SISTEM PENGUKURAN TINGKAT SIMILARITAS DOKUMEN

SKRIPSI

DENNIS AHMAD TAUFIK

10109725



PROGRAM STUDI S1
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS KOMPUTER INDONESIA
2012

LATAR BELAKANG MASALAH

1. Teknik plagiarisme paling banyak dilakukan dengan *copy paste*, cukup dengan *copy paste* semua tugas sekolah atau kuliah dapat terselesaikan dengan mudah.
2. Tindakan Plagiarisme ini dapat berdampak pada penurunan kreatifitas siswa maupun mahasiswa.
3. Dengan adanya kemudahan dalam memperoleh informasi, mahasiswa menjadi cenderung malas dalam mengembangkan kreativitas dan pemikirannya sendiri dalam mengerjakan tugas, serta cenderung untuk meniru hasil karya orang lain, bahkan menggunakan hasil karya orang lain secara keseluruhan.

IDENTIFIKASI MASALAH

1. Bagaimana merancang dan membangun suatu aplikasi pengukuran tingkat similaritas dokumen dengan tingkat akurasi yang baik?
2. Bagaimana caranya agar aplikasi dapat dan mampu mendeteksi suatu aksi plagiarisme pada suatu dokumen, meskipun memiliki perbedaan tata letak (posisi) tetapi memiliki kesamaan kata?
3. Bagaimana persentase similaritas yang dihasilkan oleh sistem terhadap dokumen yang dibandingkan?

MAKSUD DAN TUJUAN

Maksud dari penulisan tugas akhir ini adalah untuk membangun sistem pengukuran tingkat similaritas dokumen.

Tujuannya adalah :

1. Untuk merancang dan membangun suatu aplikasi pengukuran tingkat similaritas dokumen dengan tingkat akurasi yang baik
2. Aplikasi dapat mendeteksi suatu aksi plagiarisme pada suatu dokumen, meskipun memiliki perbedaan tata letak (posisi) tetapi memiliki kesamaan kata
3. Mengetahui persentase similaritas yang dihasilkan oleh sistem terhadap dokumen yang dibandingkan

BATASAN MASALAH

1. *File* yang dapat menjadi input pada aplikasi ini hanya *file* yang berekstensi .doc, .docx, .txt, dan .pdf.
2. *Output* yang akan ditampilkan pada aplikasi ini sebagai hasil akhir aplikasi, antara lain ialah:
    1. Persentase derajat kemiripan.
    2. Munculnya *tag* info "Terdeteksi Plagiat" apabila derajat tingkat kemiripannya melebihi batas toleransi, sebaliknya akan muncul *tag* info "Plagiat Tidak Terdeteksi" apabila derajat tingkat kemiripan masih dalam batas toleransi.
3. Aplikasi yang akan dibangun dapat bekerja dengan sistem *one to one document process :* satu dokumen acuan dan satu dokumen lainnya sebagai objek yang akan dibandingkan.

BATASAN MASALAH

4. Pada saat proses pengukuran, aplikasi hanya mendeteksi string teks (mencakup huruf dan angka), dan tidak memperhitungkan *equation* (rumus), tabel, dan gambar.
5. Dokumen yang dibandingkan tidak hanya dokumen yang ditulis dalam bahasa Indonesia, bisa digunakan untuk semua bahasa.
6. Aplikasi ini tidak bisa melakukan pengukuran tingkat similaritas dengan dokumen yang di proteksi.

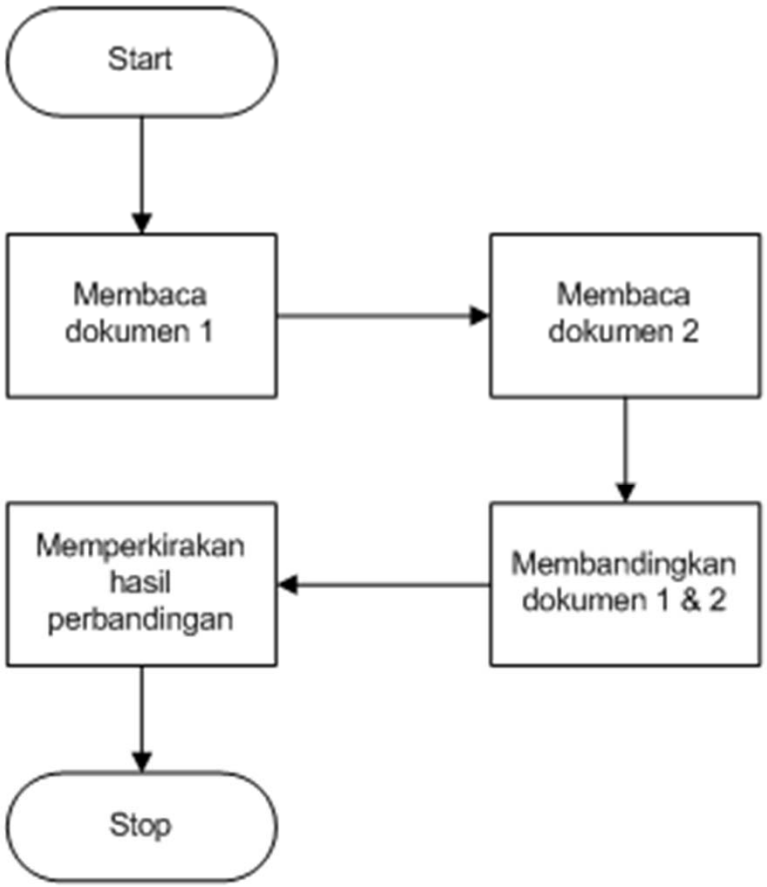
BATASAN MASALAH

7. Aplikasi ini dipersiapkan untuk berjalan hanya di satu komputer (belum sampai mengadopsi konsep *client – server*).
8. Penerapan uji coba aplikasi pengukuran tingkat similaritas dokumen ini akan diuji pada Tugas Akhir IF-UNIKOM.
9. Aplikasi ini dapat diterapkan di sekolah atau di kampus untuk mendeteksi similaritas antar dokumen.
10. Metode yang digunakan yaitu metode *fingerprinting*.

