

Pembuatan Aplikasi Web Mining Untuk Pencarian Informasi Berdasarkan Istilah Menggunakan Centroid Linkage Hierarchical Method

Nur Asmawati¹⁾, Imam Marzuki²⁾, Dwi Putri Kartini³⁾

¹⁾Mahasiswa Program Studi Teknik Elektro, Fakultas Teknik, Universitas Panca Marga

^{2),3)}Dosen Program Studi Teknik Elektro, Fakultas Teknik, Universitas Panca Marga
Jl. Yos Sudarso 107 Pabean Dringu Probolinggo 67271

Terima Naskah : 29 Juni 2014

Terima Revisi : 16 Oktober 2014

ABSTRAK

Pencarian informasi (*Information Retrieval*) berdasarkan kata kunci berguna untuk pencarian terarah dan bisa membantu ketika ingin mengetahui informasi yang berhubungan dengan kata kunci yang dicari. Begitu pula dengan pencarian terhadap dokumen yang memuat istilah tertentu yang sama maka dibutuhkan suatu aplikasi yang memberikan informasi kepada pengguna yang melakukan pencarian terhadap sekumpulan dokumen yang berisikan informasi istilah yang sama. Dalam penelitian ini, pembuatan aplikasi *Web Mining* Untuk Pencarian Informasi Berdasarkan Istilah Menggunakan *Centroid Linkage Hierarchical Method* dengan bahasa pemrograman *java*. Ketika *user* ingin mengetahui informasi berdasarkan istilah, dibutuhkan suatu sistem yang memberikan kemudahan pada *user*. Dengan menerapkan *text mining* dan metode *centroid linkage hierarchical method*.

Uji coba difokuskan untuk mengetahui nilai varian dan waktu yang diperlukan selama proses. Nilai varian digunakan untuk mendapatkan nilai *cluster* yang tepat secara otomatis dan waktu yang diperlukan selama proses digunakan untuk mengetahui lamanya waktu yang diperlukan selama proses berlangsung. Percobaan dan evaluasi dilakukan terhadap kata kunci istilah : "KPK" dengan jumlah dokumen yang berbeda yaitu 10 dokumen, 30 dokumen dan 60 dokumen. Dari uji coba dapat disimpulkan bahwa nilai varian mengalami penurunan ketika jumlah dokumen semakin banyak. Sebaliknya nilai varian akan semakin meningkat apabila dokumen semakin sedikit. Hal ini membuktikan bahwa pada dokumen yang semakin banyak maka kemungkinan terjadi pembentukan cluster secara otomatis akan semakin besar. Sedangkan nilai waktu yang diperlukan selama proses adalah meningkat seiring dengan bertambahnya jumlah dokumen. Hal ini dikarenakan ketika jumlah data semakin besar, maka tahapan proses dalam klusterisasi menggunakan metode ini akan semakin panjang.

Kata kunci : Pencarian informasi, istilah, *Text Mining*, *centroid linkage hierarchical method*

ABSTRACT

Information retrieval based on keyword searches are useful for directional and can help when you want to know information related to the keywords being searched. Similarly, the search for documents that contain certain terms are the same then needed an application that provides information to users who conduct a search of a set of documents containing information similar terms. In this final task, making the application of Web Mining Based Information To search term Centroid Linkage Hierarchical Method Using the Java programming language. When a user wants to find out information based on the term, we need a system that provides convenience to the user. By applying text mining and centroid method linkage hierarchical method.

The trial focused to determine the value of the variant and the time required for the process. Variance is used to obtain the exact value of the cluster automatically and the time required for the process used to determine the length of time required for the process. Experimentation and evaluation of the key words term "Commission" with the number of different documents that 10 documents, 30 documents and 60 documents. From the experiments it can be concluded that the variant has decreased while the number of documents more. Instead variance will increase if the fewer documents. This proves

that documents more and more the possibility of the formation of the cluster will automatically increase. While the value of the time required for the process is increased along with the increasing number of documents. This is because when a greater amount of data, then the clustering process steps in using this method will become longer.

Keyword : Information retrieval, Terms, Text Mining, centroid linkage hierarchical method

PENDAHULUAN

Semakin canggihnya perangkat untuk meng Kebutuhan program aplikasi di zaman yang serba canggih seperti sekarang ini tentunya akan menjadi sangat penting. Aplikasi tersebut dapat digunakan oleh semua orang yang memanfaatkannya sehingga memudahkan mereka untuk memenuhi kebutuhannya yaitu informasi yang *up to date* secara cepat.

