

**SISTEM PENCARIAN HASIL PENELITIAN SKRIPSI
MENGUNAKAN *WEIGHTED TREE SIMILARITY* DAN
*CONTENT BASED FILTERING***

SKRIPSI



Oleh :

NATALIA MATONDANG

20081010054

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2024**

**SISTEM PENCARIAN HASIL PENELITIAN SKRIPSI
MENGUNAKAN *WEIGHTED TREE SIMILARITY* DAN
*CONTENT BASED FILTERING***

SKRIPSI

Diajukan Untuk Memenuhi Sebagai Persyaratan Dalam Menempuh Gelar Sarjana
Komputer Program Studi Informatika



Oleh :

NATALIA MATONDANG

20081010054

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2024**

LEMBAR PENGESAHAN SKRIPSI

Judul : SISTEM PENCARIAN HASIL PENELITIAN SKRIPSI
MENGUNAKAN *WEIGHTED TREE SIMILARITY* DAN
CONTENT BASED FILTERING
Oleh : NATALIA MADTONDANG
NPM : 20081010054

Telah Diseminarkan Dalam Ujian Skripsi
Pada : Hari Kamis, Tanggal 4 Juli 2024

Mengetahui

Dosen Pembimbing

1.

Yisti Vita Via, S.ST. M.Kom

NIP. 19860425 2021212 001

Dosen Penguji

1.

Dr. Ir. Kartini, S. Kom. MT

NIP. 19611110 199103 2 001

2.

Fawwaz Ali Akbar, S.Kom. M.Kom

NIP. 19920317 2018031 002

2.

Eka Prakarsa Mandyartha, S.T., M.Kom

NIP. 19880525 2018031 001

Menyetujui

Dekan

Fakultas Ilmu Komputer

Prof. Dr. Ir. Novirina Hendrasarie, MT

NIP. 19681126 199403 2 001

Koordinator Program Studi Informatika

Fetty Tri Anggraeny, S.Kom. M.Kom

NIP. 19820211 2021212 005

SURAT PERNYATAAN BEBAS DARI PLAGIASI

Saya, mahasiswa Program Studi Sarjana Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang bertanda tangan di bawah ini:

Nama : Natalia Matondang

NPM : 20081010054

Menyatakan dengan sesungguhnya bahwa Skripsi/Tugas Akhir yang saya kerjakan berjudul:

“SISTEM Pencarian Hasil Penelitian Skripsi Menggunakan *WEIGHTED TREE SIMILARITY* DAN *CONTENT BASED FILTERING*”

bukan merupakan plagiasi sebagian atau keseluruhan dari Skripsi/Tugas Akhir/ Penelitian orang lain dari juga bukan merupakan produk dan software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/Tugas Akhir ini secara keseluruhan adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun di Institut Pendidikan lain. Bukti hasil pengecekan plagiasi dokumen ini dapat ditelusuri melalui QR Code di bawah.

Apabila di kemudian hari terbukti bahwa dokumen ini merupakan plagiasi karya orang lain, saya sanggup menerima sanksi sesuai aturan yang berlaku.

Demikian atas perhatiannya disampaikan terima kasih.

Surabaya, 15 Juli 2024

Hormat saya,



Natalia Matondang

NPM. 20081010054

SISTEM PENCARIAN HASIL PENELITIAN SKRIPSI MENGGUNAKAN *WEIGHTED TREE SIMILARITY* DAN *CONTENT BASED FILTERING*

Nama Mahasiswa : Natalia Matondang

NPM : 20081010054

Program Studi : Informatika

Dosen Pembimbing : Yisti Vita Via, S.ST. M.Kom

Fawwaz Ali Akbar, S.Kom, M.Kom

ABSTRAK

Sistem pencarian skripsi sangat dibutuhkan untuk membantu mahasiswa menemukan skripsi terdahulu serta mengurangi peluang terjadinya plagiarisme. UPNVJT telah memiliki sistem *repository* untuk menyimpan data skripsi, tetapi fitur pencarian sistem tersebut belum dapat memberikan hasil yang relevan. Hal ini tidak efisien apabila mahasiswa ingin mencari berdasarkan topik. Oleh karena itu dibutuhkan sistem pencarian untuk membantu mahasiswa menemukanskripsi yang sesuai dengan topik yang ingin dibuat.