ANALISIS KEGIATAN

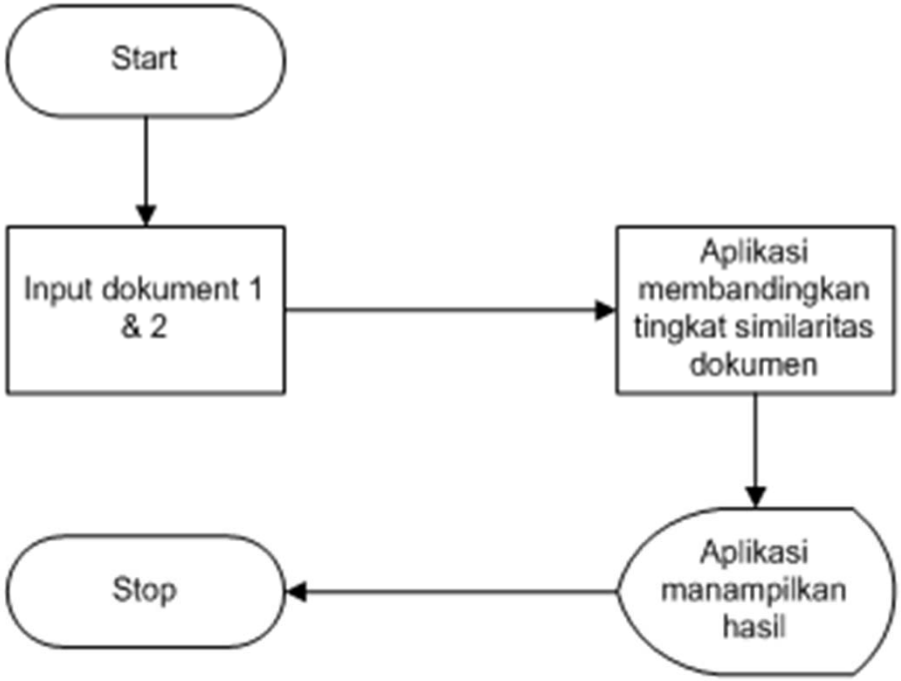
1. STUDI LITERATUR
2. ANALISIS DAN PERANCANGAN SISTEM
3. IMPLEMENTASI SISTEM
4. UJI COBA SISTEM
5. EVALUASI SISTEM DAN ANALISA HASIL UJI COBA

Sistem manual



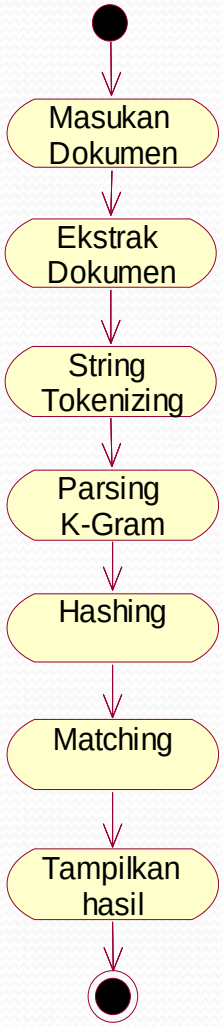
Gambar 3.2 Sistem Manual

APLIKASI SIMILARITAS DOKUMEN



Gambar 3.3 Aplikasi Similaritas Dokumen

ANALISIS ALGORITMA RABIN-KARP



ANALISIS ALGORITMA

LANGKAH PENGGUNAAN ALGORITMA RABIN-KARP :

1. CASE FOLDING
2. TOKENIZING
3. K-GRAM
4. HASHING
5. MENGHITUNG NILAI SIMILARITAS

MENGHITUNG HASHING

$$H = C_1 * a^{(k-1)} + C_2 * a^{(k-2)} + C_3 * a^{(k-3)} \ldots + C_k * a^0 \quad \ldots(1)$$

Gambar 2.4 Rumus formula rolling hash

Keterangan :

H adalah nilai *Hash*

C adalah nilai ASCII suatu karakter

a adalah basis (tidak boleh 1 dan 0)

k adalah banyaknya karakter

MENGHITUNG NILAI SIMILARITAS

$$S = \frac{N_t}{N_x + N_y}$$

Keterangan:

