

“PENCARIAN DOKUMEN BERBASIS WEB PADA DRIVE LOKAL DAN OFF-LINE WEB DENGAN MENGGUNAKAN METODE SUFFIX TREE CLUSTERING”.

Zaenal Fanani
Wiji Setyaningsih

¹ Sistem Informasi, Universitas Kanjuruhan Malang, mikazen@gmail.com

² Sistem Informasi, Universitas Kanjuruhan Malang, wiji_setiya@yahoo.co.id

ABSTRAK

Semakin kaya bahasa dan banyaknya koleksi dokumen teks, sistem pencarian merupakan hal yang sangat penting. Tetapi kebanyakan mesin pencarian tidak memberikan kepuasan dalam menemukan informasi yang dicari. Terkadang Pengguna terpaksa melakukan kerja dobel dengan memilah lagi dokumen yang sudah dicari dalam daftar yang panjang. Ini tentunya akan menyita waktu. Banyak metode yang dikembangkan untuk proses pencarian, salah satunya dengan menggunakan model *clustering* untuk mengelompokkan hasil pencarian dokumen sesuai dengan hasil yang diinginkan. Dalam pengerjaan skripsi ini, penulis menggunakan metode Suffix Tree Clustering. Adapun hasil yang didapatkan adalah menggali lebih dalam informasi, lebih singkat, efektif dan tidak memerlukan waktu lama. Penulis melakukan proses import dokumen terlebih dahulu, dan kemudian akan diproses secara *offline*. Hal ini menghindari pencarian berupa dokumen dokumen yang tersusun berdasarkan peringkat kecocokan dalam daftar yang panjang. Singkatnya tidak memperlakukan dokumen sebagai himpunan kata-kata tetapi sebagai *string*.

Kata Kunci Text Mining, Mesin Pencari, Web-Offline, Suffix Tree Clustering, Pengelompokan dokumen, String, Import Dokumen

ABSTRACT

With the richer of language and many collections of text documents, the search system is very important. But most of search engines do not give satisfaction in finding the information. Users sometimes have to do twice work with another sort of documents that have been sought in a long list. It will certainly take time. Many methods are developed for the search process, one of them uses clustering model to classify documents according to the desired results. In this thesis, the author uses the method of Suffix Tree Clustering. The results obtained are digging deeper into the information, more concise, effective and no need in long time. The author performs the process of importing file in advance, and then will be process it offline. This avoids search of structured documents based on ranked matches in a long list. This means do not treat the document as a set of words but as a string.

Likewise, according to Qualls Report (Journal) entitled "Fast and Intuitive Clustering of Web Documents" by Oren Zamir (1997), it states that the search data on a local drive or a web-offline that is generally only file names that can be searched. Users are required to find more details by themselves as well as the need to open the file, so that the search process is often long

and tedious. Based on the description above, the author entitles "The Web-Based Document Search Local Drive and Off-line Web Method Using Suffix Tree Clustering"..

Keywords: Text Mining, Search engine, Web-Offline, Suffix Tree Clustering, Grouping Document, String, Import Document

1. Pendahuluan

Pada umumnya sistem pencarian dokumen menampilkan hasil pencarian dalam daftar yang panjang berdasarkan peringkatnya. Kemudian pengguna diharuskan memilah sendiri dokumen mana yang relevan dengan topik yang ia cari dalam daftar tersebut. Sayangnya sebagian besar *search engine* menggunakan paradigma tersebut. Selain itu *search engine* juga memiliki karakteristik yaitu memiliki presisi hasil pencarian yang rendah. Kedua kelemahan *search engine* tersebut membuat pengguna cukup kesulitan untuk menemukan informasi dari dokumen yang dicari (Aulia Adi Pribadi, 2012).

