

**PEMBANGUNAN APLIKASI PENCARIAN DOKUMEN
MENGUNAKAN *TEXT MINING*
BERBASIS WEB**

TUGAS AKHIR

**Diajukan Untuk Memenuhi Sebagian Persyaratan Mencapai
Derajat Sarjana Teknik Informatika**



Oleh:

RUTH MEGA ULINA. D

09 07 05980

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ATMA JAYA YOGYAKARTA
2013**

HALAMAN PENGESAHAN

SKRIPSI BERJUDUL

**PEMBANGUNAN APLIKASI PENCARIAN DOKUMEN
MENGUNAKAN *TEXT MINING* BERBASIS WEB**

Disusun oleh :

Ruth Mega Ulina. D (NIM : 09 07 05980)

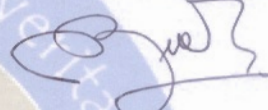
Dinyatakan telah memenuhi syarat
pada tanggal : 06 Januari 2014

Pembimbing I,



Paulus Mudjihartono S.T., M.T.

Pembimbing II,



Dr. Pranowo, S.T., M.T.

Tim Penguji:

Penguji I,



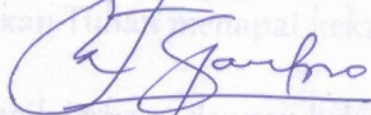
Paulus Mudjihartono S.T., M.T.

Penguji II,



Eduard Risdianto, S.T., M.T.

Penguji III,



Dr. Ir. Alb. Joko Santoso, M.T.

Yogyakarta, 06 Januari 2014
Universitas Atma Jaya Yogyakarta
Fakultas Teknologi Industri
Dekan,



Ir. B. Kristyanto, M.Eng., Ph.D.,
FAKULTAS
TEKNOLOGI INDUSTRI

MOTTO

karena Tuhanlah yang memberikan hikmat,
dari mulut-Nya datang pengetahuan dan
kepandaian.

(Amsal 2:6)

Serahkanlah perbuatanmu kepada Tuhan, maka
terlaksanalah
segala rencanamu

(Amsal 16:3)

You say: "I'm not smart enough"
God says: "I give you wisdom"

Tetapi orang-orang yang menanti-nantikan Tuhan
mendapat kekuatan baru: mereka seumpama rajawali
yang naik terbang dengan kekuatan sayapnya;
mereka berlari dan tidak menjadi lesu, mereka
berjalan dan tidak menjadi lelah.

(Yesaya 40:31)

KATA PENGANTAR

Puji dan syukur kepada Tuhan Yesus Yang Maha Baik atas limpahan berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik. Tujuan dari Pembuatan Tugas Akhir ini adalah sebagai salah satu syarat untuk mencapai derajat sarjana Teknik Informatika dari Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta.

Penulis menyadari bahwa dalam pembuatan skripsi ini tidak terlepas dari bantuan berbagai pihak yang telah menyumbangkan pikiran, tenaga, dukungan, bimbingan, dan doa kepada penulis baik secara langsung maupun tidak langsung. Oleh sebab itu, penulis mengucapkan terima kasih kepada :

1. Bapaku yang di Sorga, Jesus Christ yang telah memberikan berkatNya yang luar biasa dihidupku. Yang selalu setia serta memberikan kekuatan dan ketenangan didalam diriku, terutama dalam penyusunan Tugas Akhir ini.
2. Bapak Ir. B. Kristyanto, M.Eng., Ph.D., selaku Dekan Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta.
3. Bapak Paulus Mudjihartono, S.T., M.T., selaku Dosen Pembimbing I yang menawarkan judul skripsi kepada penulis serta dorongan dan bimbingannya hingga skripsi ini dapat diselesaikan.