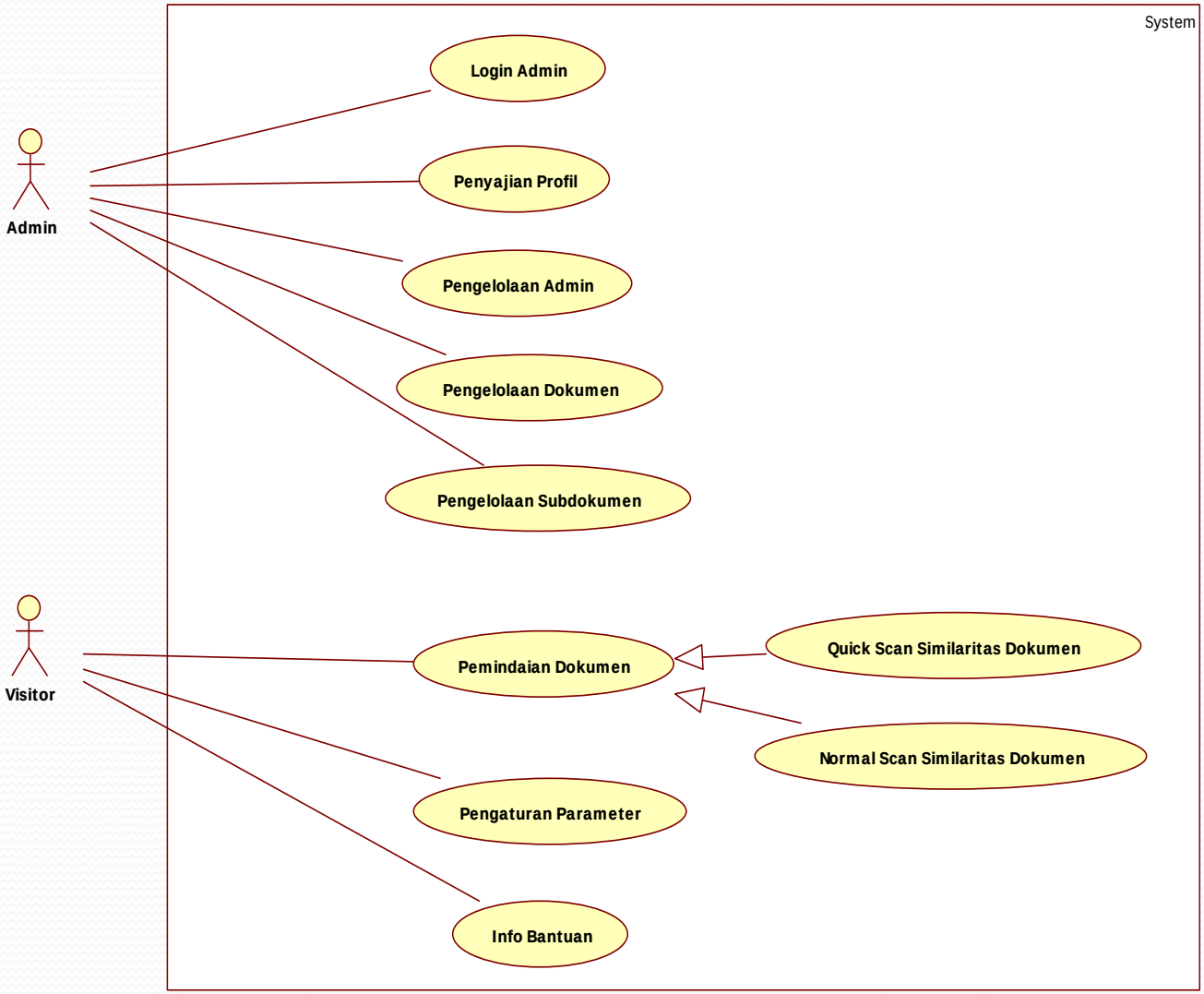
S = Similaritas

N_t = Jumlah Hashing yang sama

N_x = Total substring asli

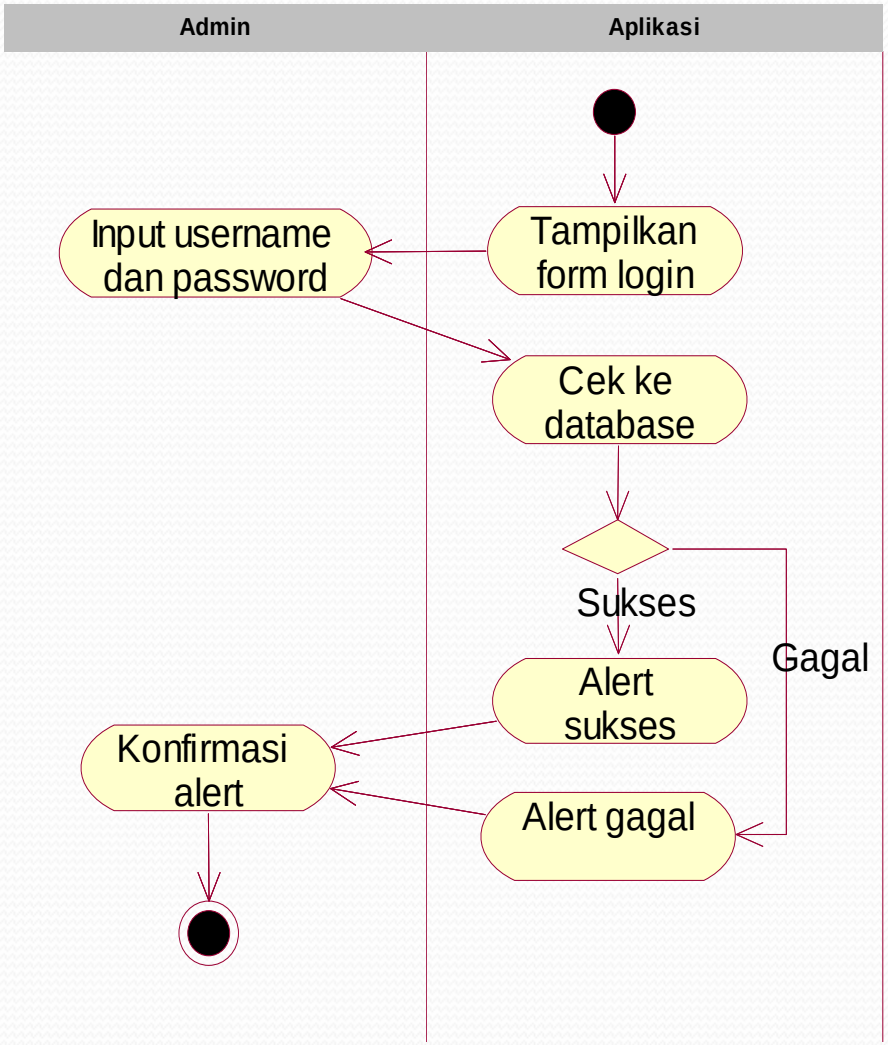
N_y = Total substring uji

USE CASE DIAGRAM

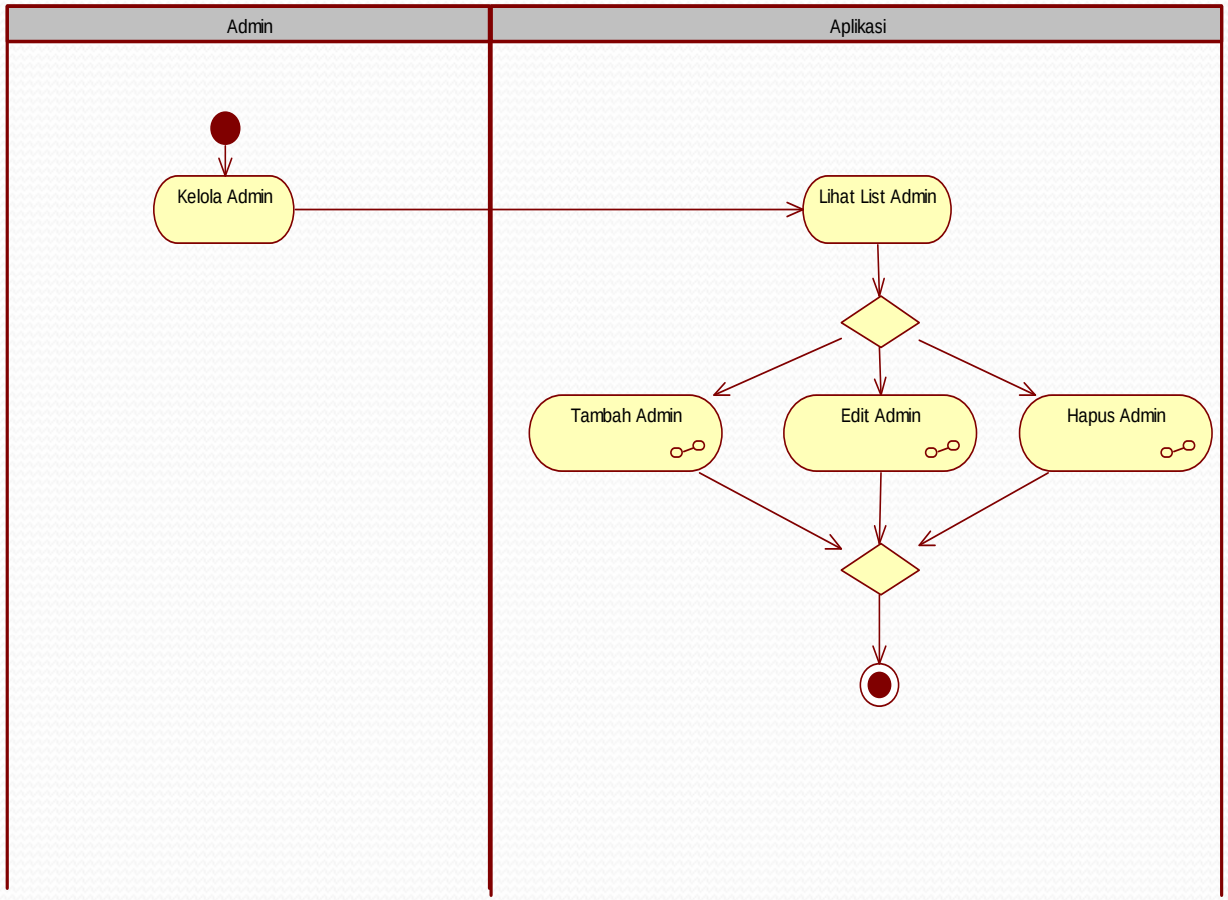


ACTIVITY DIAGRAM ADMIN

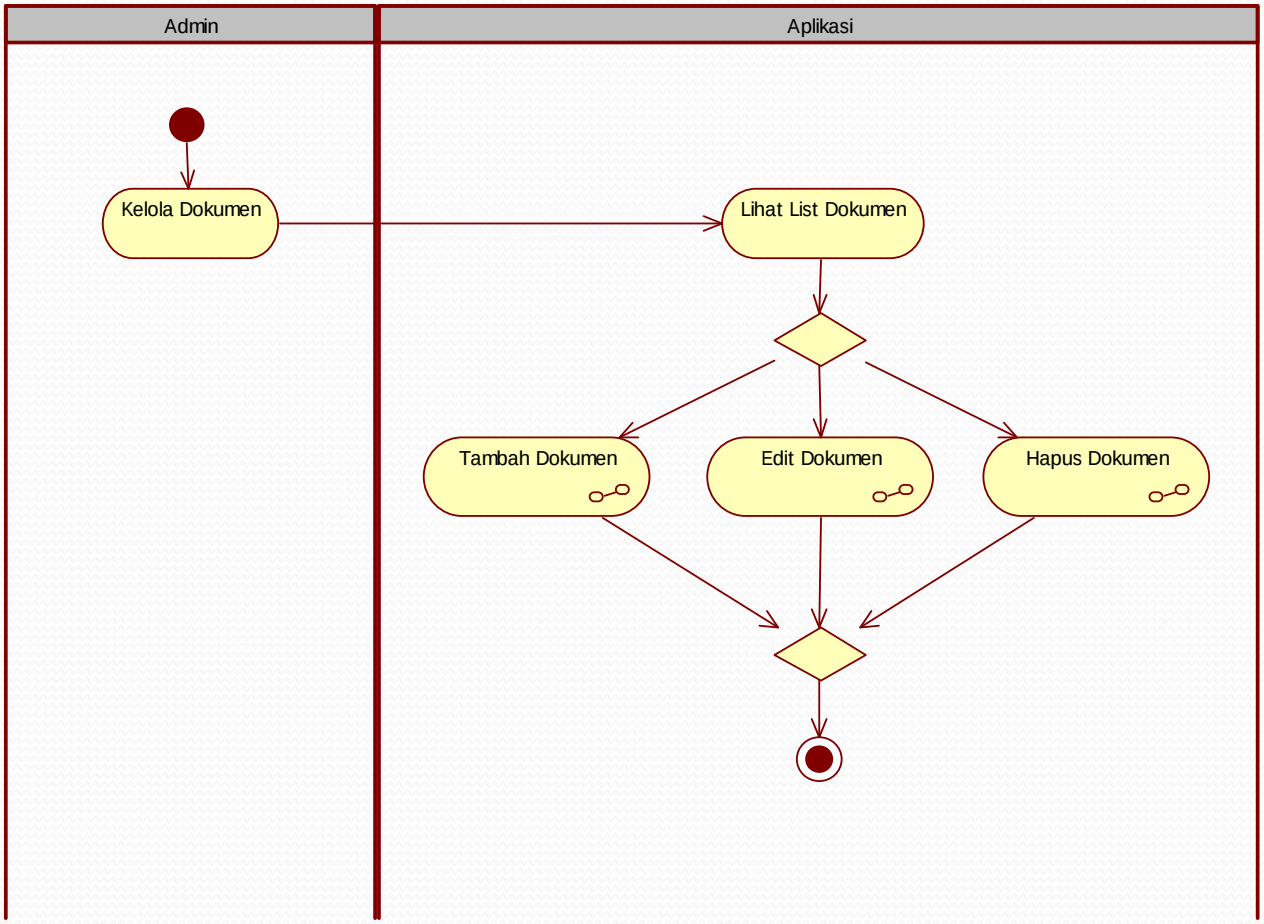
LOGIN