Pencarian informasi (*Information Retrieval*) berdasarkan kata kunci berguna untuk pencarian terarah dan bisa membantu ketika ingin mengetahui informasi yang berhubungan dengan kata kunci yang dicari. Begitu pula dengan pencarian terhadap dokumen yang memuat istilah tertentu yang sama maka dibutuhkan suatu aplikasi yang memberikan informasi kepada pengguna yang melakukan pencarian terhadap sekumpulan dokumen yang berisikan informasi istilah yang sama.

Bertitik dari masalah yang ada penulis tertarik untuk mengambil judul “ Pembuatan Aplikasi *Web Mining* Untuk Pencarian Informasi Berdasarkan Istilah Menggunakan *Centroid Linkage Hierarchical Method*”

Dalam penelitian ini akan dibahas tentang bagaimana cara untuk mendapatkan informasi istilah dari dokumen yang menggunakan telah mengalami proses *text mining* dengan tahapan antara lain *parsing*, *tokenizing*, *analisa semantik* dan bagaimana cara untuk mengelompokkan dokumen tersebut berdasarkan kata kunci istilah yang telah ditemukan. Metode yang digunakan dalam pengelompokan ini adalah metode *centroid linkage hierarchical method*. Besarnya data pada masing – masing hasil pengklasteran selanjutnya akan digunakan untuk menentukan hasil pengelompokan dokumen.

Adapun rumusan masalahnya adalah :

1. Bagaimana mengimplementasikan metode *centroid linkage hierarchical method* untuk pencarian informasi berdasarkan kata kunci istilah tertentu.

2. Bagaimana cara mengambil data – data teks yang berasal dari beberapa situs *online* yang terus *up to date*.
3. Bagaimana cara mengimplementasikan aplikasi kedalam bahasa pemrograman

Penelitian ini bertujuan untuk membuat aplikasi *web mining* untuk pencarian informasi berdasarkan istilah menggunakan *centroid linkage hierarchical method*.

Adapun batasan masalah dalam penelitian ini adalah :

1. Klasterisasi hanya menggunakan satu macam kata kunci saja yaitu kata kunci istilah tertentu.
2. Data yang digunakan sebagai data input adalah data yang diambil dari beberapa situs online yang *up to date*.

Sedangkan manfaat yang diharapkan dalam penelitian ini adalah :

1. Untuk pencarian informasi terkait istilah tertentu secara cepat.
2. Mengimplementasikan hasil study penulis

Penelitian Sebelumnya

Berdasarkan penelitian Entin Martiana “Mesin Pencari Dokumen Dengan Pengklasteran Secara Otomatis “. *Web mining* untuk pencarian berdasarkan kata kunci dengan pengklasteran otomatis adalah suatu metode pencarian dokumen dengan cara mengelompokkan atau mengklaster dokumen dari dokumen-dokumen berdasarkan kata kuncinya. Selanjutnya dilakukan pengklasteran dengan metode *centroid linkage hierarchical method* (CLHM) terhadap jumlah kata kunci yang diperoleh dari masing-masing dokumen. Dalam pengklasteran, umumnya harus dilakukan inisialisasi jumlah klaster yang ingin dibentuk terlebih dahulu, padahal pada beberapa kasus pengklasteran, user bahkan tidak tahu berapa banyak klaster yang bisa dibangun. Untuk itu, pada makalah ini diaplikasikan metode Valley Tracing sebagai constraint yang akan melakukan identifikasi terhadap pergerakan varian dari tiap

tahap pembentukan kluster dan menganalisa polanya untuk membentuk suatu kluster secara otomatis (automatic clustering). Data yang digunakan adalah data hasil dari proses text mining pada dokumen. Dari percobaan yang dilakukan dengan 424 dokumen hasilnya memberikan simpulan bahwa pada umumnya pencarian dokumen menggunakan teknik pengklasteran dengan algoritma CLHM dapat digunakan untuk mengelompokkan dokumen dengan jumlah yang tepat secara otomatis [2].

