

Pencarian Data Alumni STMIK Jakarta STI&K dengan Metode Suffix Tree

Dewi Nur Cahyani, Ire Puspa Wardhani, Susi Widayati dan Irfan

STMIK JAKARTA STI&K

Jl. BRI No. 17 Radio Dalam Kebayoran Baru – Jakarta 12140
E-mail: dewinurcahyani54@gmail.com, irewardhani@jak-stik.ac.id,
widayatiususi@gmail.co.id, irfan@jak-stik.ac.id

Abstrak

Semakin bertambahnya jumlah alumni yang ada di STMIK Jakarta STI&K mempengaruhi banyaknya jumlah informasi dan menyebabkan kesulitan dalam proses temu kembali informasi data alumni. Sistem pencarian merupakan hal yang sangat penting. Tetapi terkadang pengguna melakukan kerja berulang kali dengan memilih lagi dokumen yang sudah dicari dalam daftar panjang dan tentunya akan menyita banyak waktu. Banyak metode yang dikembangkan untuk proses pencarian, salah satunya dengan metode clustering untuk mengelompokkan dokumen sesuai dengan hasil yang diinginkan. Dalam penelitian ini penulis menggunakan metode Suffix Tree Clustering. Adapun hasil yang didapatkan adalah menggali lebih dalam informasi, lebih singkat karena tidak perlu lagi menggunakan prosedur yang panjang secara manual dengan ketepatan mendapatkan data yang diinginkan. Hal ini menghindari pencarian berupa dokumen-dokumen yang tersusun berdasarkan peringkat kecocokan dalam daftar yang panjang

Kata Kunci : Mesin Pencari, Pengelompokan Dokumen, Suffix Tree Clustering

Pendahuluan

Pada generasi serba elektronik ini, informasi telah menjadi kebutuhan vital bagi setiap organisasi, baik bagi organisasi pemerintah maupun dalam dunia bisnis, perusahaan menerima pegawai baru dengan kriteria tertentu, kemudian ingin membuktikan bawa apakah betul seseorang tersebut merupakan lulusan dari universitas / perguruan tinggi sesuai dengan ijazah yang di berikan. Informasi menjadi suatu bagian yang sangat penting untuk mendukung proses kerja administrasi dalam pelaksanaan fungsi-fungsi manajemen dari birokrasi di dalam menghadapi perubahan situasi dan kondisi yang berkembang dengan cepat, perusahaan dapat membuktikan kebutuhan data dengan melihat sendiri di daftar nama lulusan pada perguruan tinggi tersebut. Permasalahan Pencarian Data Alumni di STMIK Jakarta STI&K dengan kemudahan untuk pencarian dan penemuan kembali dokumen atau data alumni yang dibutuhkan proses scrolling lebih lama.

Pencarian dokumen menampilkan hasil pencarian dalam daftar yang panjang berdasarkan peringkatnya pada halaman web tertentu. Sehingga menimbulkan permasalahan klasik yaitu memerlukan proses page scrolling yang lebih lama pada saat proses pembacaan informasi yang ada. Kemudian pengguna diharuskan memilih sendiri dokumen mana yang relevan dengan topik yang dicari dalam daftar tersebut. Salah satu pemecahan permasalahan yang dapat dilakukan untuk mengurangi panjang halaman web tersebut adalah dengan cara melakukan pengelompokan teks berita yang ada secara tematik. Metode pengelompokan yang tepat untuk permasalahan ini adalah pengklasteran. Salah satu metode pengklasteran dokumen teks adalah Suffix Tree Clustering (STC). Metode ini mempunyai tingkat ketepatan yang sangat tinggi karena dalam pembentukan klaster berdasarkan pada phrase-shared di antara dokumen-dokumen teks yang ada.

Beberapa Metode Pengelompokan

Terdapat beberapa metode yang berhubungan dengan *clustering* pada pencarian dokumen. Diantara metode-metode tersebut adalah metode Suffix Cactus. Metode Suffix Cactus ini memiliki kesamaan struktur suffix tree dan pada suffix array. Suffix cactus dapat di deskripsikan sebagai versi compact (rapat) Suffix Tree atau Suffix Array yang diperbesar dengan informasi tambahan. Sehingga Suffix Cactus dapat disebut sebagai persilangan atau hibrid antara Suffix Tree dan Suffix Array [1].