Metode yang digunakan pada penelitian ini ialah *content based filtering* dan *weighted tree similarity*. *Content based filtering* melakukan filter berdasarkan atribut data agar hasil pencarian memiliki konten yang sesuai. *Weighted Tree Similarity* membantu meningkatkan kinerja sistem dengan membentuk pohon yang memiliki bobot dengan sehingga hasil pencarian lebih akurat. Untuk menguji kinerja sistem dilakukan pengujian menggunakan *precision* dan *recall*. Hasil pengujian menunjukkan bahwa sistem dengan menggunakan TF-IDF menghasilkan pencarian yang lebih akurat dibandingkan dengan sistem yang menggunakan TF dengan mengalami peningkatan sebanyak 28,5%. Selain itu, hasil pengujian menunjukkan bahwa Panjang query tidak berpengaruh untuk meningkatkan akurasi pencarian. Pada sistem yang menggunakan TF, tingkat akurasi terus menurun seiring dengan penambahan kata pada query. Sedangkan pada sistem yang menggunakan TF-IDF, akurasi tertinggi didapat dengan penggunaan query 3 kata.

***Kata kunci:* Pencarian, Skripsi, WTS, CBF**

KATA PENGANTAR

Dengan penuh rasa syukur kepada Tuhan Yang Maha Esa atas limpahan rahmat, kebaikan, dan anugerah-Nya, penulis dapat menyelesaikan skripsi ini yang berjudul:

“SISTEM PENCARIAN HASIL PENELITIAN SKRIPSI MENGGUNAKAN *WEIGHTED TREE SIMILARITY* DAN *CONTENT BASED FILTERING*”

Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) di Jurusan Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Penulis mendapatkan banyak dukungan, bantuan, bimbingan, dan nasihat dari berbagai pihak selama proses penulisan skripsi ini. Dengan penuh rasa hormat dan terima kasih, penulis menyampaikan apresiasi kepada semua pihak yang telah membantu dari awal hingga akhir penulisan skripsi ini.

Penulis menyadari bahwa penelitian ini masih memiliki kekurangan karena keterbatasan pengetahuan dan kemampuan penulis. Oleh karena itu, penulis menerima segala bentuk kritik dan saran dari berbagai pihak untuk penyempurnaan laporan skripsi ini sehingga dapat memberikan manfaat bagi penulis dan para pembaca.

Surabaya, 4 Juli 2024

Penulis

Natalia Matondang

UCAPAN TERIMA KASIH

Penyusunan skripsi ini dapat diselesaikan berkat bantuan, bimbingan, dukungan, dan nasihat dari berbagai pihak. Penulis dengan tulus mengucapkan terima kasih kepada Tuhan Yang Maha Esa atas belas kasih dan kebaikan-Nya yang memungkinkan penulis menyelesaikan skripsi ini. Penulis juga berterima kasih kepada semua pihak yang telah membantu, baik secara langsung maupun tidak langsung.

1. Tuhan Yang Maha Esa, yang telah memberikan banyak keajaiban dan bantuan dalam penyusunan skripsi ini.
2. Kedua Orang Tua dan Keluarga, yang telah memberikan nasihat, doa, kasih sayang, semangat, serta memenuhi kebutuhan dan fasilitas untuk memudahkan perkuliahan penulis.
3. Prof. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom, selaku Koordinator Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Yisti Vita Via, S.ST, M.Kom, selaku Dosen Pembimbing Pertama, yang telah membantu dalam penelitian skripsi serta memberikan masukan dan saran selama proses penulisan.
6. Fawwaz Ali Akbar, S.Kom, M.Kom, selaku Dosen Pembimbing Kedua, yang telah membantu dalam penelitian skripsi serta memberikan masukan dan saran selama proses penulisan.
7. Teman kos Ais, Sintya, Salma, Indy yang selalu membantu saya ketika kesulitan.
8. Naposobulung HKBP Rungkut Ekle, Icha, Lidya, Kristin, Clara, Marvin, Dela, Kenny, Sam, Dwi, Catrine, Kak Evelyn
9. Semua pihak yang telah membantu dan tidak dapat disebutkan satu per satu.

Penulis menyadari masih ada kekurangan dalam penelitian ini dan menerima kritik serta saran dengan terbuka. Semoga penelitian ini bermanfaat bagi banyak orang dan dapat dikembangkan lebih baik lagi di masa mendatang.