4. Bapak Dr. Pranowo, S.T., M.T., selaku Dosen Pembimbing II atas dorongan dan bimbingannya hingga skripsi ini dapat diselesaikan.
5. Seluruh dosen yang pernah mengajar dan membimbing penulis selama kuliah di Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta.
6. Orang tuaku tercinta, yang selalu memberikan dukungan dan doa.
7. Abang terkasih Alek Sander Groyko Damanik S.T dan Leonardo Damanik S.T, terimakasih untuk dukungan moral dan materil.
8. Kakak terkasih Dameria Damanik yang selalu memberikan semangat serta doa, dan Nora Evalina Damanik S.E, terimakasih untuk dukungan yang tidak terlihat :p
9. Adik cantik Rani Anggi Damanik yang 'agak' alay :p, terimakasih buat video-video yang menceritakan suasana rumah di pekanbaru, yang menjadikan penyemangat disaat pengerjaan Tugas Akhir ini *stuck* ☺.
10. Eda tersayang Martanti Indah Pangaribuan S.T, terimakasih untuk dukungan, dan waktu-waktu yang ada untuk liburan dan *shopping* bersama :D.
11. Keponakan-keponakan yang membanggakan, Steven, Tessa, Angel, Oscar, Gideon, Christine, dan Dedek kecil yang baru lahir, trimakasih untuk keceriaan kalian, rasa stress bou/tante hilang ♥♥.

12. Sahabatku dan Keluarga kedua selama di yogyakarta dari awal semester, Kevin, Bang Kris, Dika, Janu, Bowo, Ester, Funny, Ell, trimakasih untuk suka duka bersama selama di daerah istimewa ini ☺. Teman-teman beserta warga sekitar Lokasi KKN, terutama teman-teman di kelompok Gedang Kltuk, Rini, Ita, Aldo, Deden, Vetta, Robby, Hendra, Ribon. Terimakasih untuk kenangannya ☺. Student Staff Perpustakaan, khususnya di ruangan Local Content non nita dan jeng prima :D, terimakasih untuk terus nyemangatin aku, kalian juga harus semangat \m/. Buat pak rudy terimakasih atas izin yang bapak berikan selama Tugas Akhir dikerjakan, maaf saya suka izin pak hehehehe...
13. Laptop Acer 4750G yang berulang-ulang instal windows dalam penyelesaian Tugas Akhir ini.
14. Motor warisan AB 2263 K yang selalu menemani hari-hariku ke kampus, dari bimbingan hingga pendadaran. Yang terkadang bensinnya suka bocor tanpa sebab ☹.
15. Garden Of Prayer (GOP) tempat yang begitu nyaman dimana aku curahat dengan Tuhanku mengenai segala hal yang aku alami selama berada di daerah istimewa ini ☺.
16. Teman-teman TF angkatan 2009 atas kebersamaannya selama masa kuliah dan juga semester akhir saya ini, serta atas sharing ilmunya selama ini.
17. Semua pihak diluar sana yang tidak dapat disebutkan satu persatu, selalu menanyakan kabar Tugas Akhir. Terimakasih untuk semangat dan dorongannya, karena

tanpa kalian saya tidak akan pernah terpacu untuk menyelesaikan Tugas Akhir ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna karena adanya keterbatasan waktu dan pengetahuan yang dimiliki penulis. Oleh karena itu segala kritik dan saran yang bersifat membangun sangat diharapkan. Akhir kata, semoga skripsi ini dapat berguna dan bermanfaat bagi semua pihak.

Yogyakarta, Desember 2013

Penulis

HALAMAN PERSEMBAHAN

THIS THESIS ARE DEDICATED FOR:

MY SAVIOUR JESUS CHRIST...

MY MOM & MY DAD

ALL MY BROTHERS & SISTERS...

and also to

ALL MY BEST FRIENDS...

Daftar Isi

KATA PENGANTAR	iv
Daftar Isi	ix
Daftar Gambar	xi
Intisari	xiv
BAB I	1
PENDAHULUAN	1
I.1. Latar Belakang Masalah	1
I.2. Rumusan Masalah	2
I.3. Tujuan	3
I.4. Batasan Masalah	3
I.5. Metode Penelitian	3
I.6. Sistematika Penulisan	4
BAB II	7
TINJAUAN PUSTAKA	7
BAB III LANDASAN TEORI	11
III.1 Penambangan Teks (Text Mining)	11
III.2 Algoritma TF-IDF	15
III.3 Cosine Similarity	16
BAB IV	21
ANALISIS DAN PERANCANGAN SISTEM	21
Gambar 2.2 Class Diagram	46
BAB V	70
IMPLEMENTASI SISTEM, DAN PENGUJIAN SISTEM	70
V.1 Definisi Sistem	70