Metode Suffix Tree Clustering (STC) memiliki dua kata kunci yaitu, pertama menggunakan parse dasar pembentukan kelompok, dan kedua menggunakan suatu definisi kelompok sederhana. Suffix Tree Clustering ini memiliki dua langkah utama [2]: Pertama adalah pencarian shared dan parse untuk semua dokumen berita yang dikoreksi. Disebut shared parse sebagai parse cluster atau base cluster, yang ditemukan dengan suatu struktur data yang dinamakan suffix tree. Kedua, Mengkombinasikan base cluster-basecluster ke dalam suatu cluster tertentu. Penggabungan antara dua base cluster didasarkan pada jumlah dokumen yang melakukan overlap diantara kedua base cluster tersebut. Suatu parse yang dimaksud dalam konteks algoritma ini adalah urutan satu atau lebih kata-kata.

Suffix Tree merupakan salah satu dari struktur data yang penting pada stringology. Suffix Tree adalah struktur index yang dibentuk dari string yang memungkinkan banyak query yang cepat terhadap berbagai string. Apa yang membuat Suffix Tree menarik adalah waktu dan ukuran konstruksi yang linear terhadap panjang teks [3]. Suffix Tree banyak diterapkan pada beragam aplikasi. Aplikasi tersebut merupakan implementasi Suffix Tree sebagai indeks teks statis berukuran besar yang memungkinkan pencarian cepat. Tipe pencarian dasarnya adalah pencocokan (matching) string seperti pencarian untuk kejadian dari sebuah pola string pada sebuah teks.

Metode Indexing dan Query

Indexing Retrieval merupakan teknik untuk dokumen teks yang banyak menggunakan teknik IR (Indexing dan Retrieval) karena [4]: Pertama, keberadaan sebagian besar dokumen

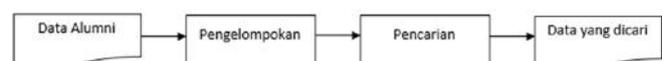
teks banyak digunakan di dalam organisasi seperti perpustakaan. Kedua, Teks adalah penting sebagai sumber informasi untuk organisasi. Untuk pengefisiensi penggunaan penyimpanan informasi dalam naskah sangat diperlukan sistem IR. Dan teks dapat digunakan untuk kebutuhan media lain seperti Audio, Gambar, dan Video. Tujuan dari suatu sistem IR adalah untuk mendapat kembali materi relevan dari suatu database dokumen merupakan jawaban atas query dari pemakai. Kebanyakan dari sistem IR yang komersil saat ini dapat digolongkan pada sistem IR boolean atau sistem pencarian text-pattern. Query pencarian text pattern adalah string atau sebuah ungkapan bersifat reguler.

Pengelompokan (Clustering)

Metode Clustering adalah metode yang memiliki kemampuan untuk menganalisis dan mengelompokkan secara otomatis dokumen-dokumen. Teknik Clustering ini umumnya menggunakan kata dan dokumen yang biasanya dianggap sebagai kumpulan kata-kata tanpa urutan atau yang disebut bag of word. Suffix Tree Clustering (STC) adalah algoritma pertama yang menggunakan frasa (multi word term) sehingga prosesnya lebih sederhana dibandingkan algoritma yang lain. [4]

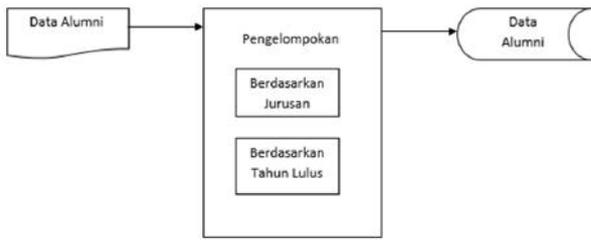
Metode Penelitian

Secara umum metode penelitian dalam penelitian ini mengikuti skema penelitian Pencaian Data Alumni STMIK Jakarta STI&K mengacu pada gambar 1.