DAFTAR ISI

SURAT PERNYATAAN BEBAS DARI PLAGIASI.....	ii
ABSTRAK	iii
KATA PENGANTAR	iv
UCAPAN TERIMA KASIH.....	v
DAFTAR ISI.....	vi
DAFTAR TABEL	ix
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan	3
1.4. Manfaat	4
1.5. Batasan Masalah.....	4
BAB II TINJAUAN PUSTAKA	5
2.1. Penelitian terdahulu	5
2.2. Sistem Pencarian	8
2.3. Sistem Temu Kembali Informasi.....	8
2.4. Text Preprocessing.....	9
2.5. Pembobotan TF-IDF	10
2.6. Content Based Filtering	12
2.7. Cosine Similarity	12
2.8. Weighted Tree Similarity.....	12
2.9. Precision Score	13
2.10. Recall	14

2.11. NLTK (Natural Language Toolkit).....	14
2.12. Pandas	14
2.13. Numpy.....	15
2.14. Sastrawi.....	15
2.15. Pickle.....	16
2.16. Flask	16
BAB III METODOLOGI.....	17
3.1. Tahapan Penelitian.....	17
3.2. Analisis Masalah.....	17
3.3. Pengumpulan Data.....	18
3.4. Preprocessing Data.....	19
3.4.1. Case folding.....	20
3.4.2. Tokenizing.....	22
3.4.3. Filtering	24
3.4.4. Stemming.....	26
3.5. Perancangan Sistem	27
3.5.1. Content Based Filtering.....	27
3.5.2. Weigted Tree Similarity	34
3.6. Skenario Pengujian	37
BAB IV HASIL DAN PEMBAHASAN	39
4.1. Implementasi Program.....	39
4.1.1. Import Library	39
4.1.2. Preprocessing Data.....	40
4.1.3. Content Based Filtering.....	47
4.1.4. Weighted Tree Similarity	56

4.2. Pengujian.....	59
4.3. Implementasi Antar Muka	213
4.3.1. Tampilan Awal	213
4.3.2. Tampilan Hasil Pencarian	214
4.2.3. Tampilan Detail Skripsi.....	214
BAB V KESIMPULAN DAN SARAN.....	216
5.1. Kesimpulan	216
5.2. Saran	216
DAFTAR PUSTAKA	217

DAFTAR TABEL

Tabel 3. 1. Contoh Case folding	20
Tabel 3. 2 Contoh Tokenizing.....	22
Tabel 3. 3 Contoh Filtering	24
Tabel 3. 4 Contoh Stemming	26
Tabel 3. 5 Contoh Pembobotan Judul TF	28
Tabel 3. 6 Contoh Pembobotan Abstrak TF	30
Tabel 3. 7 Contoh Pembobotan Judul TF-IDF.....	31
Tabel 3. 8 Contoh Pembobotan Abstrak TF-IDF.....	33
Tabel 4. 1 Hasil Pencarian 'Sistem Informasi Enterprise' dengan TF.....	60
Tabel 4. 2 Hasil Pencarian ' Analisis dan Pengujian Sistem ' dengan TF	66
Tabel 4. 3 Hasil Pencarian ' Sistem Informasi Geografis ' dengan TF.....	73
Tabel 4. 4 Hasil Pencarian ' Jaringan dan Keamanan Informasi ' dengan TF.....	79
Tabel 4. 5 Hasil Pencarian ' Komputasi Cerdas ' dengan TF	83
Tabel 4. 6 Hasil Pencarian ' Komputasi Visual ' dengan TF.....	84
Tabel 4. 7 Hasil Pencarian ' Data Mining ' dengan TF	85
Tabel 4. 8 Hasil Pencarian ' Robotika' dengan TF	89
Tabel 4. 9 Hasil Pencarian ' iot' dengan TF.....	90
Tabel 4. 10 Hasil Pencarian ' Animasi dan Gim ' dengan TF	91
Tabel 4. 11 Hasil Pencarian ' Solusi Sistem Informasi' dengan TF	91
Tabel 4. 12 Hasil Pencarian ' Manajemen Sistem Informasi ' dengan TF	97
Tabel 4. 13 Hasil Uji Kategori Skripsi dan Pembobotan TF	103
Tabel 4. 14 Hasil Pencarian 'Sistem Informasi Enterprise' dengan TF-IDF	104
Tabel 4. 15 Hasil Pencarian ' Analisis dan Pengujian Sistem ' dengan TF-IDF ..	105
Tabel 4. 16 Hasil Pencarian ' Sistem Informasi Geografis ' dengan TF-IDF.....	107
Tabel 4. 17 Hasil Pencarian ' Jaringan dan Keamanan Informasi ' dengan TF-IDF	108
Tabel 4. 18 Hasil Pencarian ' Komputasi Cerdas ' dengan TF-IDF	109
Tabel 4. 19 Hasil Pencarian ' Komputasi Visual ' dengan TF-IDF	110
Tabel 4. 20 Hasil Pencarian ' Data Mining ' dengan TF-IDF.....	111
Tabel 4. 21 Hasil Pencarian ' Robotika' dengan TF-IDF	112
Tabel 4. 22 Hasil Pencarian ' iot' dengan TF-IDF.....	113