Melanjutkan riset yang dilakukan oleh Aulia Adi Pribadi dan Entin Martiana K. (2012), "Pencarian Judul TA Menggunakan *Text Mining dan Metode Suffix Tree*", yang melakukan proyek mengembangkan sebuah aplikasi pengelompokan dokumen berbasis web dengan metode *suffix tree clustering*. Aplikasi membutuhkan *input* pencarian dan akan menghasilkan *output* berupa *cluster* yang di dalamnya terdapat dokumen yang bersesuaian. *Cluster* ini bisa bertingkat-tingkat tergantung dari kata atau phrase yang mungkin bisa dibedakan lagi pada cluster induk yang sama. *Cluster-cluster* yang dihasilkan inilah yang ditampilkan kepada pengguna. Selanjutnya pada cluster terakhir yang dipilih akan menampilkan kumpulan dokumen yang masing-masing terdiri dari judul dan cuplikan.

Riset yang dilakukan oleh Aulia Adi Pribadi dan Entin Martiana K, ini berbasis web online. Kelemahan dari sistem online tentunya tergantung pada koneksi internet dan juga waktu pencarian. Dalam riset tersebut hanya sebatas pencarian judul TA,

berdasar input kalimat atau judul yang dicari. Hasil pencarian pun terbatas hanya nama *file*-nya saja, dengan presisi hasil pencarian rendah, user harus memilah sendiri dengan membuka satu-satu *file* yang ditemukan. Tentunya ini kurang efektif dan memerlukan waktu tidak sedikit dan mungkin akan menjemukan bagi user tersebut. Berdasarkan uraian diatas, maka penulis mengangkat judul "Pencarian Dokumen Berbasis Web Pada *Drive* Lokal Dan *Off-line Web* Dengan Menggunakan Metode *Suffix Tree Clustering*"

2. Tinjauan Pustaka

Ada sebuah cara yang cukup membantu untuk mendapatkan hasil pencarian dokumen dengan tepat dan mudah. Yaitu dengan menggunakan model *clustering* untuk mengelompokkan hasil pencarian dokumen sesuai dengan topik yang terkait. Sehingga Penulis mengembangkan menjadi proyek pencarian data pada *drive* lokal maupun *web-offline*. Dengan cara melakukan proses import terlebih dahulu sehingga bisa menggali lebih dalam dan proses pencarian dapat lebih singkat, efektif dan tidak memerlukan waktu lama. Proses offline menghindari pencarian berupa dokumen dokumen yang tersusun berdasarkan peringkat kecocokan dalam daftar yang panjang. Singkatnya tidak memperlakukan dokumen sebagai himpunan kata-kata tetapi sebagai string. Demikian juga menurut Qualls Report (Jurnal) berjudul "*Fast and Intuitive Clustering of Web Documents*" oleh Oren Zamir (1997), menyatakan bahwa pencarian data pada *drive* lokal maupun *web-offline* umumnya hanya nama file yang bisa dicari dan pengguna diharuskan mencari sendiri lebih detail serta harus membuka *file*

tersebut, sehingga proses pencarian seringkali lama dan menjemukan.

3. Pembahasan

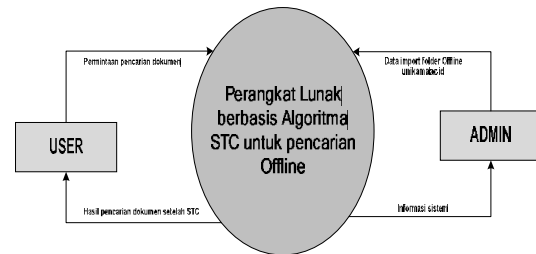
Analisis sistem didefinisikan sebagai penguraian dari sistem utama ke dalam sub-sistem dengan tujuan untuk mengidentifikasi permasalahan-permasalahan yang ada dan kebutuhan-kebutuhan yang diperlukan agar dapat diusulkan dan diciptakan sistem baru yang lebih baik. Analisis terhadap sistem lama yang sedang berjalan perlu dilakukan sebagai dasar dari perancangan sistem baru, agar dapat dibuat sistem yang lebih efektif dan efisien.