Gambar 1: Skema penelitian

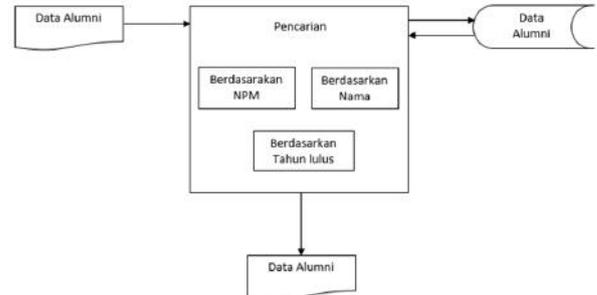
Awal proses yaitu pengelompokan data alumni berdasarkan suffix kelompok tertentu dan menghasilkan kelompok-kelompok data berdasarkan kategorinya, yaitu tahun lulus dan jurusan. Proses kedua adalah Pencarian data yaitu proses penemuan kembali data alumni dalam database berdasarkan kelompoknya dilihat pada gambar 2.



Gambar 2: Pengelompokan data alumni

Pada proses pencarian data alumni didasarkan atas nama, npm dan tahun lulus dan menghasilkan data yang diinginkan. Dapat dilihat pada gambar 3.

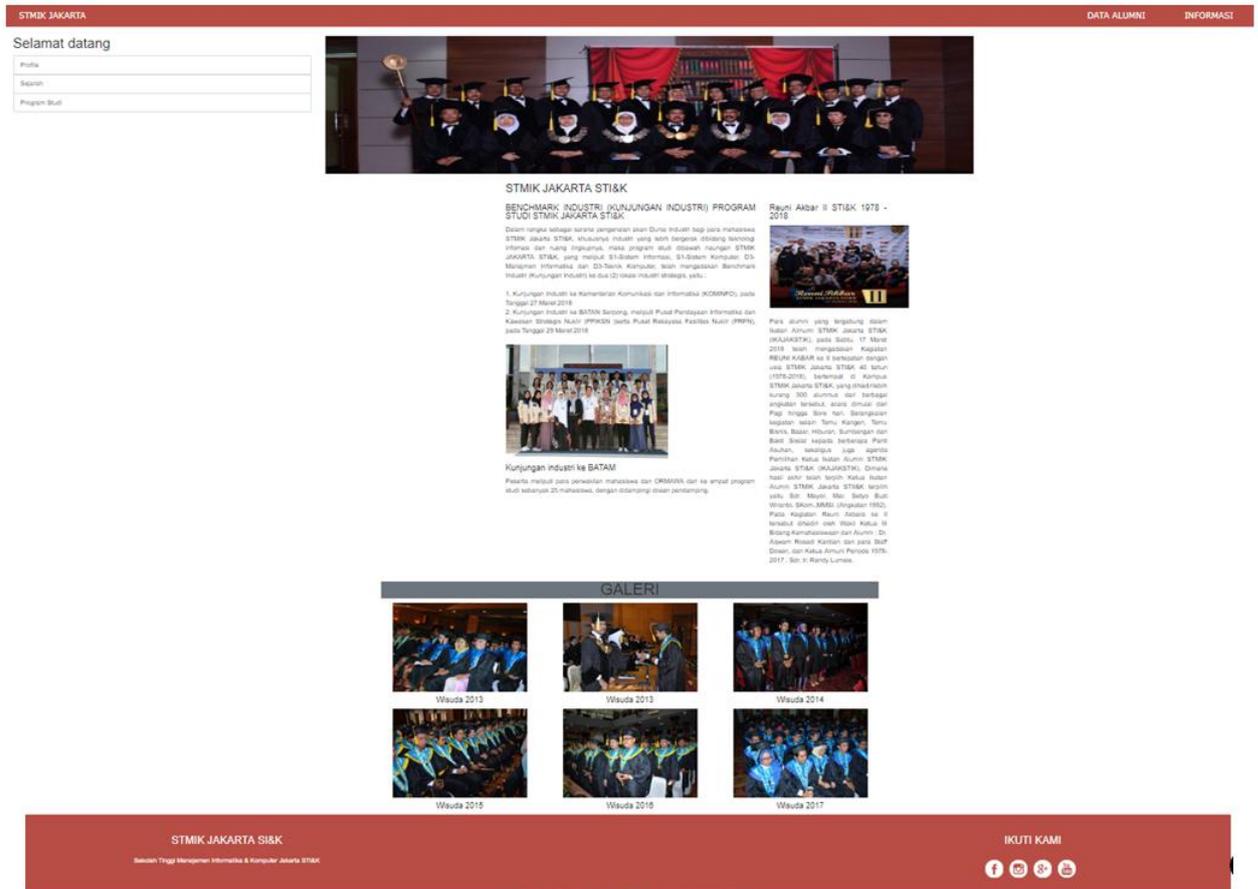
muncul sebagai tampilan awal saat website ini. Pada gambar 5 merupakan halaman profile dimana pada halaman ini pengunjung dapat mengetahui lebih jauh tentang kampus STMIK Jakarta STI&K. Pada gambar 6 diperlihatkan tentang pengunjung yang dapat melihat sejarah kampus STMIK Jakarta STI&K.