Tabel 4. 23 Hasil Pencarian ' Animasi dan Gim ' dengan TF-IDF	113
Tabel 4. 24 Hasil Pencarian ' Solusi Sistem Informasi' dengan TF-IDF	114
Tabel 4. 25 Hasil Pencarian ' Manajemen Sistem Informasi ' dengan TF-IDF....	116
Tabel 4. 26 Hasil Uji Kategori Skripsi dan Pembobotan TF-IDF.....	117
Tabel 4. 27 Hasil Pencarian ' Jaringan' dengan TF	118
Tabel 4. 28 Hasil Pencarian ' Desain' dengan TF	119
Tabel 4. 29 Hasil Pencarian ' Pengujian' dengan TF	120
Tabel 4. 30 Hasil Pencarian ' Website' dengan TF	125
Tabel 4. 31 Hasil Pencarian ' Android' dengan TF	126
Tabel 4. 32 Hasil Uji dengan Panjang Query 1 Kata dan Pembobotan TF.....	127
Tabel 4. 33 Hasil Pencarian ' Jaringan' dengan TF-IDF	127
Tabel 4. 34 Hasil Pencarian ' Desain' dengan TF-IDF.....	128
Tabel 4. 35 Hasil Pencarian ' Pengujian' dengan TF-IDF.....	129
Tabel 4. 36 Hasil Pencarian ' Website' dengan TF-IDF.....	130
Tabel 4. 37 Hasil Pencarian ' Android' dengan TF-IDF	132
Tabel 4. 38 Hasil Uji dengan Panjang Query 1 Kata dan Pembobotan TF-IDF..	132
Tabel 4. 39 Hasil Pencarian ' Deteksi Penyakit' dengan TF	133
Tabel 4. 40 Hasil Pencarian ' Desain UI/UX' dengan TF	134
Tabel 4. 41 Hasil Pencarian ' Pengujian Sistem' dengan TF.....	134
Tabel 4. 42 Hasil Pencarian ' Keamanan Jaringan' dengan TF	142
Tabel 4. 43 Hasil Pencarian ' Analisis Sentimen' dengan TF	143
Tabel 4. 44 Hasil Uji dengan Panjang Query 2 Kata dan Pembobotan TF.....	145
Tabel 4. 45 Hasil Pencarian ' Deteksi Penyakit' dengan TF-IDF	145
Tabel 4. 46 Hasil Pencarian ' Desain UI/UX' dengan TF-IDF	147
Tabel 4. 47 Hasil Pencarian ' Pengujian Sistem' dengan TF-IDF.....	147
Tabel 4. 48 Hasil Pencarian ' Keamanan Jaringan' dengan TF-IDF	150
Tabel 4. 49 Hasil Pencarian ' Analisis Sentimen' dengan TF-IDF	151
Tabel 4. 50 Hasil Uji dengan Panjang Query 2 Kata dan Pembobotan TF-IDF..	152
Tabel 4. 51 Hasil Pencarian ' Sistem Prediksi Gambar' dengan TF.....	152
Tabel 4. 52 Hasil Pencarian ' Analisis Penerimaan Aplikasi' dengan TF	158
Tabel 4. 53 Hasil Pencarian ' Analisis Kepuasan Sistem' dengan TF.....	162
Tabel 4. 54 Hasil Pencarian ' Pengembangan Gim Edukasi' dengan TF	167