Berdasarkan penelitian Nadia Damayanti 2011 “Temu Kembali Informasi Berdasarkan Lokasi Pada Dokumen Yang Dikelompokkan Menggunakan Metode Centroid Linkage Hierarchical” Pencarian informasi berdasar kata kunci dapat membantu pengguna ketika ingin mengetahui informasi yang berhubungan dengan kata kunci yang dicari. Begitupun dengan pencarian informasi ketika pengguna ingin mengetahui kelompok dokumen yang memuat lokasi tertentu yang sama. Karenanya dibutuhkan suatu sistem yang memberikan informasi kepada pengguna yang ingin melakukan pengelompokkan dokumen berdasar lokasi tertentu yang sama. Dengan menerapkan temu kembali informasi dalam mencari kata kunci lokasi pada dokumen dan metode Centroid Linkage Hierarchical sebagai metode pengelompokkan data diharapkan dapat memberikan hasil yang optimal dalam mengelompokkan dokumen dan hanya akan mengambil informasi yang mempunyai tingkat kepentingan tinggi [1].

Informasi

Informasi adalah data yang diolah menjadi bentuk yang lebih berguna bagi yang menerimanya. Kualitas informasi tergantung dari tiga hal yaitu informasi harus akurat, tepat pada waktunya dan relevan [4].

Temu Kembali Informasi

Temu kembali informasi merupakan suatu proses pencarian informasi yang didasarkan pada satu kata kunci yang diharapkan dapat memenuhi keinginan *user* dari kumpulan dokumen yang ada. Selain itu temu kembali informasi merupakan pembelajaran tentang sistem pengindeksan, penvarian dan mengingat data khususnya.

Crawler

Web Crawler atau biasa juga disebut *web spider* memiliki tugas untuk meng-*crawl*

(merayapi) seluruh informasi dari suatu *website*. *Spider* akan menggali seluruh data dari suatu *website* termasuk didalamnya komponen *website* seperti : meta data, *keywords* dan lain sebagainya. Kemudian *web spider* akan mengumpulkan atau mengindeks seluruh *website* ke dalam data *base search engine*. Sampai pada akhirnya halaman *website* akan di tampilkan pada *Search Engine Result Page (SERP)*. Sederhananya seperti ini : Ketika Anda mengetikkan *query* atau kata kunci pada search engine, mesin akan mencari database index dan memberikan daftar halaman *web* yang sesuai dengan kriteria yang Anda cari. Proses *web crawler* dalam mengunjungi setiap dokumen web disebut web crawling atau spidering.

Fungsi *Web Crawler / Web Spider*:

1. *Web crawler* biasa digunakan untuk membuat salinan sebahagian atau keseluruhan halaman *web* yang telah dikunjunginya agar dapat diproses lebih lanjut oleh system pengindeksan.
2. *Web crawler* dapat digunakan untuk proses pemeliharaan sebuah *website*, seperti memvalidasi kode html sebuah web.

Web crawler juga digunakan untuk memperoleh data yang khusus, seperti mengumpulkan alamat email.

Text Mining

Text mining adalah proses penemuan akan informasi sebelumnya tidak terungkap dengan memproses dan menganalisa data dalam jumlah yang besar. Hasil yang diharapkan adalah informasi yang baru yang tidak terungkap jelas sebelumnya.

Text mining juga menghadapi masalah yang sama yaitu jumlah data yang besar, dimensi yang tinggi, data dan struktur yang terus berubah. Berbeda dengan *data mining* yang utamanya memproses data terstruktur, data yang digunakan *text mining* pada umumnya dalam bentuk tidak terstruktur. Akibatnya *text mining* mempunyai tantangan tambahan yang tidak ditemui di *data mining*, seperti struktur teks yang kompleks dan tidak lengkap, arti yang tidak jelas dan tidak standart, dan bahas yang berbeda ditambah translasi yang tidak akurat. Dikarenakan struktur data ditujukan agar mudah diproses komputer secara otomatis, *pre-processing data* di *data mining* jauh lebih mudah dilakukan dari pada di *text mining*.

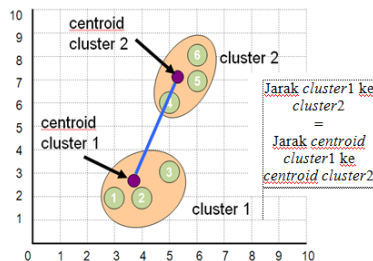
Dalam penelitian ini *text mining* yang digunakan hanyalah *tokenizing* yaitu proses penghilang tanda baca pada kalimat yang ada

dalam dokumen sehingga menghasilkan kata-kata yang berdiri sendiri-sendiri [1].