Gambar 3: Pencarian data alumni

Hasil dan Pembahasan

Halaman utama pada aplikasi ini terlihat pada gambar 4 merupakan halaman pertama



Gambar 4: Tampilan Halaman Utama



Gambar 5: Tampilan Halaman Profile

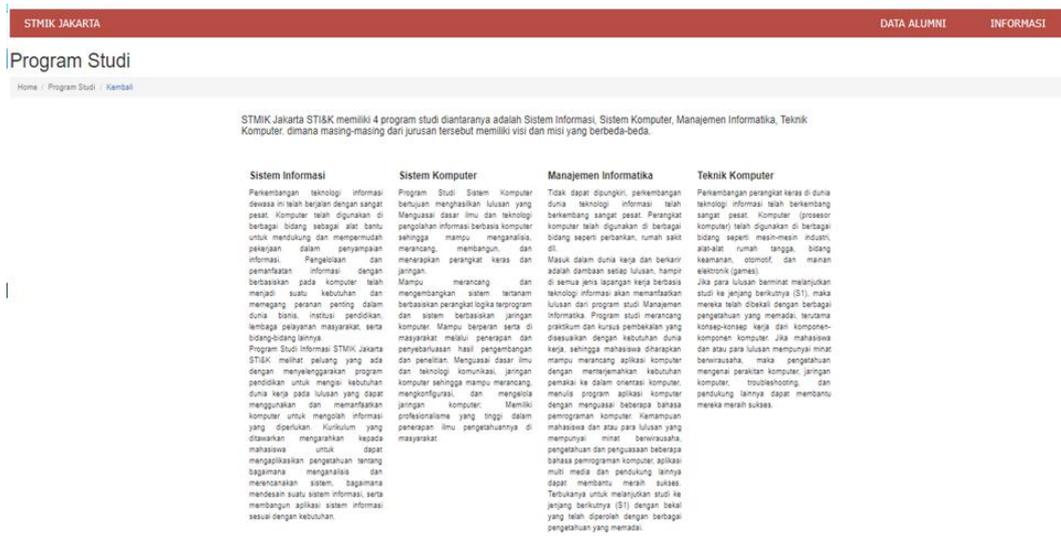


Gambar 6: Tampilan Halaman Sejarah

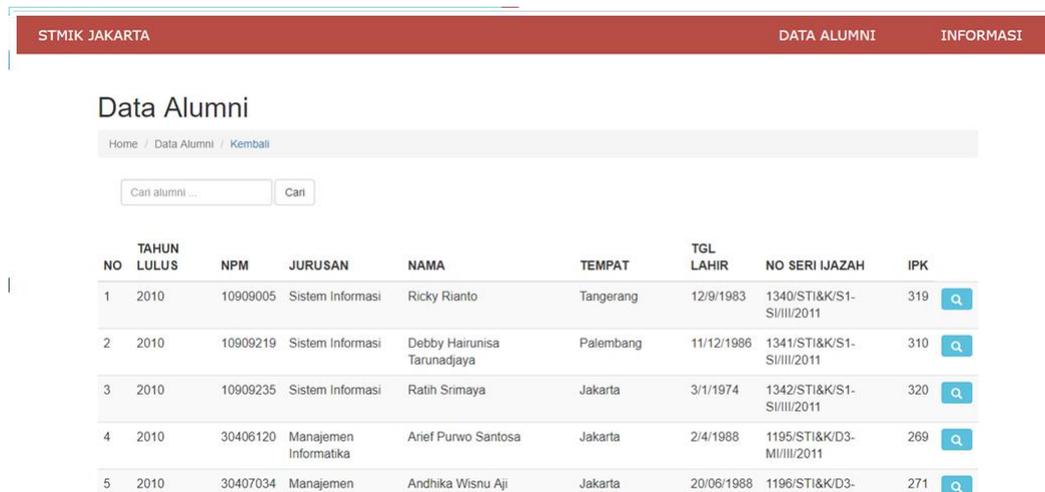
Pada gambar 7 terlihat bahwa tampilan halaman ini pengunjung dapat melihat apa saja program studi yang ada di kampus STMik Jakarta STI&K. Pada gambar 8 memperlihatkan tampilan halaman pengunjung yang dapat melihat data alumni lebih rinci tanpa harus login terlebih dahulu. Pada gambar 9 merupakan halaman dimana pengunjung dapat mencari data alumni lebih mudah dengan pilihan button search, gambar 10 memperlihatkan bahwa pengunjung dapat mengetahui informasi lebih detail tentang kampus STMik

