

**PENGUJIAN TINGKAT KEMIRIPAN PADA DOKUMEN DENGAN
MENGUNAKAN TABEL FREKUENSI DAN BARCHART**

Tugas Akhir



Oleh :

Juan Robert Sirait

22074361

Program Studi Teknik Informatika
Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana Yogyakarta
Tahun 2012

**PENGUJIAN TINGKAT KEMIRIPAN PADA DOKUMEN DENGAN
MENGUNAKAN TABEL FREKUENSI DAN BARCHART**

Tugas Akhir



Diajukan kepada Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai salah satu syarat dalam memperoleh gelar
Sarjana Komputer



Disusun oleh :
Juan Robert Sirait
22074361

Program Studi Teknik Informatika
Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana Yogyakarta
Tahun 2012

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

Pengujian Tingkat Kemiripan pada Dokumen dengan Menggunakan Tabel Frekuensi dan Barchart.

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 14 Agustus 2012



JUAN ROBERT SIRAIT
22074361

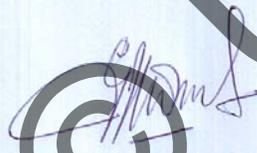


HALAMAN PERSETUJUAN

Judul : Pengujian Tingkat kemiripan Pada Dokumen Dengan
Menggunakan Tabel Frekuensi Dan Barchart
Nama : Juan Robert Sirait
NIM : 22074361
Mata Kuliah : Tugas Akhir
Kode : TIW276
Semester : Genap
Tahun akademik : 2011/2012

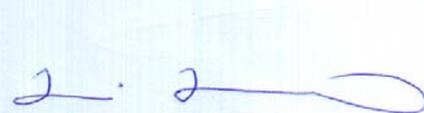
Telah diperiksa dan disetujui
Di Yogyakarta,
Pada tanggal ...14...Agustus 2012

Dosen Pembimbing I



(Drs. R. Gunawan Santosa, M.Si)

Dosen Pembimbing II



(Ir. Sri Suwarno, M.Eng)

HALAMAN PENGESAHAN

PENGUJIAN TINGKAT KEMIRIPAN PADA DOKUMEN DENGAN MENGUNAKAN TABEL FREKUENSI DAN BARCHART.

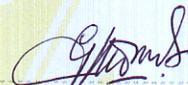
Oleh: JUAN ROBERT SIRAIT / 22074361

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 31 Juli 2012

Yogyakarta, 14 Agustus 2012
Mengesahkan,

Dewan Penguji:

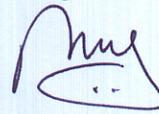
1. Drs. R. Gunawan Santosa, M.Si.
2. Ir. Sri Suwarno, M.Eng.
3. Willy Sudiarto Raharjo, SKom.,M.Cs
4. Aloysius Airlangga Bajuadji, S.Kom., M.Eng.



Dekan

(Drs. Wimmie Handiwidjojo, MIT.)

Ketua Program Studi



(Nugroho Agus Haryono, M.Si)

UCAPAN TERIMA KASIH

Puji dan syukur penulis haturkan kehadiran Tuhan Yang Maha Esa atas limpahan rahmat, kasih, berkat dan anugerah Nya yang tak berkesudahan, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul Pengujian Tingkat Kemiripan pada Dokumen dengan Menggunakan Tabel Frekuensi dan *Barchart*.

Tugas Akhir ini merupakan kelengkapan dan pemenuhan dari salah satu syarat dalam memperoleh gelar Sarjana Komputer. Selain itu bertujuan melatih mahasiswa untuk dapat menghasilkan suatu karya yang dapat dipertanggung jawabkan secara ilmiah, sehingga dapat bermanfaat bagi penggunaannya.

Dalam menyelesaikan pembuatan program dan laporan Tugas Akhir ini, penulis telah banyak menerima bimbingan, saran dan masukan dari berbagai pihak, baik secara langsung maupun secara tidak langsung. Untuk itu dari lubuk hati yang paling dalam pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. **Bpk Drs. R. Gunawan Santosa, M.Si** selaku dosen pembimbing I atas bimbingan, petunjuk, masukan dan semangat yang diberikan kepada penulis.
2. **Bpk Ir. Sri Suwarno, M.Eng** selaku dosen pembimbing II yang telah memberikan bimbingannya dengan sabar, petunjuk, ide serta sangat memotivasi penulis selama pengerjaan Tugas Akhir ini.
3. **Bapak (alm)** dan **Mama** atas senyuman, kesabaran, dan cintanya dalam mendukung penulis untuk menyelesaikan Tugas Akhir ini, maaf kalau selama ini aku merepotkan kalian.
4. **K'letta, K'Rani, Anti** dan **Ucok**, terimakasih untuk doa dan dukungannya yang membuat penulis mampu menyelesaikan Tugas Akhir ini.
5. Sahabat-sahabat karib penulis **Umbu, Yudi, Arris, Junior, Adin, Ivan, Budi**, dan semua teman-teman seperjuangan yang telah mendukung penulis dalam menyelesaikan Tugas Akhir yang tidak bisa disebutkan satu-persatu.

