dokumen yang memiliki kesamaan dan perhitungan yang akurat serta dapat meningkatkan efektivitas teknik *clustering*. Hasil dari pengelompokan dokumen sejenis dengan proses *clustering* diharapkan akan membantu pengguna menemukan informasi yang relevan, lebih cepat, dan akan memungkinkan untuk pencarian dokumen pada arah yang lebih tepat.

Kata Kunci :

Analisis Cluster, Vector Space Model, Term Frequency and Inverse Document Frequency dan *Clustering Hirarki Agglomerative* Algoritma.

STMIK GI MDP

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Pernyataan Penyusunan Skripsi

Kami, Lili Andriani,
Kgs. M. Iqbal

Dengan ini menyatakan bahwa Skripsi yang berjudul :

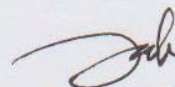
**PENERAPAN METODE *CLUSTERING HIRARKI AGGLOMERATIVE*
UNTUK KATEGORISASI DOKUMEN
PADA *WEBSITE SMA NEGERI 2 PALEMBANG***

adalah benar hasil karya kami dan belum pernah diajukan sebagai karya ilmiah,
sebagian atau seluruhnya, atas nama kami atau pihak lain.

Penulis,



Lili Andriani
2007250026



Kgs. M. Iqbal
2007250079

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Tingginya penggunaan internet juga telah memacu pesatnya pertumbuhan dan pertukaran informasi. Tidak hanya dalam dunia maya, tetapi jumlah informasi dalam bentuk *text* juga semakin banyak digunakan di berbagai institusi dan sekolah. Jumlah dokumen elektronik yang semakin besar merupakan sumber informasi yang berharga. Semakin banyak dokumen yang ada maka semakin banyak juga data yang tidak mudah untuk ditelusuri dan dikelompokkan sesuai dengan kategori yang ada. kesalahan pada kategorisasi masih bisa terjadi oleh admin sehingga dapat kemungkinan terjadi kesalahan informasi.

Melihat keadaan tersebut, maka penulis ingin membuat suatu Fungsi kategorisasi data secara otomatis yang akan mempermudah dalam hal penentuan data yang diinginkan oleh admin. Dengan hanya mengetikkan judul, deskripsi dan memasukkan *source* dari dokumen maka dokumen akan otomatis terkategorisasi ke dalam topik yang bersangkutan. Begitu juga dengan pengguna yang menggunakan *website* ini untuk mencari informasi yang diperlukan lebih terarah, di mana saat pengguna menginput kata kunci, maka akan langsung diarahkan kepada artikel yang telah tersedia pada

Website SMA Negeri 2 Palembang ini saja, sehingga pencarian lebih cepat dan tepat.

Pengelompokan (*clustering*) dokumen merupakan sebuah cara yang dapat digunakan untuk mempermudah pencarian dokumen dalam *database*. *Clustering* merupakan salah satu metode dalam data *mining* yang bisa digunakan untuk mengelompokkan data. *Clustering* merupakan proses pengelompokan data sehingga semua anggota dari bagian data memiliki kemiripan berdasarkan perhitungan jarak antara kata dalam judul dokumen. Salah satu aplikasi dari *clustering* adalah *document clustering*. Tahapan *clustering* adalah: representasi dokumen, penggunaan *cluster* algorithm, dan evaluasi. Yang diberikan sebagai masukan dalam dokumen *clustering* adalah *source* dari dokumen dan *similarity* yang akan digunakan untuk menemukan beberapa *cluster* dari dokumen yang saling terkait satu sama lainnya. Dalam membentuk dokumen *clustering* ada beberapa metode yang dapat digunakan, tetapi pada umumnya metode yang sering digunakan yaitu: metode *Clustering Hirarki Agglomerative* yang merupakan salah satu bagian dari metode hirarki. Berdasarkan hal tersebut, maka penulis mencoba akan melakukan pengelompokan dokumen berdasarkan metode hirarki yang akan diimplementasikan pada *website* SMA Negeri 2 Palembang.