Jakarta STI&K.

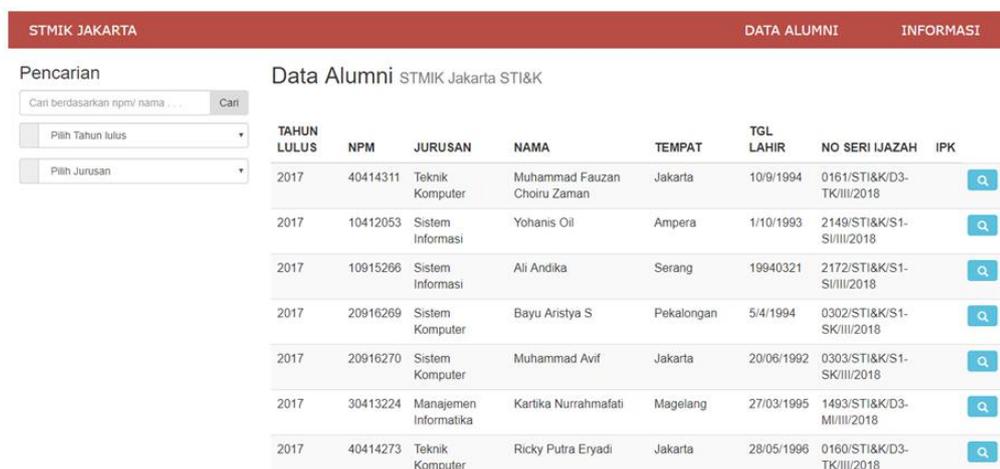
Halaman login merupakan halaman admin untuk akses masuk dimana dapat mengelola data alumni. Pada tampilan masuk jika username dan password diisi dengan benar, maka website akan menampilkan halaman sesuai dengan hak aksesnya. Dan jika sebaliknya, username dan password diisi salah maka tidak dapat masuk ke dalam halaman hak akses tersebut. Tampilan halaman login dapat dilihat pada gambar 11.



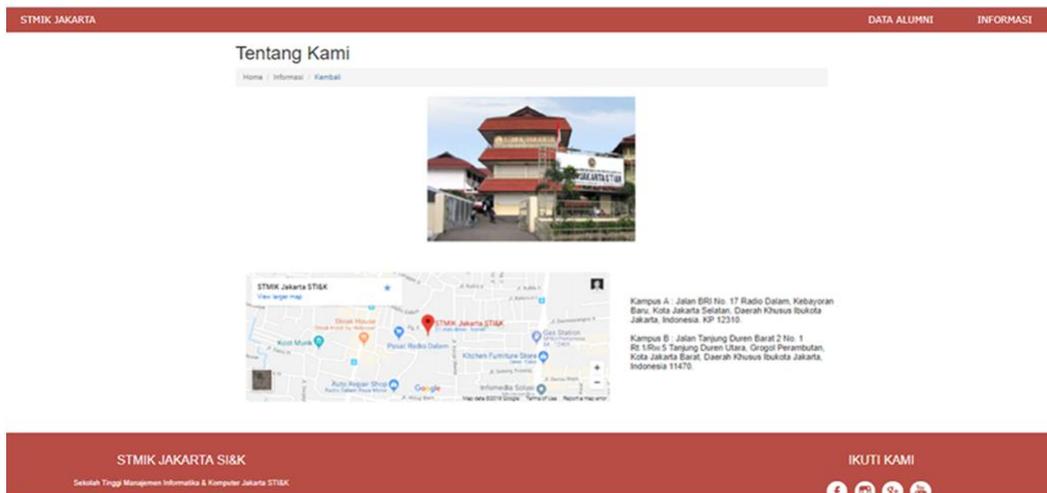
Gambar 7: Tampilan Halaman Program Studi