V.2	Implementasi Sistem	71
V.2.7	Antarmuka Display Dokumen.....	75
V.3	Hasil Pengujian Perangkat Lunak	78
IV	Tabel 5.1 Hasil Pengujian Perangkat Lunak	78
BAB VI	84
KESIMPULAN DAN SARAN	84
VI.1	Kesimpulan	84
VI.2	Saran	84
Daftar Pustaka	85
LAMPIRAN	87

Daftar Gambar

Gambar 3.2: Proses Floding	13
Gambar 3.3: Proses Tokenizing dan Filter	14
Gambar 3.4: Proses Stemming	14
Gambar 3.5: Ilustrasi Algoritma Penambahan	15
Gambar 1. Arsitektur Perangkat lunak ARS	22
Gambar 2. Use Case Diagram	24
Gambar 3. Entity Relationship Diagram (ERD)	39
Gambar 2.1.1: Login	40
Gambar 2.1.2.1 : Entri Data Dokumen	40
Gambar 2.1.2.2 : Hapus Data Dokumen	41
Gambar 2.1.3.1 : Entri Data Keyword	41
Gambar 2.1.4.3: Hapus Data Keyword	42
Gambar 2.1.4.1 : Entri Data Kamus	42
Gambar 2.1.4.2 : Edit Data Kamus	43
Gambar 2.1.4.3: Hapus Data Kamus	43
Gambar 2.1.5.1 : Entri Data Bad Words	44
Gambar 2.1.5.2 : Hapus Data Bad Words	44
Gambar 2.6.4 : Display Dokumen	45
Gambar 2.3.2 Antarmuka Halaman Administr	61
Gambar 2.3.3a Antarmuka Halaman Pengelolaan Data Dokumen ...	62
Gambar 2.3.3b Antarmuka Add Data Dokumen	62
Gambar 2.3.3c Antarmuka Hapus Data Dokumen	63

Gambar 2.3.4a Antarmuka Halaman Pengelolaan Data Kamus	64
Gambar 2.3.4b Antarmuka Add Data Kamus	64
Gambar 2.3.4c Antarmuka Halaman Ubah Data Kamus	65
Gambar 2.3.4d Antarmuka Halaman Hapus Data Kamus	65
Gambar 2.3.5b Antarmuka Halaman Tambah Data Bad Words	67
Gambar 2.3.5c Antarmuka Hapus Data Bad Words	67
Gambar 2.3.5 Halaman Utama User	68
Gambar 2.3.6 Antarmuka Hasil Pencarian	68
Gambar 5.1 Antarmuka Login	71
gambar 5.2 antarmuka pengelolaan dokumen	72
Gambar 5.3 Antarmuka Pengelolaan Kamus	72
Gambar 5.4 Antarmuka Kata Tidak Penting	73
Gambar 5.5a Antarmuka Keyword	73
Gambar 5.5b Antarmuka Keyword	74
Gambar 5.6 Antarmuka Pengguna	75
Gambar 5.7 Antarmuka Display Dokumen	75

Daftar Tabel

Tabel 2.1 perbandingan sistem yang telah ada dengan yang akan dibangun	10
Tabel 3.1 Perhitungan TF/IDF	17
Tabel 5.3 Hasil Pengujian Perangkat Lunak	78

Intisari

Pada saat ini informasi sangatlah mudah didapat salah satunya adalah dari internet kita dapat mendapatkan informasi yang sangat luas. Dengan banyaknya informasi yang didapat maka akan menyulitkan dalam menemukan dokumen seperti yang diinginkan.

Pada prinsipnya, penyimpanan dokumen teks dan proses pencarian kembali dokumen tersebut sifatnya sederhana, selama ada kumpulan dokumen yang disimpan dan pengguna yang memberikan pertanyaan ataupun kebutuhan

Penelitian ini menggunakan metode cosine similaritas untuk menghitung kesamaan kata dalam dokumen. Algoritma yang dikembangkan diuji dengan menggunakan dokumen teks. Hasil uji menunjukkan bahwa algoritma dapat digunakan untuk menghitung tingkat similaritas (kesamaan) dokumen berdasarkan kata kunci yang diinputkan oleh pengguna. Pemilihan kata kunci sangat mempengaruhi hasil pencarian dokumen teks.

Kata kunci: similaritas, dokumen teks