Klasterisasi dan Centroid Linkage Hierarchical Method

Cluster yaitu menemukan kumpulan objek – objek dalam satu kelompok sama (atau punya hubungan) dengan yang lain dan berbeda (atau tidak berhubungan) dengan objek – objek dalam kelompok lain. Tujuan dari analisa cluster adalah meminimalkan jarak di dalam cluster dan memaksimalkan jarak antar cluster. Sedangkan klasterisasi adalah proses pengelompokan objek atau data kedalam grup yang anggotanya memiliki kesamaan tertentu. Klasterisasi merupakan metode penganalisaan data yang sering dimasukkan sebagai salah satu metode data mining, yang tujuannya adalah mengelompokkan data dengan karakteristik yang sama ke suatu wilayah yang sama dan data dengan karakteristik yang berbeda ke wilayah yang yang sama dan data dengan karakteristik yang berbeda ke wilayah yang lain.

Centroid linkage hierarchical method adalah proses pengklasteran yang didasarkan pada jarak antar centroidnya. Metode ini baik untuk clustering dengan data set distribution [1]



Gambar 1 Ilustrasi Algoritma Centroid Linkage Hierarchical Method

Algoritma centroid linkage hierarchical method adalah sebagai berikut :

1. Diasumsikan setiap data dianggap sebagai cluster. Jika n = jumlah data dan c = jumlah klaster, maka c = n.
2. Menghitung jarak antar klaster.
3. Mencari dua klaster yang mempunyai jarak centroid antar klaster yang paling minimal dan digabungkan (merge) kedalam klaster baru sehingga c = c-1.
4. Kembali ke langkah 3, dan diulangi sampai dicapai klaster yang diinginkan.

Varian Cluster

Berdasarkan penelitian yang dikerjakan oleh Winda Ayu Irianto analisa cluster bisa diperoleh dari kepadatan suatu cluster yang dibentuk sehingga secara relatif dapat dikelompokkan dalam bentuk yang homogen. Kepadatan suatu cluster dapat ditentukan dengan variance within cluster (V_w) dan variance between cluster (V_b) dimana varian tiap tahap pembentukan cluster dapat dihitung dengan rumus :

$$V_c = \frac{1}{n_c - 1} \sum_{i=1}^{n_c} (x_i - \bar{y})^2 \dots \dots \dots (2.1)$$

Dimana, V_c = varian pada cluster c
 C = 1..k, dimana k=jumlah cluster
 n_c = jumlah data pada cluster c
 x_i = data ke -i pada suatu cluster
 \bar{y} = rata - rata dari data pada suatu cluster

Selanjutnya dari nilai varian di atas, kita dapat menghitung nilai variance within cluster (V_w) dengan rumus :

$$V_w = \frac{1}{N - c} \sum_{i=1}^c (n_i - 1) \cdot V_i^2 \dots \dots \dots (2.2)$$

Dimana, N = jumlah semua data
 n_i = jumlah data cluster i
 V_i = varian pada cluster i

Dan nilai variance between cluster (V_b) dengan rumus :

$$V_b = \frac{1}{c - 1} \sum_{i=1}^c n_i (\bar{y}_i - \bar{y})^2 \dots \dots \dots (2.3)$$

Dimana, \bar{y} = rata-rata dari y_i

Salah satu metode yang digunakan untuk menentukan cluster yang ideal adalah batasan variance, yaitu dengan menghitung kepadatan cluster berupa variance within cluster (V_w). Cluster yang ideal mempunyai V_w minimum yang mempresentasikan interna homogeneity dan maksimum V_b yang menyatakan external homogeneity.

$$V = \frac{V_w}{V_b} \dots \dots \dots (2.4)$$

Java

Java merupakan bahasa berorientasi objek (OOP) yaitu cara ampuh dalam pengorganisasian dan pengembangan perangkat lunak

Precision

Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. *Precision* merupakan perbandingan jumlah dokumen relevan yang didapatkan sistem dengan jumlah seluruh dokumen yang terambil oleh sistem baik relevan maupun tidak relevan. Menurut pengertian ini dapat dinyatakan dalam persamaan

$$P = \frac{TP}{TP+FP} \dots\dots\dots(2.5)$$

Dimana :

P (Precision) = Tingkat *precision* pencarian

TP (True Positive) = Dokumen relevan yang ditemukan

FP (False Positive) = Dokumen tidak relevan yang ditemukan

Recall

Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. *Recall* merupakan perbandingan jumlah dokumen relevan yang didapatkan sistem dengan jumlah seluruh dokumen relevan yang ada dalam koleksi dokumen (terambil ataupun tak terambil sistem). Menurut pengertian ini dapat dinyatakan dalam persamaan

$$P = \frac{TP}{TP+FN} \dots\dots\dots(2.6)$$

Dimana :