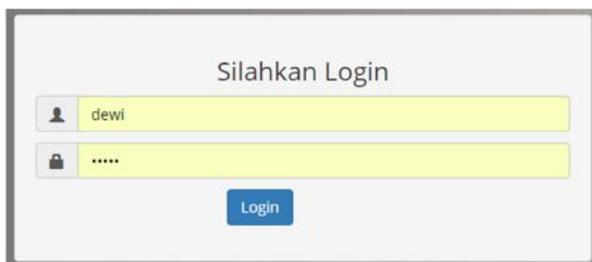
Gambar 8: Tampilan Halaman Data Alumni



Gambar 9: Tampilan Halaman Pencarian



Gambar 10: Tampilan halaman Informasi



Gambar 11: Tampilan Halaman login

Pada gambar 12 merupakan halaman utama admin yang langsung dapat melihat data alumni dan dapat menggunakan hak aksesnya, gambar 13 merupakan halaman untuk menambah data alumni sedangkan gambar 14 merupakan halaman edit data alumni yang dapat merubah data-data yang pernah diinput dan memperbaharui data.

NO	TAHUN LULUS	NPM	JURUSAN	NAMA	TEMPAT	TGL_LAHIR	NO SERI IJAZAH	IPK	Opsi
1	2010	10909005	Sistem Informasi	Ricky Rianto	Tangerang	12/9/1983	1340/STI&K/S1-SI/III/2011	319	[Edit] [Delete]
2	2010	10909219	Sistem Informasi	Debby Hairunisa Tarunadjaya	Palembang	11/12/1986	1341/STI&K/S1-SI/III/2011	310	[Edit] [Delete]
3	2010	10909235	Sistem Informasi	Ratih Srimaya	Jakarta	3/1/1974	1342/STI&K/S1-SI/III/2011	320	[Edit] [Delete]
4	2010	30406120	Manajemen Informatika	Arief Purwo Santosa	Jakarta	2/4/1988	1195/STI&K/D3-MI/III/2011	269	[Edit] [Delete]
5	2010	30407034	Manajemen Informatika	Andhika Wisnu Aji	Jakarta	20/06/1988	1196/STI&K/D3-MI/III/2011	271	[Edit] [Delete]
6	2010	30407188	Manajemen Informatika	Dwi Purnama Sari	Jakarta	19890616	1197/STI&K/D3-MI/III/2011	375	[Edit] [Delete]
7	2010	40407102	Teknik Komputer	Mayreza Saputra	JAKARTA	8/5/1989	0109/STI&K/D3-TK/III/2011	359	[Edit] [Delete]

Gambar 12: Halaman Utama Admin

Gambar 13: Tambah Data Alumni

Gambar 14: Halaman Edit Data Alumni

Tampilan Halaman Pencarian

Pada penelitian ini telah diuji coba dengan mengambil beberapa sampel dari 6503 data alumni yang ada mulai dari tahun 2010 sampai 2017. Dengan perhitungan sampel dengan menggunakan rumus slovin sebagai berikut :

$$n = N / (N(e)^2 + 1)$$

dengan keterangan : n = sampel N = populasi e = batas toleransi kesalahan

Diketahui data populasi adalah 6503, dan tingkat kesalahan yang di kehendaki adalah 9% maka jumlah sampel yang akan digunakan adalah $n = 6503 / (0.09)^2 + 1 = 124,46$ dibulatkan menjadi 124. Dengan demikian,