1.2 Rumusan Masalah

Adapun yang menjadi rumusan masalah dari penelitian ini adalah “Bagaimana mengimplementasikan dan mengukur efektifitas dari metode *clustering hirarki agglomerative* dalam mengkatagorisasikan/mengelompokan suatu dokumen pada *website* SMA Negeri 2 Palembang?”

1.3 Ruang Lingkup

Agar pembahasannya tidak meluas, untuk memberi batas pada ruang lingkup penelitian sekaligus memberi fokus pada penyelesaian, maka kami berikan beberapa batasan yaitu:

- a. Penelitian ini akan menitikberatkan pada pengimplementasian metode *clustering hierarchical agglomerative* dalam menggolongkan dokumen.
- b. Berita yang ada di *website* merupakan informasi yang memang bebas untuk dipublikasikan kesemua orang.
- c. Dokumen untuk pengujian sistem berasal tentang materi pelajaran, contoh-contoh soal pelajaran serta kegiatan yang berkaitan tentang SMA Negeri 2 Palembang saja.
- d. Dokumen yang ada pada *website* SMA Negeri 2 Palembang ini dapat berupa *file* berbentuk *doc*, *pdf*, *txt*, *html*, dan gambar (*jpg*, *gif* dll).
- e. *Keyword* sudah diketahui dengan demikian pada tahap *preprocessing* yang akan dilakukan perhitungan kemunculan *keyword* pada setiap judul dari dokumen.

- f. Deskripsi ditulis oleh admin dimana deskripsi tersebut menggambarkan isi dari dokumen tersebut.
- g. Dokumen materi dan soal-soal pelajaran yang ada pada *website* ini hanya dikhususkan untuk siswa SMA Negeri 2 Palembang saja.
- h. Kategorisasi yang dimaksud tidak mempunyai label (nama) karena pada saat jalannya proses *clustering*, komputer tidak bisa memberi nama.

1.4 Tujuan dan Manfaat

Adapun tujuan serta manfaat dari pembuatan fungsi pengkategorisasian terhadap dokumen pada *website* SMA Negeri 2 Palembang ini adalah:

1.4.1 Tujuan Penelitian

Untuk mengimplementasikan dan mengukur efektifitas metode *Clustering Hirarki Agglomerative* dalam mengkategorisasikan atau mengelompokkan suatu dokumen kedalam topik-topik yang sesuai pada data di *website* SMA Negeri 2 Palembang secara otomatis.

1.4.2 Manfaat Penelitian

Dengan adanya proses *clustering* dokumen yang menerapkan metode *Clustering Hirarki Agglomerative* pada *website* SMA Negeri 2 Palembang ini, maka diharapkan dapat memudahkan *admin* dalam mengkatagorisasikan data serta mempermudah pengguna dalam mencari dokumen berdasarkan tingkat kemiripan antara dokumen yang tersedia dengan kata kunci yang dicari oleh pengguna.

1.5 Metodologi Penelitian

Dalam penulisan skripsi ini, adapun langkah-langkah dalam metodologi ini antara lain :

1. Studi Litelatur dan Pengumpulan Data

Tahapan yang dilakukan diawali dengan melakukan studi pustaka dari *text book* dan artikel-artikel guna memberi pemahaman yang fundamental akan konsep yang digunakan pada banyak alternatif penyelesaian pada *Clustering Hirarki Agglomerative*. Selain itu dilakukan studi literatur yang diperoleh dari internet. Selain itu dilakukan dengan cara membaca buku-buku, dan jurnal ilmiah yang berhubungan dengan metode pencarian data dan *programming* untuk mendapatkan materi yang dapat dijadikan landasan dan referensi bagi penyusunan skripsi ini.

2. Perancangan Sistem dan Analisis

Pada tahap ini dilakukan pemahaman keinginan *user* dan menganalisis permasalahan lebih mendalam dengan melihat beberapa faktor yaitu pendefinisian masalah, tujuan dan pengembangan sistem. penulis juga mempelajari metode pendekatan hirarki dan melakukan analisis terhadap algoritma yang akan digunakan yaitu *Clustering Hirarki Agglomerative*.