R (Recall) = Tingkat *recall*

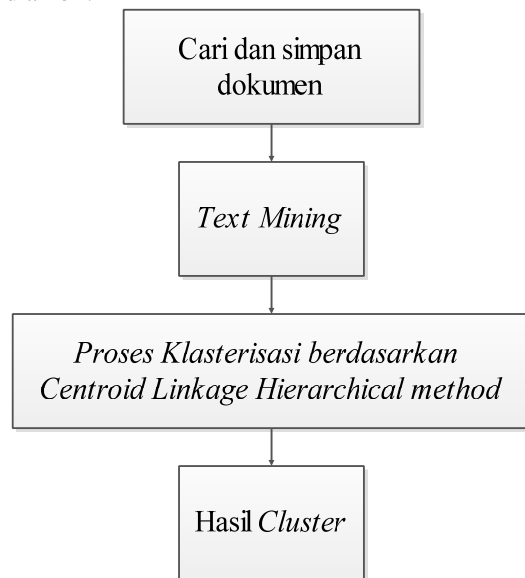
FN (False Negative) = Dokumen tidak relevan yang tidak ditemukan

METODE

Pada penelitian ini pengambilan dokumen berasal beberapa situs online. Dokumen yang telah diambil secara langsung akan tersimpan dalam format txt. Dokumen yang telah diambil akan melalui tahap temu kembali informasi yaitu *parsing*, *tokenizing*, dan *analisa sematik*. Setelah itu akan dihasilkan kumpulan kata istilah hasil dari temu kembali informasi. Maka akan diketahui berapa jumlah kata kunci istilah pada masing

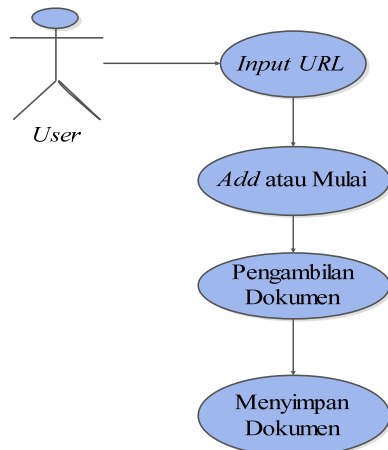
– masing dokumen. Setelah itu akan dilakukan klusterisasi dengan metode *centroid linkage hierarchical method* dan akan dihitung varian dari masing – masing pembentukan *cluster* yang dihasilkan untuk ditentukan *cluster* mana yang pembentukannya merupakan *cluster* terbaik. Dari hasil tersebut kemudian dilakukan pencocokan terhadap kata kunci yang dimasukkan oleh *user* dan dokumen yang dicari akan ditampilkan dalam *form* hasil pencarian.

Inputan dalam aplikasi ini disini adalah kata kunci yang dimasukkan oleh *user*. Kata kunci tersebut akan dicocokkan dengan sejumlah dokumen yang didapatkan dari berbagai situs online. Pada gambar 2 dibawah ini merupakan gambaran secara garis besar dari pencarian informasi di dalam dokumen yang telah diambil.



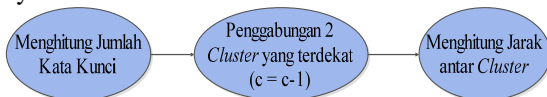
Gambar 2 Blok Diagram Utama

Gambar 3 di bawah ini merupakan *use case diagram* proses pencarian dan penyimpanan dokumen yang diambil dari situs *online* (rssDetik, rssOkezone, rssRepublika, dan rssTempo) yang *up to date*.



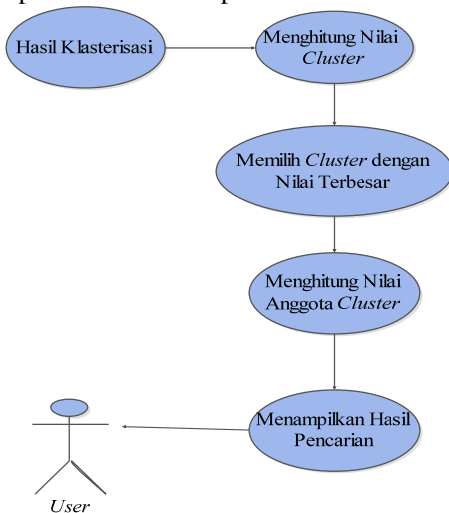
Gambar 3 Use Case Diagram Pencarian dan Penyimpanan Dokumen secara online

Gambar 4 di bawah ini merupakan proses klasterisasi menggunakan *Centroid Linkage Hierarchical Method (CLHM)*. Kata kunci istilah yang di *input* kan oleh *user* akan dicari jumlahnya kemudian dijumlahkan lalu akan ditentukan *cluster* nya.