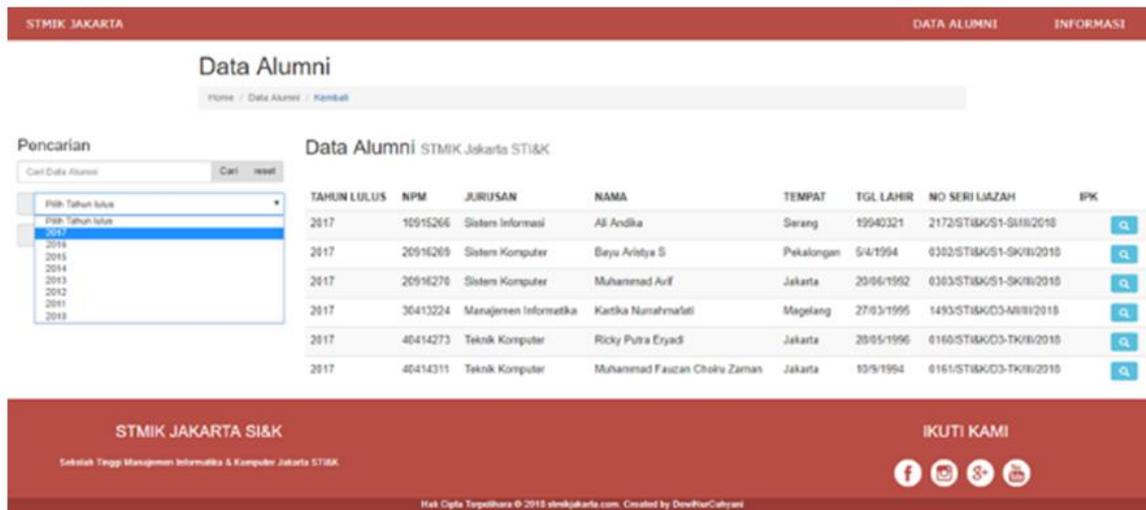
jumlah sampel yang dibutuhkan adalah 124 data dengan nilai akurasi 91%

Hasil Pencarian Berdasarkan Tahun Lulus

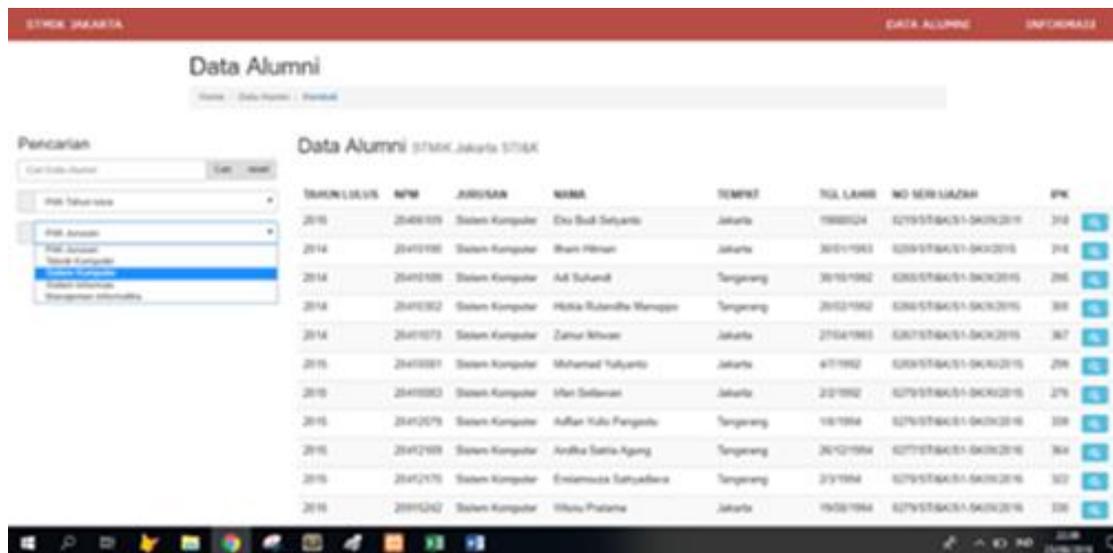
Pencarian data alumni dapat dilakukan dengan beberapa pencarian. Pada gambar 15 merupakan hasil pencarian berdasarkan tahun lulus sesuai dengan pengelompokannya.

Hasil Pencarian Berdasarkan Jurusan

Pada gambar 16 merupakan hasil pencarian berdasarkan Jurusan sesuai dengan pengelompokannya.