6. Pihak-pihak lain yang tidak dapat penulis sebutkan satu per satu, sehingga Tugas Akhir ini dapat terselesaikan dengan baik.

Penulis menyadari bahwa program dan laporan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari pembaca sekalian sehingga suatu saat penulis dapat memberikan karya yang lebih baik lagi.

Akhir kata penulis ingin meminta maaf bila ada kesalahan baik dalam penyusunan laporan maupun yang pernah penulis lakukan sewaktu membuat program Tugas Akhir. Sekali lagi penulis mohon maaf yang sebesar-besarnya dan semoga ini bermanfaat bagi kita semua.

Yogyakarta, 23 Juli 2012

Penulis



INTISARI

Plagiasi adalah pemakaian dan penggunaan ide atau hasil karya orang lain yang diklaim sebagai hasil karyanya sendiri. Pemakaian istilah plagiasi juga biasanya dipakai untuk penggunaan karya orang lain tanpa menyebutkan sumbernya atau identitas si pembuat karya tersebut. Oleh karena itu diperlukan suatu sistem yang mampu membantu dalam mengkalkulasi frekuensi kemunculan kata-kata yang ada pada suatu *document* untuk menganalisis tingkat kemiripannya. Dalam penelitian ini akan dianalisis kemiripan *document* untuk diketahui tingkat kemiripan yang terdapat diantara *document*, untuk itu digunakan Tabel Frekuensi dan *Barchart*.

Penggunaan Tabel Frekuensi dan *Barchart* diterapkan pada proses pencarian tingkat kemiripan *document* yaitu rata-rata kedekatan dua *document* yang didapat dari proses perbandingan. Selain penggunaan Tabel Frekuensi dan *Barchart* yang diterapkan juga mampu menampilkan informasi dari data yang dipakai berupa frekuensi kata-kata dan juga grafik batang.

Penerapan pengujian tingkat kemiripan pada dokumen dengan menggunakan Tabel Frekuensi dan *Barchart* dapat berjalan dengan baik. Pengujian yang telah dilakukan menunjukkan bahwa pengujian tersebut mampu menemukan tingkat kemiripan *document* dan membandingkan tingkat kemiripan tersebut diantara *document*.

Kata Kunci : *Tabel Frekuensi dan Barchart, Document, perbandingan, Kemiripan*

DAFTAR ISI

PERNYATAAN KEASLIAN TUGAS AKHIR	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH	vi
INTISARI.....	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
BAB 1 PENDAHULUAN	
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	1
1.3. Batasan Masalah	2
1.4. Tujuan Penelitian	2
1.5. Metode Penelitian	2
1.6. Sistematika Penulisan	3
BAB 2 LANDASAN TEORI	
2.1. Tinjauan Pustaka	5
2.2. Landasan Teori.....	6
2.2.1 Dokumen Teks.....	6
2.2.2 Removing Stop Words.....	9
2.2.3 Normalisasi Karakter Huruf.....	9
2.2.4 Rumusan Statistik	10
2.2.5 Ukuran Statistik.....	11
2.2.6 Tabel Frekuensi.....	12
2.2.7 Barchart.....	13
BAB 3 ANALISIS DAN PERANCANGAN SISTEM	
3.1. Pemilihan Bahasa Pemrograman	15