ACTIVITY DIAGRAM KELOLA ADMIN

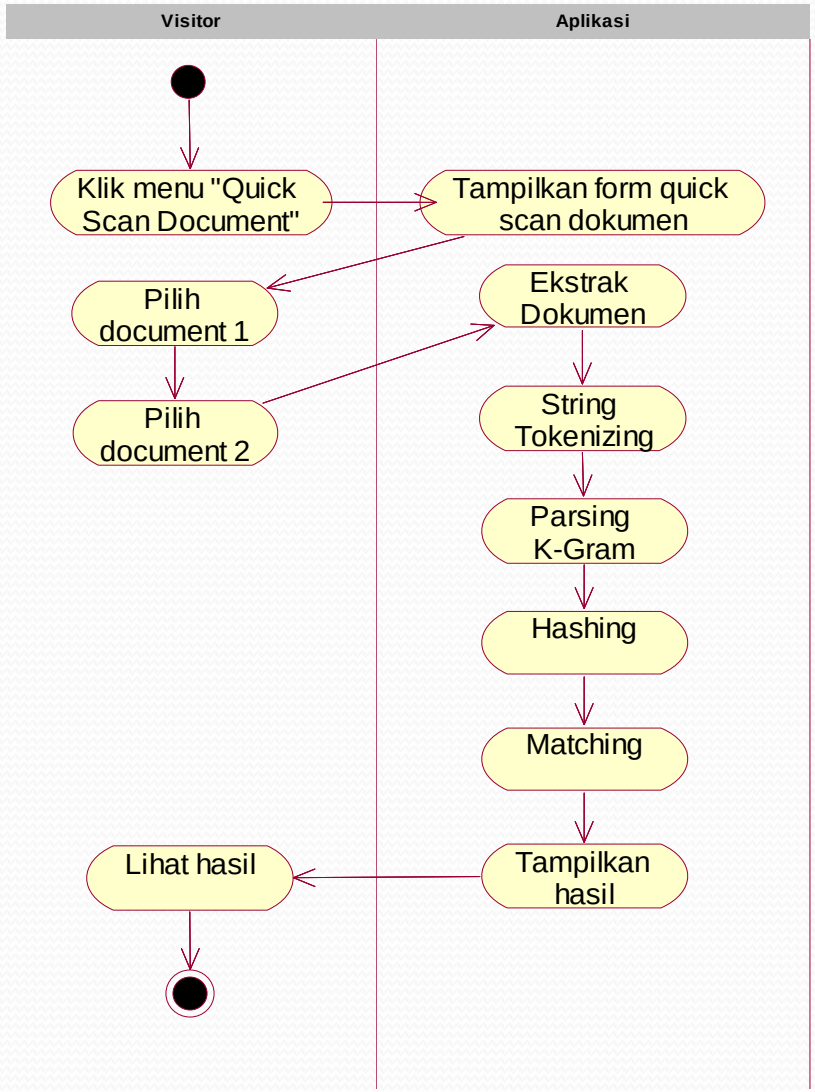


ACTIVITY DIAGRAM KELOLA DOKUMEN

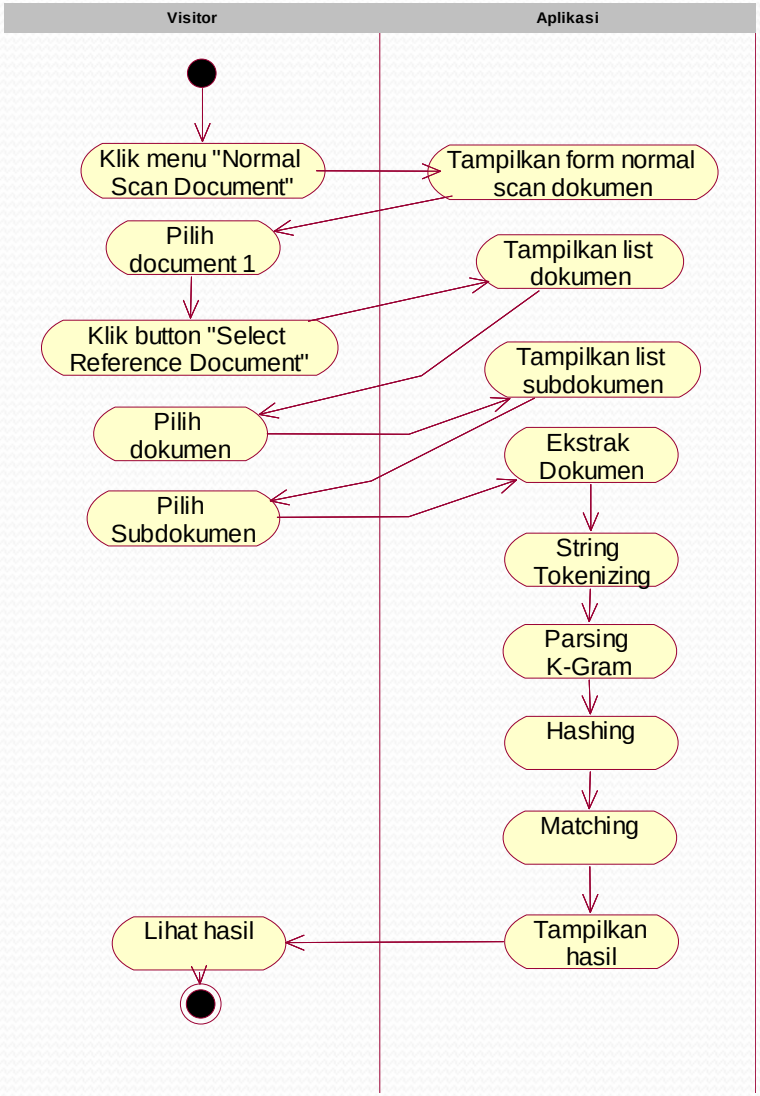


ACTIVITY DIAGRAM VISITOR

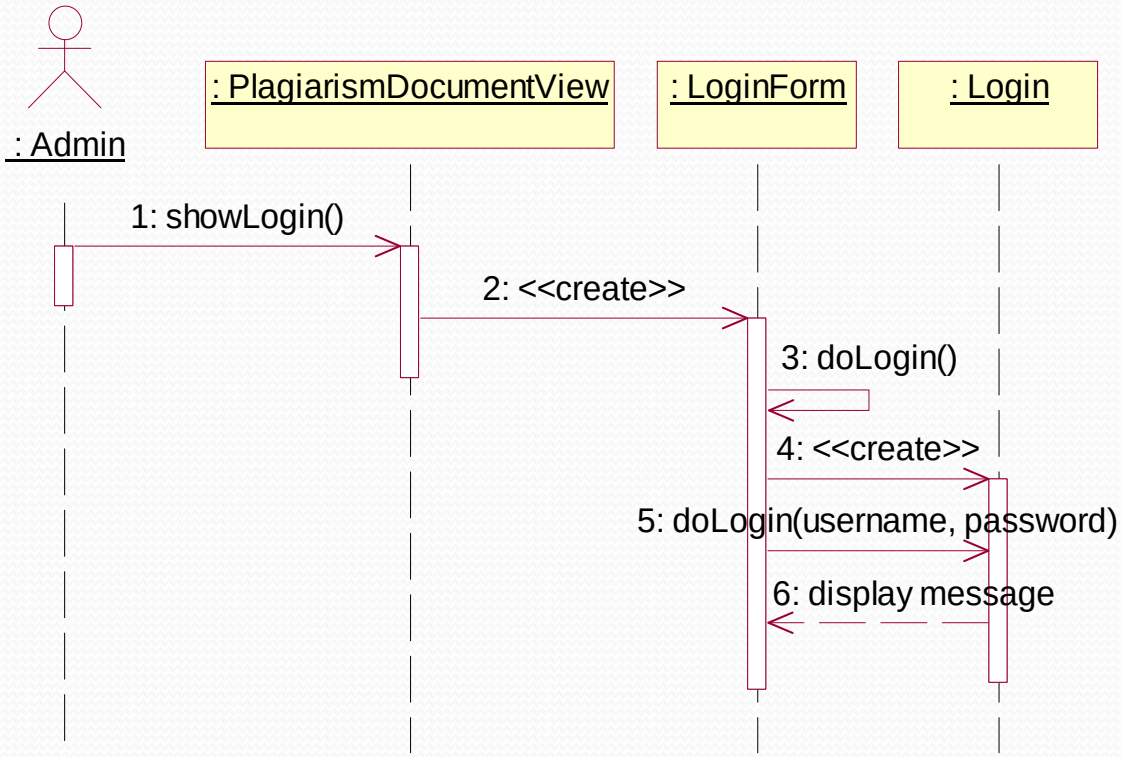
QUICK SCAN DOKUMEN



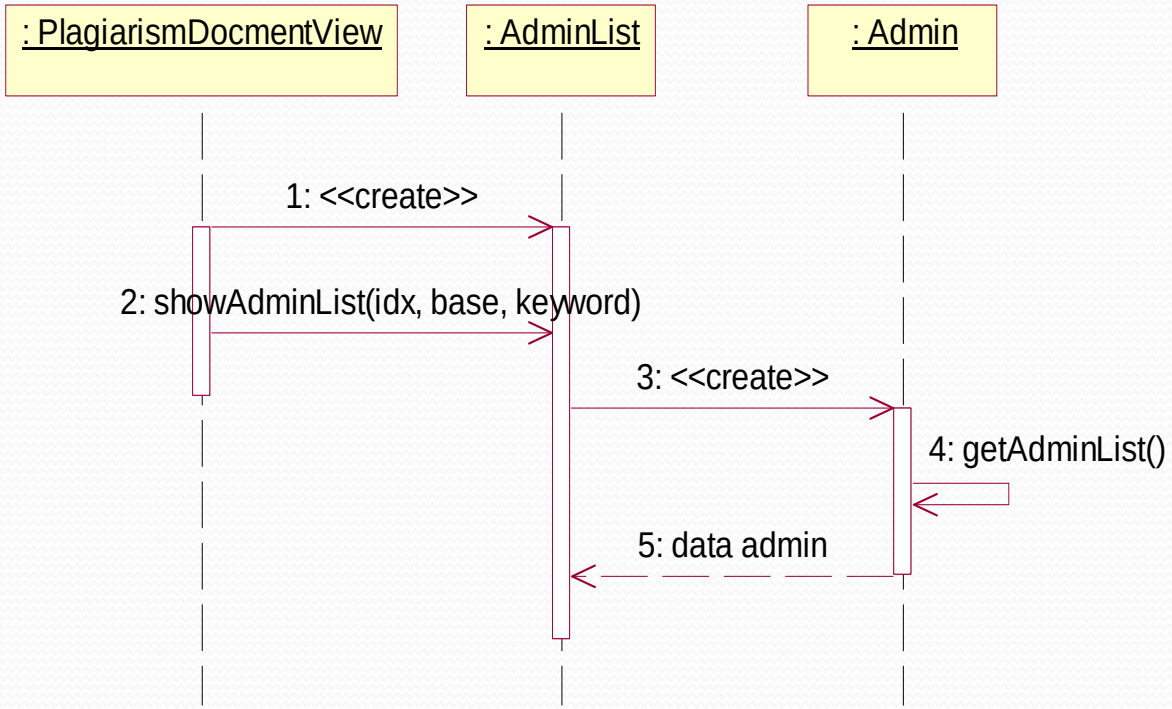
ACTIVITY DIAGRAM VISITOR
NORMAL SCAN DOKUMEN



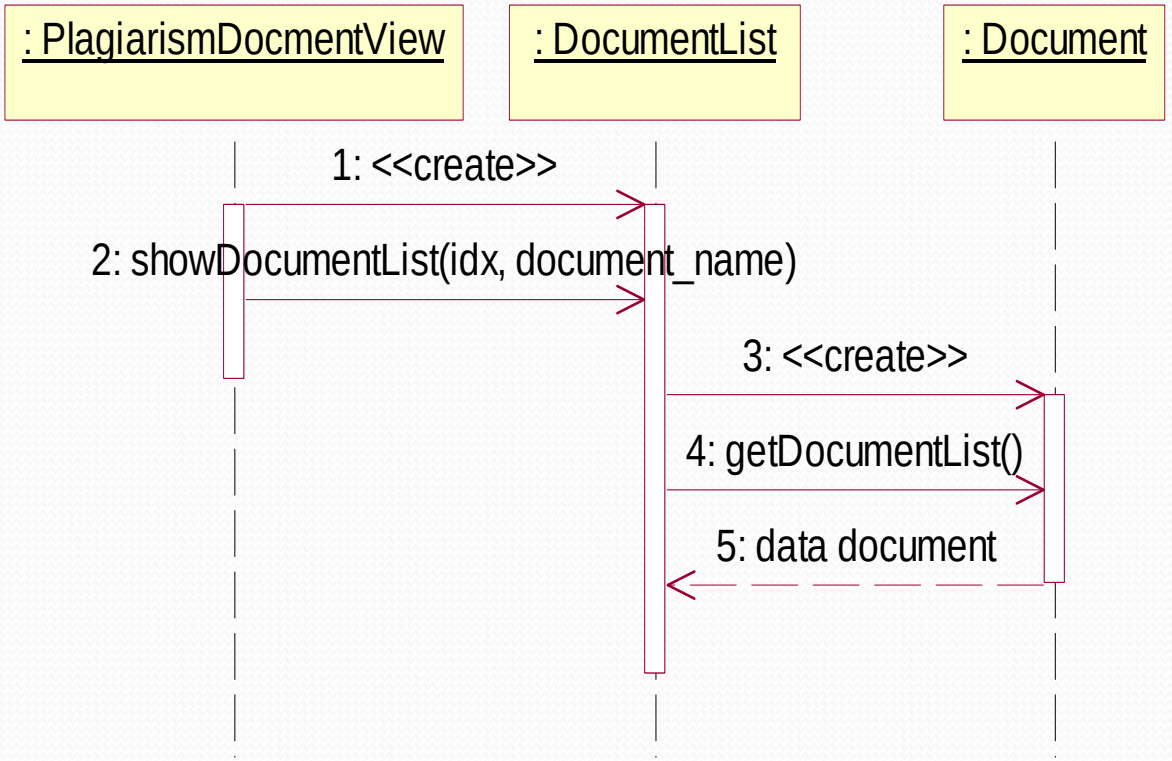