Gambar 4 Use Case Diagram Clustering dengan Centroid Linkage Hierarchical Method

Gambar 5 di bawah ini merupakan *Use Case Diagram* Hasil Pencarian Dokumen. Pada proses klasterisasi ini akan dihitung nilai *cluster* yang mempunyai nilai *cluster* terbesar kemudian akan ditampilkan dalam hasil pencarian.



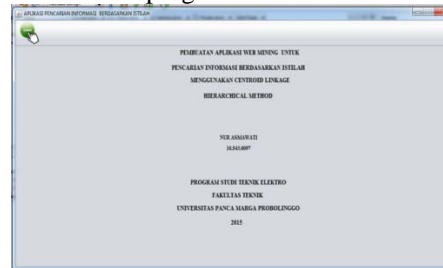
Gambar 5 Use Case Diagram Hasil Pencarian Dokumen

HASIL DAN PEMBAHASAN

Dalam bagian ini membahas tentang aplikasi yang dibuat yang terdiri dari tampilan, cara kerja serta pembahasannya.

Halaman Utama Aplikasi

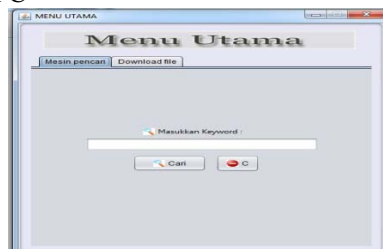
Pada gambar 6 di bawah ini ditunjukkan GUI halaman utama aplikasi . di dalam tampilan halaman utama aplikasi ini terdapat menu yang di dalamnya berisi menu proses dan menu keluar. Pada menu proses berisi menu utama yang di dalamnya terdapat form mesin pencari dan form *download file* serta menu keluar berisi perintah untuk keluar dari program.



Gambar 6 Halaman utama aplikasi

Halaman Pencarian Dokumen

Pada gambar 7 di bawah ini merupakan tampilan pencarian dokumen. Pertama *user* terlebih dahulu menginputkan kata kunci istilah yang akan dicari pada *text field* kemudian klik tombol cari yang berfungsi untuk memproses pencarian dokumen. Sedangkan untuk mengkosongkan *text field* cukup dengan menekan tombol C



Gambar 7 Halaman untuk Mesin Pencarian Istilah Tertentu

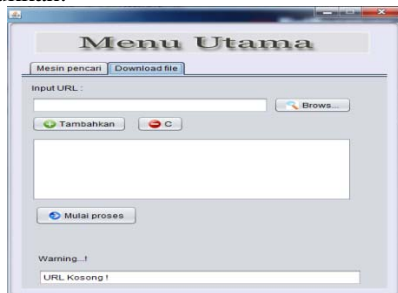
Keterangan untuk gambar 7

1. *Text field* masukkan *keyword* berfungsi sebagai tempat *input* kata kunci istilah yang akan di cari.
2. Tombol cari berfungsi untuk melakukan pencarian.

3. Tombol C berfungsi untuk menghapus atau mengkosongkan *text field*.

Halaman Download File

Pada gambar 8 di bawah ini merupakan tampilan menu *download file* yang berfungsi untuk mengambil dokumen secara *online*. Pada *text file* masukkan alamat *url* yang akan diambil dokumennya. Lalu tekan tombol tambah yang berfungsi untuk memasukkan alamat yang akan dicari dan untuk melakukan pengecekan apakah *url* yang dimasukkan merupakan *url rss* atau bukan *rss*. Selanjutnya tekan tombol proses, tunggu beberapa detik lalu hasil *pendownloadan* dokumen akan berlangsung. Sedangkan *text area* digunakan untuk menampilkan alamat *url* yang ada di dalam *list* yang telah disimpan. Tombol C berfungsi untuk mengkosongkan atau membersihkan alamat inputan *url* yang telah dimasukkan.