3.2. Gambaran Kerja Sistem	15
3.3. Perancangan Antarmuka Sistem	17
3.3.1 Perancangan Form Input	17
3.3.2 Perancangan Form Output	18
3.4. Perancangan Form Tambahan	19
3.4.1 Perancangan Form Utama	19
3.4.2 Perancangan Form About	20
3.4.3 Perancangan Form keluar	20
3.5 Spesifikasi Sistem	23
3.5.1 Perangkat Lunak.....	23
3.5.2 perangkat keras.....	23
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM	
4.1. Implementasi Sistem	24
4.1.1 Implementasi Rancangan Tampilan	24
4.1.1.1 Form Utama	24
4.1.1.2 Form Input	25
4.1.1.3 Form Output	27
4.1.1.4 Form Tambahan.....	31
4.1.1.4.1 Form Dinamis	31
4.1.1.4.2 Form About.....	34
4.2. Analisis Sistem	35
4.2.1. Pengujian Pada Topik Olahraga.....	35
4.2.2. Pengujian Pada Topik Ekonomi.....	36
4.2.3. Pengujian Pada Topik Politik.....	37
4.2.4. Pengujian Pada Topik Hiburan	38
4.2.5. Pengujian Pada Topik Kesehatan.....	39
4.2.6. Pengujian Pada Topik Teknologi	40
4.2.7. Pengujian Pada Topik Kuliner	41
BAB 5 KESIMPULAN DAN SARAN	
5.1. Kesimpulan	42

5.2. Saran	42
DAFTAR PUSTAKA	44

© UKDW

DAFTAR GAMBAR

Gambar	Keterangan	Halaman
2.1	Pencarian Dokumen yang Mirip	8
2.2	Proses Pengelompokan Dokumen	8
2.3	Proses Normalisasi	10
2.4	Tabel Frekuensi dan Matriks A	13
2.5	Barchart	14
3.1	Flowchart Program Secara Umum	16
3.2	Flowchart Algoritma Jarak antar Dokumen	17
3.3	Perancangan Form Input	18
3.4	Perancangan Form Output	19
3.5	Perancangan Form Utama	19
3.6	Perancangan Form About	20
3.7	Perancangan Menu Keluar	20
4.1	Form Utama Aplikasi	25
4.2	Form Tampilan Input	25
4.3	Form tampilan Input	26
4.4	Form Tampilan Input Document	26
4.5	Command Browse Document	27
4.6	Tampilan Saat menampilkan Document	27
4.7	Command Menampilkan Document	28
4.8	Tampilan Token	28
4.9	Command Proses Tokenisasi	28
4.10	Command tabel Frekuensi	29
4.11	Barchart	29
4.12	Commnad Barchart	30
4.13	Command Perbandingan Kemiripan Document	31
4.14	Tampilan Form Stopwords	32
4.15	Tampilan Form Jenis Kata	32
4.16	Tampilan Form Kata	33
4.17	Tampilan Form Document	34
4.18	Tampilan Form About	34

DAFTAR TABEL

Tabel	Keterangan	Halaman
2.1	Tabel Stopwords	9
2.2	Tabel Data	14
3.1	Tipe Data dari Tabel Dokumen	21
3.2	Tipe Data dari tabel Jenis Kata	22
3.3	Tipe Data dari Tabel Kata	22
3.4	Tipe Data dari Tabel Stopwords	23
4.1	Tabel Frekuensi	29
4.2	Tabel Perhitungan Document	31
4.3	Pengujian Topik 1	35
4.4	Pengujian Topik 2	36
4.5	Pengujian Topik 3	37
4.6	Pengujian Topik 4	38
4.7	Pengujian Topik 5	39
4.8	Pengujian Topik 6	40
4.9	Pengujian Topik 7	41



BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam kehidupan sehari-hari, masyarakat modern tentu sudah terbiasa dengan dokumen teks. Dokumen teks merupakan suatu karakter-karakter yang menyatakan kata-kata atau lambang-lambang untuk berkomunikasi oleh manusia dalam bentuk tulisan. Namun terkadang beberapa dokumen memiliki kemiripan dalam isinya. Kemiripan isinya ditinjau dari kesamaan kata yang terkandung di dalam dokumen tersebut. Kemiripan dokumen yang terjadi dapat menimbulkan dugaan plagiasi atau penjiplakan.

Plagiasi adalah pemakaian dan penggunaan ide atau hasil karya orang lain yang diklaim sebagai hasil karyanya sendiri. Pemakaian istilah plagiasi juga biasanya dipakai untuk penggunaan karya orang lain tanpa menyebutkan sumbernya atau identitas si pembuat karya tersebut. Oleh karena itu diperlukan suatu sistem yang mampu membantu dalam mengkalkulasi frekuensi kemunculan kata-kata yang ada pada suatu dokumen untuk menganalisis tingkat kemiripannya.

Dalam penelitian ini akan dianalisis kemiripan dokumen untuk diketahui tingkat kemiripan yang terdapat diantara dokumen, untuk itu digunakan tabel *Frekuensi* dan *Barchart*.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah di atas, rumusan masalah dari penelitian ini adalah bagaimana menerapkan tabel *Frekuensi* dan *Barchart* dalam pengujian untuk dapat menemukan tingkat kemiripan dokumen, yaitu menemukan tingkat persentase kosa kata yang mirip diantara dua dokumen.