SEQUENCE DIAGRAM ADMIN LOGIN



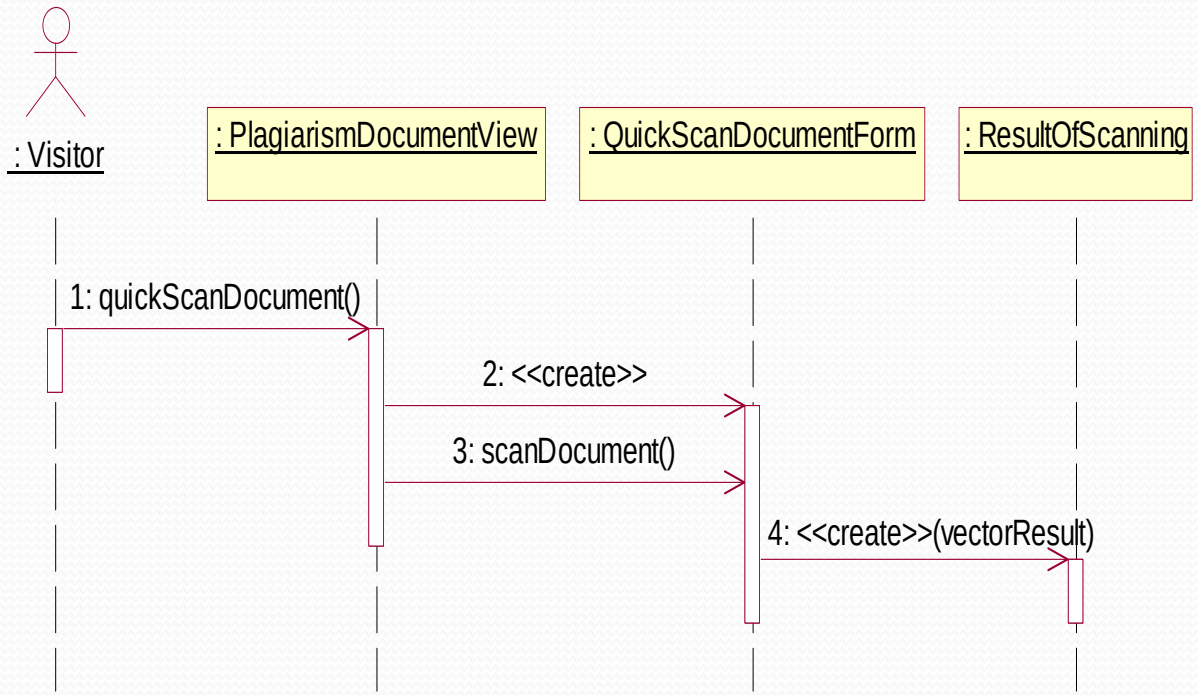
SEQUENCE DIAGRAM KELOLA ADMIN



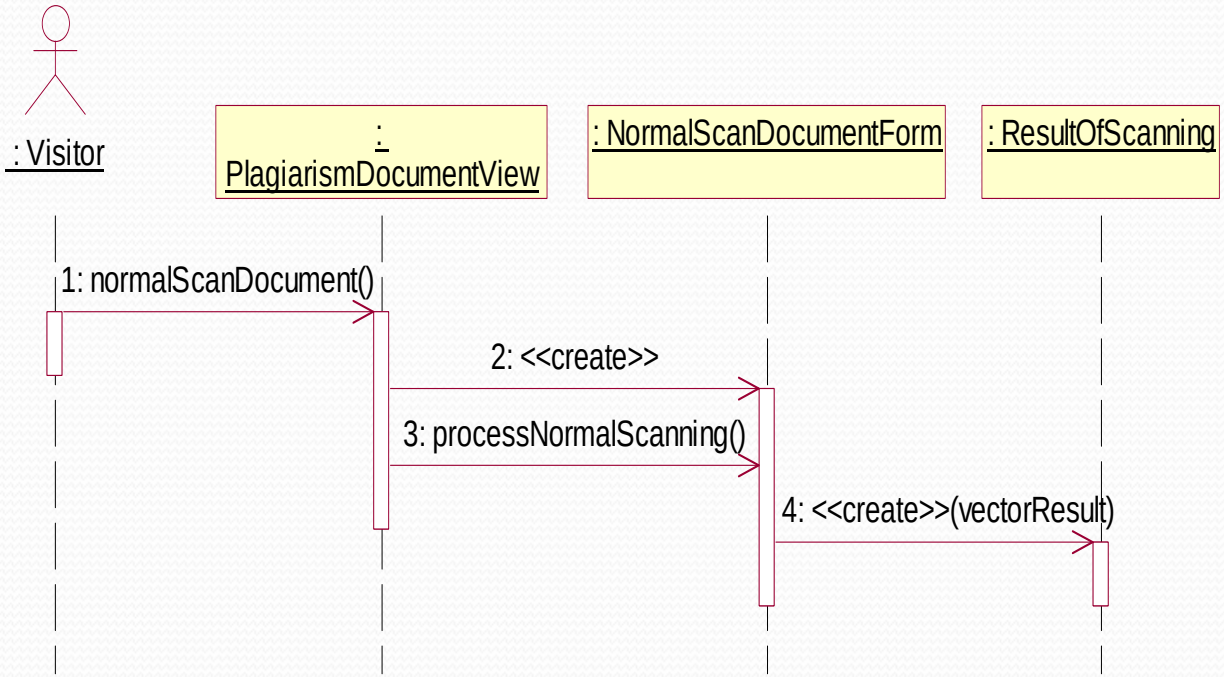
SEQUENCE DIAGRAM KELOLA DOKUMEN



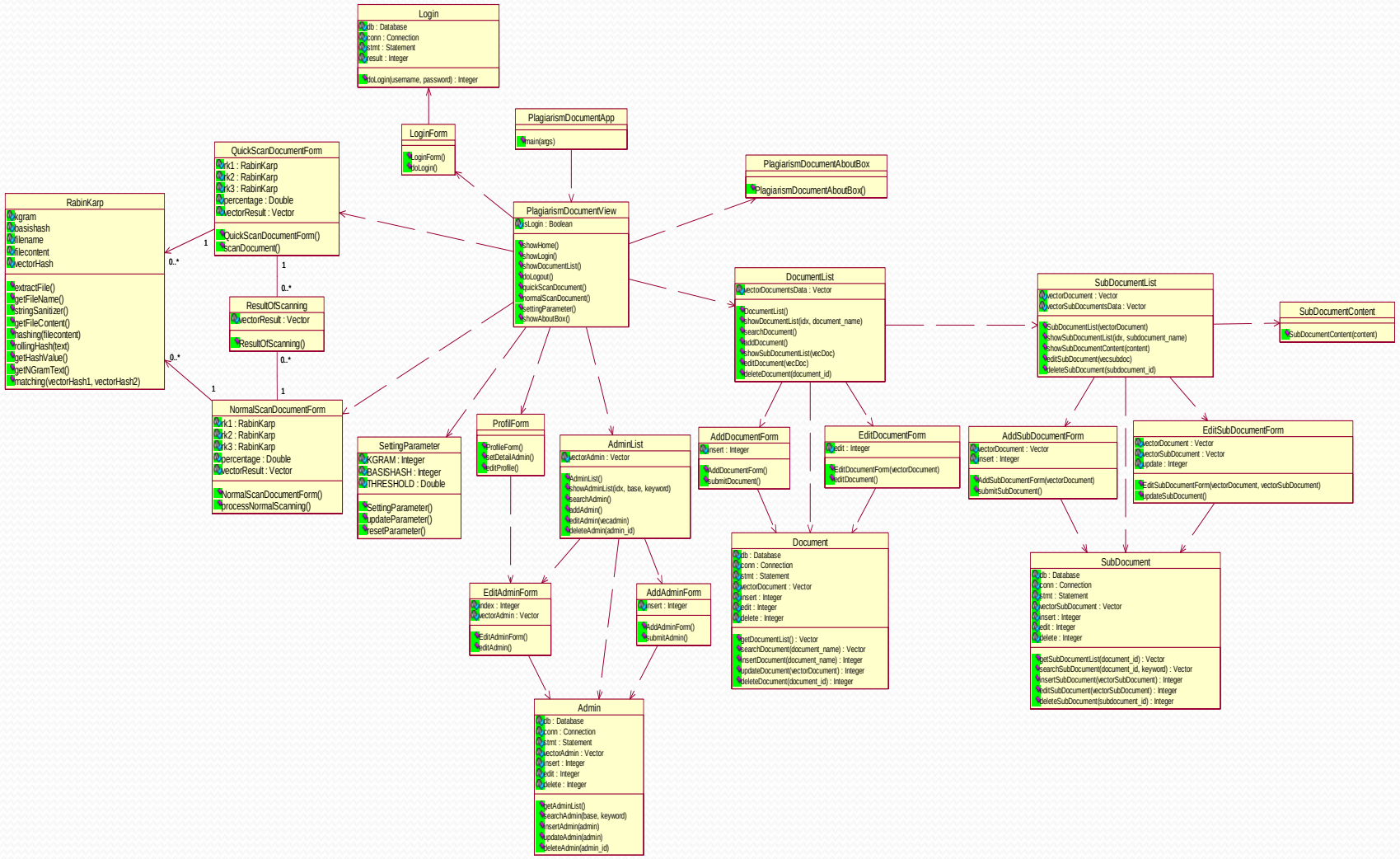
SEQUENCE DIAGRAM VISITOR QUICK SCAN DOKUMEN



SEQUENCE DIAGRAM VISITOR NORMAL SCAN DOKUMEN