Gambar 8 Halaman untuk Download File penulis

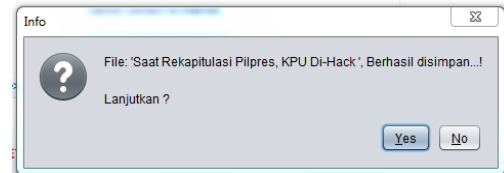
Keterangan untuk gambar 8 adalah sebagai berikut :

1. *Text field* untuk *input url* berfungsi untuk tempat penginputan *rss* yang akan dicari.
2. Tombol *browse..* berfungsi untuk mencari *rss* yang telah tersimpan.
3. Tombol tambah berfungsi untuk menambahkan *rss* yang telah di inputkan.
4. Tombol C berfungsi untuk menghapus alamat *rss* pada *text field input url*.
5. Tombol mulai proses berfungsi untuk memulai proses *pendownloadan rss* yang telah di inputkan.
6. *Text field* pada *warning* berfungsi untuk memberikan keterangan apakah *rss* yang di inputkan berhasil ditambahkan atau tidak

Halaman Tampilan Info

Pada gambar 9 di bawah ini merupakan halaman tampilan info yang menyajikan informasi mengenai file dokumen yang akan di *download*, di

dalam info ini berisi 2 pilihan perintah yaitu *yes* dan *no*. Perintah *yes* disini berfungsi untuk melanjutkan *download* dokumen yang akan diambil sedangkan perintah *no* berfungsi untuk menghentikan *download* dokumen yang akan di *download*.



Gambar 9 Halaman tampilan info

Halaman Hasil Pencarian Dokumen

Pada gambar 10 di bawah ini merupakan halaman hasil pencarian dokumen yang telah di *download* dan di simpan.



Gambar 10 Hasil Pencarian

Keterangan untuk gambar 10 adalah sebagai berikut :

1. *Cluster optimum* menunjukkan jumlah *cluster* yang terbentuk dari pencarian dokumen.
2. Nilai varian berfungsi untuk menunjukkan jumlah *cluster*.
3. *Cluster anggota* berfungsi untuk menunjukkan jumlah dokumen.

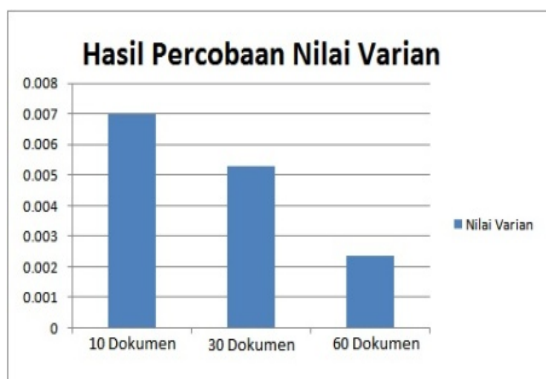
4. Beban berfungsi untuk menunjukkan jumlah kata kunci yang di cari.
5. Jumlah anggota *cluster* berfungsi untuk menunjukkan jumlah anggota *cluster* yang terbentuk pada *cluster* tersebut.
6. Waktu yang diperlukan selama proses menunjukkan lamanya waktu yang ditempuh selama pencarian dokumen berlangsung mulai dari peng inputan kata kunci istilah yang dicari sampai ditemukannya hasil pencarian

Uji Coba dan Evaluasi

Dari hasil percobaan dapat dituliskan dalam tabel 1.

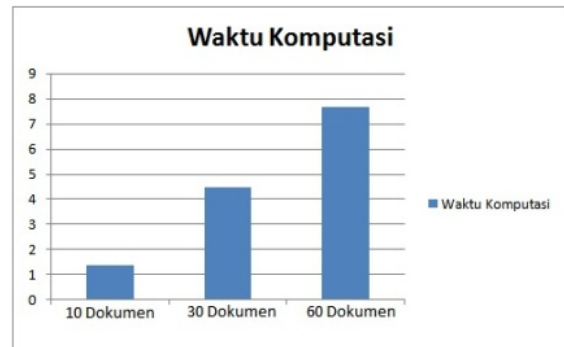
Tabel 1 hasil Percobaan

No	Jumlah Dokumen	Nilai Varian	Waktu yang diperlukan
1.	10	0,00698	1,386detik
2.	30	0.00529	4,464 detik
3.	60	0,00238	7,644 detik



Gambar 11 Grafik hasil percobaan nilai varian

Berdasarkan grafik yang dihasilkan, nilai varian mengalami penurunan ketika jumlah dokumen semakin banyak. Sebaliknya nilai varian akan semakin meningkat apabila dokumen semakin sedikit. Hal ini membuktikan bahwa pada dokumen yang semakin banyak maka kemungkinan terjadi pembentukan cluster secara otomatis akan semakin besar.