3. Perencanaan Sistem

Pada tahap ini, penulis mulai melakukan perancangan dari hasil dua tahap sebelumnya, yaitu mengimplementasikan metode hirarki dan membuat *flowchart*.

4. Implementasi

Pada tahap ini, hasil perancangan mulai akan dibuat yaitu Perancangan dan pengembangan perangkat lunak berdasarkan analisa yang sudah dilakukan dan mengimplementasikan perhitungan dengan metode hirarki yang dibuat dalam bentuk koding program.

5. Pengujian Sistem

Pada tahap ini, dilakukan pengujian terhadap sistem yang telah jadi dengan menggunakan data-data yang telah ada. Hasil pengujian ini kemudian dijadikan dasar untuk membuat perbaikan-perbaikan yang diperlukan untuk menghasilkan sistem yang diharapkan. Pengujian sistem dilakukan dengan mencoba mengetikkan berbagai kata kunci dipencarian data.

6. Perbaikan atau Penambahan Data

Apabila terdapat kesalahan setelah pengujian, maka sistem tersebut akan diperbaiki dan data pada *database* akan ditambah sehingga dokumen yang berkaitan tentang materi pelajaran, contoh-contoh soal, dan dokumen lainnya yang berkaitan tentang SMA Negeri 2 Palembang ini semakin lengkap.

7. Analisa dan Simpulan

Pada tahap ini, penulis memuat analisa dan simpulan dari sistem yang telah selesai diuji. Membuat suatu kesimpulan dari pengujian sistem penelitian akhir dengan membandingkan apakah hasilnya seperti yang diharapkan pada tujuan penelitian akhir sebelumnya.

1.6 Sistematika Penulisan

Sistematika penulisan merupakan gambaran mengenai bab-bab yang disusun oleh penulis dalam laporan skripsi ini. Penulisan skripsi terdiri dari lima bab, dimana tiap bab terdiri dari beberapa sub bab. Susunan garis besar sistematika penulisan skripsi dapat dilihat di bawah ini.

BAB 1 PENDAHULUAN

Pada bab ini dijelaskan tentang latar belakang pengambilan judul skripsi, lingkup materi yang akan dibahas, rumusan masalah, tujuan dan manfaat dari penelitian, metodologi yang digunakan, serta sistematika penulisan.

BAB 2 LANDASAN TEORI

Pada bab ini dijelaskan tentang teori umum dan teori khusus secara mendalam mengenai istilah-istilah bidang ilmu yang terkait dalam perancangan fungsi pengkatagorisasian dokumen pada *website* SMA Negeri 2 Palembang dengan metode hirarki.

BAB 3 ANALISIS RANCANGAN DAN ALGORITMA PROGRAM

Pada bab ini akan diuraikan tentang spesifikasi perangkat lunak dan perangkat keras yang digunakan untuk penelitian, metodologi yang digunakan, struktur data yang digunakan, rancangan layar dan diagram alir.

BAB 4 IMPLEMENTASI DAN ANALISIS PROGRAM

Pada bab ini berisi hasil dan pembahasan dari fungsi pengkategorisasian dokumen yang dirancang, meliputi kelebihan/keunggulan yang diperoleh, prosedur dalam pengujian program, dan menganalisis hasil uji coba tersebut berdasarkan fungsionalitas terhadap efektifitas penerapan metode *Clustering Hirarki Agglomerative* terhadap fungsi pengkategorisasian dokumen dalam *website* SMA Negeri 2 Palembang ini

BAB 5 PENUTUP

Pada bab ini berisi rangkuman hasil analisa mengenai perancangan “Penerapan metode hirarki untuk kategorisasi dokumen pada SMA Negeri 2 Palembang” dalam bentuk kesimpulan dan saran yang dapat dijadikan sebagai acuan untuk pengembangan aplikasi ini lebih lanjut.