Riset lama yang dilakukan oleh Aulia Adi Pribadi dan Entin Martiana K (2012), ini berbasis web online. Kelemahan dari sistem online tentunya tergantung pada koneksi internet dan juga waktu pencarian. Dalam riset tersebut hanya sebatas pencarian judul TA, berdasar input kalimat atau judul yang dicari. Hasil pencarian pun terbatas hanya nama file-nya saja, dengan presisi hasil pencarian rendah, user harus memilah sendiri dengan membuka satu-satu file yang ditemukan. Tentunya ini kurang efektif dan memerlukan waktu tidak sedikit dan mungkin akan menjemukan bagi user tersebut.

Ada sebuah cara yang cukup membantu untuk mendapatkan hasil pencarian dokumen dengan tepat dan mudah. Yaitu dengan menggunakan model clustering untuk mengelompokkan hasil pencarian dokumen sesuai dengan topik yang terkait. Sehingga Penulis mengembangkan menjadi proyek pencarian data pada drive lokal maupun web-offline. Dengan cara melakukan proses import terlebih dahulu sehingga bisa menggali lebih dalam dan proses pencarian dapat lebih singkat, efektif dan tidak memerlukan waktu lama. Proses offline menghindari pencarian berupa dokumen-dokumen yang tersusun berdasarkan peringkat kecocokan dalam daftar yang panjang.

3.1 Desain Sistem

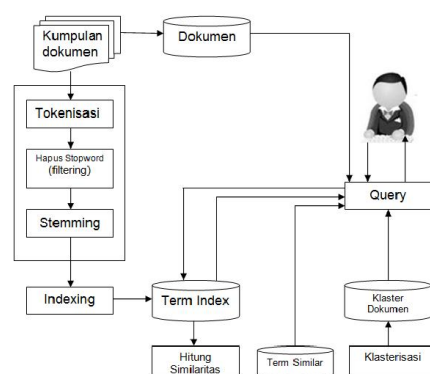
Diagram konteks terdapat bagian-bagian penting yaitu, beberapa entitas, proses dan arus data. Berikut adalah diagram konteksnya :



Gambar 1 Diagram Konteks

Gambar 2 Blok Sistem

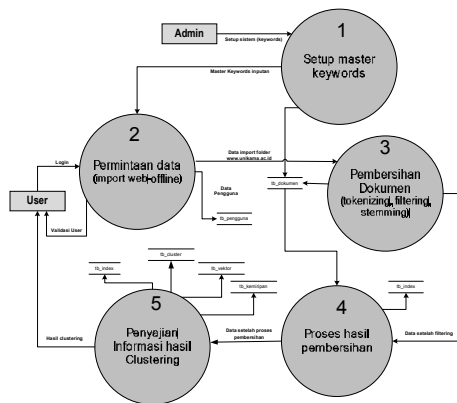
Pada gambar diatas, Proses *text mining* dan pembobotan maupun menentukan cluster akan dijalankan saat user menginputkan *query* pada form pencarian, proses *text mining* akan selalu mengolah data aktual, dengan demikian jika terdapat perubahan pada dokumen, perhitungan yang dihasilkan oleh program adalah perhitungan yang bersifat aktual.



Gambar 3 Arsitektur pencarian dokumen berbasis STC

Pada Gambar di atas menjelaskan bahwa aliran proses *Suffix Tree Clustering* dimulai dari proses *Tokenizing* yaitu memecah paragraf menjadi kalimat, kemudian menjadi kata.

Sedangkan *Stopword (Filtering)* adalah mengambil kata-kata yang penting, dilanjutkan para proses *Stemming*, yaitu mengambil kata dasar dari kata penting dari proses sebelumnya. Kemudian berlanjut pada proses *Analyzing*, mengelompokkan kata dan melakukan pembobotan sesuai muatan. Muatan inilah yang nantinya akan diolah berdasar *index* untuk proses pencarian secara *offline* menggunakan metode *Suffix Tree Clustering*.