1.3 Batasan Masalah

Pada proposal tugas akhir ini, Penulis membatasi permasalahan pada :

- a. Rumusan Statistik yang digunakan berupa, rata-rata, jarak antar dua dokumen, tabel *Frekuensi* dan *Barchart*
- b. Jenis kata yang dipakai yaitu kata benda, kata kerja, kata sifat, kata keterangan, kata ganti, kata bilangan, kata tugas
- c. Dokumen yang diuji berupa dokumen yang berbentuk teks, berbahasa indonesia, dan yang berformat *.txt

1.4 Tujuan Penelitian

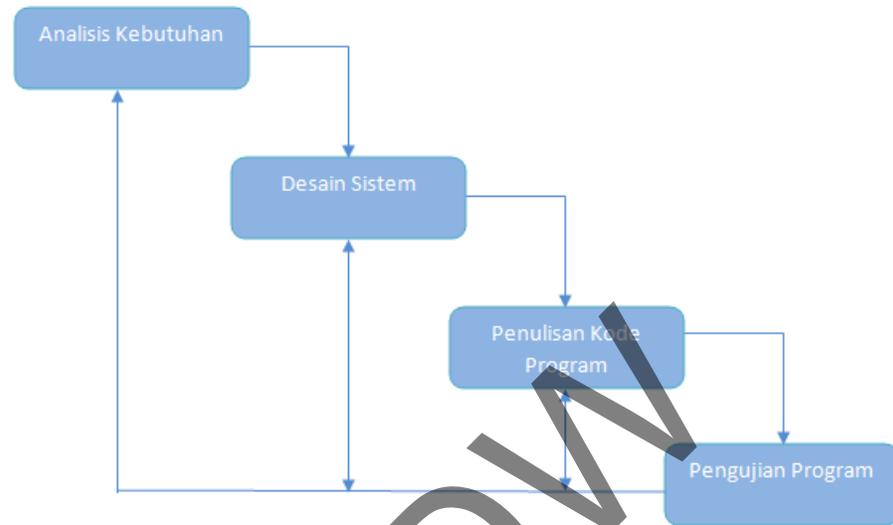
Tujuan utama dari penulisan tugas akhir ini adalah untuk menghasilkan sebuah program uji kemiripan dokumen dengan menggunakan tabel *Frekuensi* dan *Barchart*, sehingga dapat membantu untuk menemukan tingkat kemiripan pada *document*.

1.5 Metode Penelitian

Metode penelitian yang akan digunakan pada pembuatan program pengujian tingkat kemiripan dokumen adalah sebagai berikut :

- a. Studi pustaka, dilakukan dengan cara mencari informasi dan materi dari buku maupun internet yang berhubungan dengan tabel *Frekuensi* dan *Barchart* beserta rumusan statistik dan penerapannya pada pembuatan program, *user interface* yang baik untuk digunakan, serta buku-buku pemrograman *Visual Basic.Net*.
- b. Implementasi tabel *Frekuensi* dan *Barchart* beserta rumusan statistik, berfungsi sebagai basis kecerdasan pada sistem sehingga sistem mampu menyelesaikan permasalahan yang ada.
- c. Metodologi Penyelesaian Masalah, dengan menggunakan metode *Waterfall*, karena dengan menggunakan metode ini tahap demi tahap yang dilalui harus menunggu selesainya tahap sebelumnya dan berjalan berurutan. Metode ini dimulai secara urut, yaitu mulai dari level kebutuhan sistem lalu menuju ke

tahap analisis, desain, *coding*, dan testing. Secara umum tahapan pada model *Waterfall* dapat dilihat pada Gambar 1.1 :



Gambar 1.1 Model *Waterfall*

1.6 Sistematika Penulisan

Penulisan laporan tugas akhir ini terbagi menjadi 5 bagian utama. Bab 1 merupakan pendahuluan, yang berisi tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metode penulisan dan sistematika penulisan. Bab 1 juga memberikan gambaran umum mengenai penelitian tugas akhir ini.

Bab 2 merupakan tinjauan pustaka, yang berisi tentang landasan teori yang mendasari perancangan dan pembuatan program. Pada bab ini landasan teori akan membahas mengenai tabel *Frekuensi* dan *Barchart* serta rumusan dari statistik dan penerapannya dalam uji kemiripan dokumen.