Gambar 15: Hasil Pencarian Berdasarkan Tahun Lulus

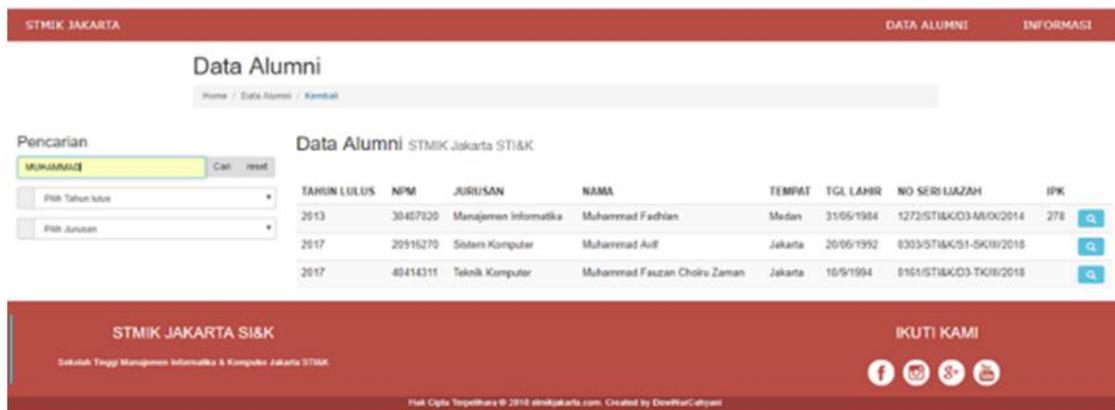


Gambar 16: Hasil Pencarian Berdasarkan Jurusan

Hasil Pencarian Berdasarkan Nama / Penutup NPM

Pada gambar 17 merupakan hasil pencarian berdasarkan data sesuai data yang ada pada database. Kemudian apabila data yang dicari merupakan data mahasiswa yang belum lulus/ mahasiswa aktif maka ada keterangan 'Belum Lulus'. Serta apabila data yang dicari bukan merupakan alumni / mahasiswa aktif STMIK Jakarta STI&K, maka terdapat pemberitahuan bahwa data yang dicari tidak terdaftar sebagai mahasiswa STMIK Jakarta STI&K.

Dari penelitian ini terlihat kelebihan dari metode Suffix Tree Clustering dengan penggunaan metode pengelompokan data lebih efisien berdasarkan karakteristik yang sama ke suatu wilayah yang sama. Penggunaan frasa (muti word term) prosesnya lebih sederhana dibandingkan dengan penggunaan algoritma lain. Dan aplikasi ini dapat memberikan kemudahan dalam proses pencarian data alumni secara rinci tanpa memerlukan login terlebih dahulu. Penggunaan sampel sebanyak 124 dari 6503 data yang berada pada database alumni pada penelitian ini memiliki nilai akurasi dalam ketepatan pencarian sebesar 91%.



Gambar 17: Hasil Pencarian Berdasarkan Nama / NPM

Hasil pengujian terhadap pengguna alumni dengan jumlah 20 responden yang memberi tanggapan terhadap aplikasi memiliki nilai presentase sebesar 62% dan termasuk dalam kategori “Baik”, sehingga penggunaan aplikasi ini dapat disimpulkan dapat memberikan kemudahan dan ketepatan kepada pengguna dalam proses pencarian data alumni secara detail tanpa harus login terlebih dahulu. Hasil Pengujian terhadap responden 20 alumni sebagai pengguna dengan tanggapan terhadap aplikasi yang telah dibuat masuk kedalam kategori “Baik”.

Daftar Pustaka

- [1] 1. Jogyanto Hartono, “Analisis dan Desain Sistem Informasi Pendekatan Terstruktur

Teori dan Praktik Aplikasi Bisnis”, Andi, Yogyakarta, 2008.

- [2] Novan S, ”Implementasi Aplikasi Information Retrieval Untuk Pendeteksian dan Klasifikasi Berita Kejadian Berbahasa Indonesia Berbasis Web”, Tugas akhir, Jurusan Teknik informatika Fakultas Teknologi Informasi ITS Surabaya, 2001.
- [3] Kusumaya, “Pengembangan Suffix Tree Clustering untuk Comparative Text Mining”, Tesis. Bandung : Institute Teknologi Bandung, 2007.
- [4] Haryanto, “Identifikasi Naskah Dokumen Teks dengan Metode Indexing”, Jurnal Telematika, UPN Veteran Yogyakarta, 2012.