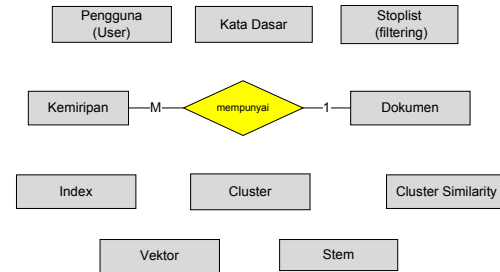


Gambar 4 DFD Level 1

Secara garis besar dalam DFD level 1 mempunyai lima proses yang merupakan proses utama pada sistem, yaitu proses admin yang melakukan setup master keywords, kemudian proses permintaan data yaitu melakukan download atau import data web-offline (dalam proyek ini mengambil sample data www.unikama.ac.id). Admin adalah seorang yang mempunyai hak akses penuh dalam sistem ini.

Kemudian proses menunjukkan bahwa aliran proses *Suffix Tree Clustering* dimulai dari proses *Tokenizing* yaitu memecah paragraf menjadi kalimat, kemudian menjadi kata. Kemudian proses *Filtering* dan *Stemming*. Kemudian *Analyzing*, mengelompokkan kata dan melakukan

pembobotan sesuai muatan. Kemudian proses terakhir yaitu *Kombinasi Base Cluster* yaitu menyesuaikan makna dari satu kata dengan yang lain berdasar kluster dasar.



Gambar 5 ERD

Pembangunan basis data memerlukan 10 tabel, 5 tabel (*tb_index*, *tb_cluster*, *tb_cluster_similarity*, *tb_vektor*, *tb_kemiripan*) berfungsi sebagai tabel proses pengolahan teks yang digunakan oleh algoritma *Suffix Tree Clustering* atau proses teks mining, dan 5 tabel digunakan untuk menyimpan data-data keperluan website *offline* dalam hal ini www.unikama.ac.id.

Berikut ini adalah implementasi antarmuka yang digunakan:

Prosedur penggunaan aplikasi, sebagai berikut :

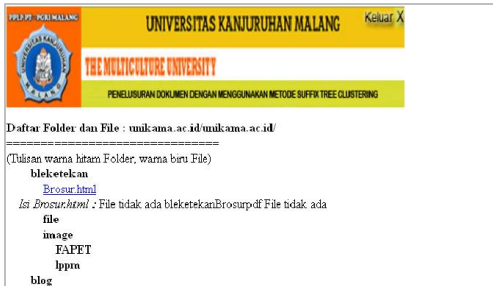
- Mengaktifkan browser apa saja
- Mengetik : localhost/stc/index.php
- Login dengan username admin, bisa mengatur user/ pengguna aplikasi

a. Halaman *Login Username*

Gambar 6 Login Username

Pada halaman ini akan menampilkan proses login username, yang terdiri dar 2 (dua) hak akses, yaitu sebagai admin atau user biasa. Keduanya memiliki perbedaan wewenang dalam aplikasi pencarian ini.

b. Halaman Proses *Import*



Gambar 7 Proses *Import*

Pada halaman ini akan menampilkan proses import dari www.unikama.ac.id, yaitu menampilkan file dan folder serta sub folder.

c. Halaman Penelusuran Dokumen



Gambar 8 Penelusuran Dokumen

Pada halaman ini akan menampilkan tentang pencarian dokumen, yang sebelumnya user harus melakukan proses secara berurutan dari import dokumen Offline hingga proses pencarian.

d. Halaman Pencarian



Gambar 9 Pencarian Sesuai Inputan

Pada halaman ini menampilkan proses pencarian dalam *web offline*. Kalimat

yang dicari : Menjalin kerjasama dengan organisasi alumni universitas Kanjuruhan Malang dalam rangka memberdayakan alumni, yang dilaksanakan oleh *file*.

e. Halaman *Preprocessing*



Gambar 9 *Preprocessing*

Pada halaman ini akan menampilkan preprocessing menghapus semua kata penghubung dan kata ganda menjadi satu kata yang unik

4. Kesimpulan

Berdasarkan hasil percobaan dan analisa yang dilakukan pada Tugas Akhir ini, maka dapat disimpulkan bahwa :