Bab 3 akan membahas tentang perancangan sistem, yang meliputi pemilihan bahasa pemrograman, perancangan basis data, perancangan antar muka (*interface*) serta perancangan input output program.

Bab 4 berisi tentang implementasi dan analisis sistem, yang meliputi tampilan sistem yang telah dibuat lengkap beserta penjelasan dan analisis dari sistem yang telah dibuat.

Bab 5 merupakan bagian terakhir yang berisi kesimpulan dan saran. Kesimpulan merupakan jawaban dari pertanyaan penelitian yang ditanyakan dalam perumusan masalah, sementara saran berisi kesimpulan yang perlu ditindak lanjuti atau direalisasikan dikemudian hari.

© UKDW

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan perancangan sistem pada Bab 3 dan implementasi serta pengujian sistem pada Bab 4, maka dapat diperoleh kesimpulan sebagai berikut:

1. Penggunaan Tabel Frekuensi dan Barchart yang diterapkan pada aplikasi pengujian *Document* mampu menemukan tingkat kemiripan yang diharapkan, yakni rata-rata kedekatan dua dokumen.
2. Rumus *jarak antar dokumen* yang diterapkan membutuhkan waktu yang cukup lama. Hal ini disebabkan karena proses perhitungan harus menunggu sistem untuk *me-load* semua data yang ada, sehingga mempengaruhi kinerja sistem.
3. Document yang dikatakan mirip oleh sistem belum tentu mirip. Hal ini dikarenakan sistem hanya melihat berdasarkan frekuensi dan jenis katanya.
4. Semakin besarnya *database* yang mencapai 11.000 korpus atau semakin banyaknya *korpus* yang disimpan akan membuat kecepatan sistem menjadi berkurang mendekati hampir 2 menit, hal ini dikarenakan sistem harus *me-load* semua data yang ada di *dabase* untuk ditampilkan.

5.2 Saran

Penggunaan Tabel Frekuensi dan *Barchart* untuk menemukan tingkat kemiripan pada *document* membutuhkan waktu yang cukup lama. Hal ini terjadi bila *database* menyimpan banyak data / *korpus*. Selain itu, aplikasi akan menjadi sedikit berat bila program harus *me-load* data yang jumlahnya diatas puluhan ribu untuk ditampilkan. Oleh karena itu diperlukan pengembangan dengan

menggunakan algoritma lain yang memiliki metode pencarian yang lebih efisien misalnya *Jaro-Winkler Distance*.

Aplikasi perbandingan *document* ini hanya memiliki kemampuan untuk mencari kemiripan kata berdasarkan frekuensinya dan jenis katanya. Oleh karena itu, untuk pengembangan lebih lanjut misalnya menambahkan konsep *konteks* pada aplikasi, hal ini bisa dijadikan bahan pertimbangan.

© UKDW

DAFTAR PUSTAKA

- A.H.Nasoetion, A.Rambe. (1984). *Teori Statistika Untuk Ilmu-ilmu Kuantitatif*. Jakarta : Bhatara Karya Aksara.
- Anto, Dajan. (1986). *Pengantar Metode Statistik Jilid 1*. Jakarta : Inter Cipta Prajasa.
- Dodit Supriyanto. (2010). *Membuat Aplikasi Desktop menggunakan MySQL & VB.NET*. Jakarta: mediakita.
- Evelin, Tjong. (2001). *Implementasi Text Mining untuk Mendeteksi Kemiripan Dokumen*. Yogyakarta : Universitas Kristen Duta Wacana.
- Harip, Santoso. (2005). *Membangun Aplikasi.NET menggunakan VB.NET 2005*. Jakarta : Elex Media Komputindo.
- Han, J., Kamber, M. (2001). *Data Mining Concepts and Techniques*.
- Kantardzic M. (2003). *Data Mining Concepts, Models, Methods and Algorithms*.
- Ledy, Agusta (2001). *Perbandingan Algoritma Nazief & Adriani dengan Algoritma Porter untuk Proses Stemming pada Teks Berbahasa Indonesia*. Yogyakarta : Universitas Kristen Duta Wacana.
- Sudaryono, A. Saefullah, U. Rahardja. (2010). *Statistik Deskriptif for IT*. Yogyakarta : Andi offset.
- Tanty, Susanty. (1998). *Implementasi Algoritma K-Mean Pada Pencarian Document Teks*. Yogyakarta : Universitas Kristen Duta Wacana.
- Yesti, Meidayati. (2004). *Analisis Tingkat Kemiripan Dokumen*. Yogyakarta : Universitas Kristen Duta Wacana.