CLASS DIAGRAM

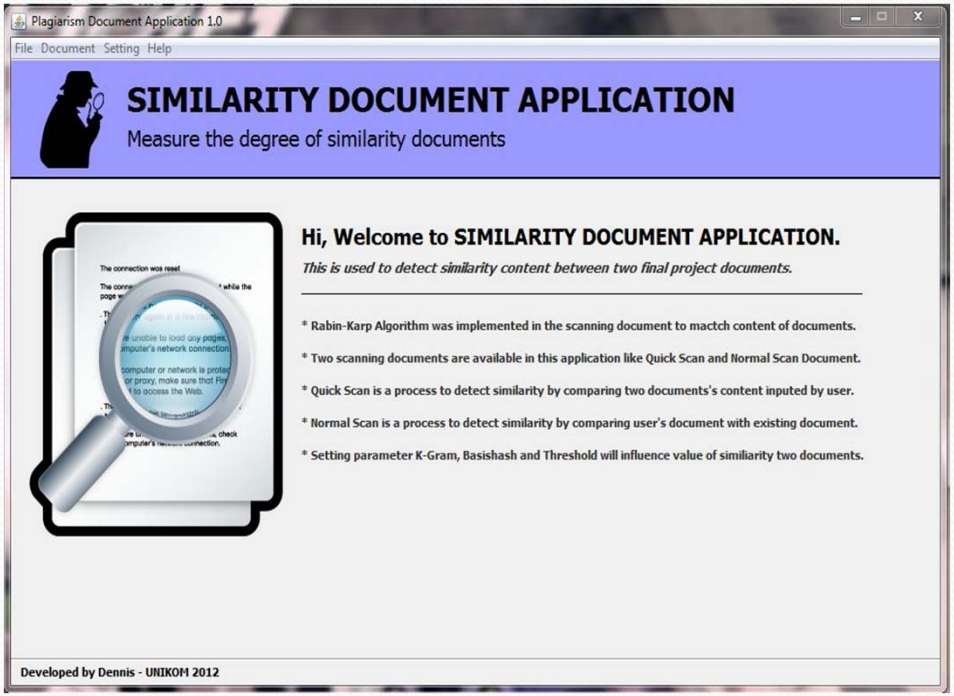


IMPLEMENTASI ANTARMUKA

LOGO

HEADER

ABOUT PLAGIARISM DOCUMENT APPLICATION



DEMO PROGRAM

KESIMPULAN DAN SARAN

KESIMPULAN:

1. Aplikasi Pengukuran Tingkat Similaritas Dokumen sudah cukup mudah digunakan.
2. Performansi dari aplikasi sudah cukup cepat untuk memproses atau menjalankan fungsinya.
3. Aplikasi ini belum mampu optimal membantu memudahkan proses mendeteksi tindakan plagiarisme pada dokumen di Perguruan Tinggi.
4. Dalam beberapa kondisi tertentu algoritma Rabin-Karp belum dapat secara tepat mengukur tingkat similaritas dokumen sehingga diperlukan algoritma lain yang lebih baik lagi.

KESIMPULAN DAN SARAN

SARAN:

1. Aplikasi ini diharapkan dapat diterapkan di sekolah atau di kampus dalam memeriksa similaritas dokumen.
2. Aplikasi ini akan lebih bagus lagi jika diterapkan tidak hanya menggunakan satu algoritma.
3. Aplikasi ini akan lebih bagus lagi jika bisa menentukan persamaan makna *sinonim* kata.

TERIMA KASIH

谢谢