1. Telah dibuat aplikasi pencarian dokumen berbasis web pada drive lokal dan off-line web dengan menggunakan metode *Suffix Tree Clustering* \menggunakan bahasa pemrograman PHP dan MySQL dengan fitur pencarian, pembuatan index, pembuatan bobot dan beberapa fitur tambahan lainnya.
2. Aplikasi yang dibuat dapat digunakan untuk melakukan pencarian dokumen berbasis web, melakukan penerapan metode *Suffix Tree Clustering* untuk pengelompokan dokumen hasil pencarian pada drive lokal dan web off-line secara efektif.
3. Sistem yang sudah dibangun dapat mengefisienkan proses pencarian dokumen secara efektif dan optimal.
4. STC merupakan metode untuk melakukan *index*. Ini memungkinkan pencarian kata bisa dilakukan lebih cepat.

5. Nilai STC menentukan makna dari inputan user.

5. Saran

Berikut merupakan beberapa saran untuk pengembangan sistem di masa yang akan datang, berdasar pada hasil perancangan, implementasi, dan uji coba yang telah dilakukan :

1. Kiranya perlu dilakukan pengembangan aplikasi pencarian dokumen yang lebih efektif dan efisien, seperti dengan menambahkan dan menggali lebih dalam lagi mengenai data-data dan kriteria-kriteria lainnya yang berhubungan dengan metode *Suffix Tree Clustering*
2. Sistem ini hanya dapat digunakan dengan bahasa yang baku, diharapkan di masa mendatang akan bisa lebih dikembangkan lagi yang juga dapat menggunakan bahasa yang tidak baku.
3. Menyempurnakan dan dilakukan pengembangan program sejenis dengan permasalahan yang lebih luas.
4. Pada sistem ini, bahasa yang harus digunakan user harus menggunakan bahasa Indonesia baku. Jika tidak, maka data solusi yang ditampilkan tidak sesuai dengan harapan user.

6. Daftar Pustaka

Adi Nugroho, ST., MMSI. 2004. *Konsep Pengembangan Sistem Basis Data*. Informatika . Bandung.

Alfa Firdaus, ST., MT. 2013. *Pusat Pengembangan Bahan Ajar UMB*. <http://juliadi.wikispaces.com>. (di akses Tanggal, 17 Maret 2014)

Ayuliana, ST., MMSI. 2011. <http://Thesis.binus.ac.id>. (di akses Tanggal, 17 Maret 2014)

Eva Yulianti, S.Kom.,M.Cs. 2011. *Sistem Pengambilan Keputusan*. <http://sisfo.itp.ac.id>. (di akses Tanggal, 17 Maret 2014)

Fatansyah. 1999. *Basis Data*. Informatika . Bandung.

Hengki Prihatna. 2005. *Webmaster Profesional*.Komputindo. Jakarta.

Kadir, A. 2001. *Dasar Pemrograman Web Dinamis Menggunakan PHP*. ANDI .Yogyakarta.

Kendall dan Kendall. 2003. *Analisis dan Perancangan Sistem Jilid 1 Edisi Kelima*. PT. Prenhalindo . Jakarta.

Khoirul Anwar, S.T., M.T. 2012. *Metode TOPSIS* [http:// staff.pradnya.ac.id](http://staff.pradnya.ac.id). (di akses Tanggal, 17 Maret 2014)

Nanang. 2013. *Modul Ajar Testing dan Implementasi*. Universitas Kanjuruhan Malang.

Noviyanto, ST. 2011. *Sistem Penunjang Keputusan* <http://viyan.staff.gunadarma.ac.id>. (di akses Tanggal, 17 Maret 2014)

Rumahweb.2014.Nilai. <http://rumahweb.com>. (di akses Tanggal, 17 Maret 2014)

Saliman. 2012. *Mengenal Decision Support System (DSS)*. [http:// staff.uny.ac.id](http://staff.uny.ac.id). (di akses Tanggal, 17 Maret 2014)

Sumaryanto. 2011. *Upaya Pengambilan Keputusan Yang Tepat*. <http://staff.uny.ac.id>. (di akses Tanggal, 17 Maret 2014)

Wiji Setyaningsih, M.Kom. 2013. *Modul Ajar Decision Support System*. Universitas Kanjuruhan Malang.