Gambar 12 Grafik waktu komputasi atau waktu yang diperlukan selama proses

Sedangkan nilai waktu yang diperlukan selama proses adalah meningkat seiring dengan bertambahnya jumlah dokumen. Hal ini dikarenakan ketika jumlah data semakin besar, maka tahapan proses dalam klasterisasi menggunakan metode ini akan semakin panjang.

Precisian dan Recall

Tabel 2 Hasil pengujian dengan *Centroid Linkage Hierarchical Method*

No	Jumlah dokumen	Precisian	Recall	Waktu yang diperlukan
1	10 dokumen	0,33	0,125	0,822 detik
2	30 dokumen	0,25	0,030	2,842 detik
3	60 dokumen	0,25	0,035	6,401 detik
Rata – rata		0,27	0,06	3,355 detik

Dari tabel 2 didapatkan *precisian, recall* dan waktu yang diperlukan pada sejumlah dokumen yang berbeda. Dari tabel tersebut didapat nilai rata – rata *Precisian* 0,27, *Recall* 0,06 dan waktu yang diperlukan 3,355 detik

Dari sejumlah dokumen yang berbeda terlihat bahwa nilai *precision* dan *recall* menunjukkan angka yang tetap. Namun nilai waktu komputasi meningkat seiring dengan bertambahnya jumlah dokumen. Hal ini dikarenakan ketika jumlah data semakin besar, maka tahapan proses dalam klasterisasi menggunakan metode ini akan semakin panjang.

SIMPULAN

Berdasarkan hasil dan pembahasan yang telah dilakukan oleh penulis, maka dapat diambil kesimpulan :

1. Berdasarkan hasil uji coba dan evaluasi yang dilakukan oleh penulis, dengan menggunakan *centroid linkage hierarchical method* dapat digunakan untuk pencarian informasi berdasarkan istilah tertentu baik berupa dokumen yang telah tersimpan maupun dokumen yang akan di *download* melalui *rss*.
2. Lamanya waktu pencarian tergantung dari jumlah dokumen yang tersimpan.
3. Lamanya waktu yang diperlukan untuk mendownload dokumen tergantung dari koneksi *internet* yang tersedia
4. Nilai varian mengalami penurunan ketika jumlah dokumen semakin banyak. Sebaliknya nilai varian akan semakin meningkat apabila dokumen semakin sedikit. Hal ini membuktikan bahwa pada dokumen yang semakin banyak maka kemungkinan terjadi pembentukan cluster secara otomatis akan semakin besar.
5. Sedangkan nilai waktu yang diperlukan selama proses adalah meningkat seiring dengan bertambahnya jumlah dokumen. Hal ini dikarenakan ketika jumlah data semakin besar, maka tahapan proses dalam klasterisasi menggunakan metode ini akan semakin panjang.

DAFTAR PUSTAKA

- [1] Damayanti, Nadia, "Temu Kembali Informasi Teks Berdasarkan Lokasi Pada Dokumen Yang Dikelompokkan Menggunakan Metode Centroid Linkage Hierarchical Method", *Penelitian D4, Politeknik Elektronika Negeri Surabaya*, 2011
- [2] E. Martiana, N. R. Muftada, U. Aguseta, "Mesin Pencari Dokumen Dengan Pengklasteran Secara Otomatis", *Politeknik Elektronika Negeri Surabaya, Telkomnika*, 2010
- [3] Hermawawati, fajar Astuti. 2013. " *Data Mining*", Yogyakarta : Penerbit Andi
- [4] H.M.Jogiyanto. 1997. " Sistem Informasi berbasis Komputer Edisi 2 Konsep dasar dan

Komponen "Yogyakarta : BPFE – Yogyakarta

- [5] Syarief, Mulkan. 2012. " *Berbagai Project Java dengan IDE NetBeans*", Yogyakarta : Penerbit Andi
- WK, Nikodemus. 2011. "Paling dicari *Java Source Code*", Yogyakarta : Penerbit Andi
- [6] http://Pengertian%20Dan%20Fungsi%20Web%20Crawler%20_%20blog%20mas%20hendra.html.