Tabel 4. 55 Hasil Pencarian ‘Sistem Pendukung Keputusan’ dengan TF.....	170
Tabel 4. 56 Hasil Uji dengan Panjang Query 3 Kata dan Pembobotan TF.....	175
Tabel 4. 57 Hasil Pencarian ‘Sistem Pendeteksi Gambar’ dengan TF-IDF.....	175
Tabel 4. 58 Hasil Pencarian ‘Analisis Penerimaan Aplikasi’ dengan TF-IDF	176
Tabel 4. 59 Hasil Pencarian ‘Analisis Kepuasan Sistem’ dengan TF-IDF	178
Tabel 4. 60 Hasil Pencarian ‘Pengembangan Gim Edukasi’ dengan TF-IDF	180
Tabel 4. 61 Hasil Pencarian ‘Sistem Pendukung Keputusan’ dengan TF-IDF....	180
Tabel 4. 62 Hasil Uji dengan Panjang Query 3 Kata dan Pembobotan TF-IDF..	182
Tabel 4. 63 Hasil Pencarian ‘Rancang Aplikasi Berbasis Android’ dengan TF..	182
Tabel 4. 64 Hasil Pencarian ‘Sistem Informasi Berbasis Website’ dengan TF ...	185
Tabel 4. 65 Hasil Pencarian ‘Pendeteksi Penyakit Melalui Gambar’ dengan TF	191
Tabel 4. 66 Hasil Pencarian ‘Evaluasi Keamanan Sistem Jaringan’ dengan TF .	192
Tabel 4. 67 Hasil Pencarian ‘Analisis Penerimaan Sistem Informasi’ dengan TF	198
Tabel 4. 68 Hasil Uji dengan Panjang Query 4 Kata dan Pembobotan TF.....	204
Tabel 4. 69 Hasil Pencarian ‘Rancang Aplikasi Berbasis Android’ dengan TF-IDF	204
Tabel 4. 70 Hasil Pencarian ‘Sistem Informasi Berbasis Website’ dengan TF-IDF	206
Tabel 4. 71 Hasil Pencarian ‘Pendeteksi Penyakit Melalui Gambar ‘ dengan TF- IDF	207
Tabel 4. 72 Hasil Pencarian ‘Evaluasi Keamanan Sistem Jaringan’ dengan TF-IDF	209
Tabel 4. 73 Hasil Pencarian ‘Analisis Penerimaan Sistem Informasi’ dengan TF- IDF	210
Tabel 4. 74 Hasil Uji dengan Panjang Query 4 Kata dan Pembobotan TF-IDF..	212

DAFTAR GAMBAR

Gambar 3. 1. Tahap Penelitian.....	17
Gambar 3. 2 Kumpulan Skripsi di Ruang Baca FASILKOM.....	18
Gambar 3. 3 Data Skripsi	19
Gambar 3. 4 Tahap Preprocessing Data.....	19
Gambar 3. 5 Proses Content Based Filtering	27
Gambar 3. 6 Contoh Tree Judul Q dan Judul D1	34
Gambar 3. 7 Contoh tree skripsi	36
Gambar 4. 1 Dataset dari DataSkripsi.xlsx	41
Gambar 4. 2 Dataset sebelum Case Folding	42
Gambar 4. 3 Dataset setelah Case Folding.....	43
Gambar 4. 4 Hasil Proses Tokenizing.....	44
Gambar 4. 5 Hasil Proses Filtering	45
Gambar 4. 6 Hasil Proses Steaming.....	47
Gambar 4. 7 Hasil Perhitungan TF Judul.....	48
Gambar 4. 8 Hasil Perhitungan TF Abstrak.....	49
Gambar 4. 9 Hasil Perhitungan DF Judul	50
Gambar 4. 10 Hasil Perhitungan DF Abstrak	51
Gambar 4. 11 Hasil Perhitungan IDF Judul	52
Gambar 4. 12 Hasil Perhitungan IDF Abstrak.....	52
Gambar 4. 13 Hasil TF-IDF Judul	53
Gambar 4. 14 Hasil TF-IDF Abstrak	54
Gambar 4. 15 Hasil Penghitungan Cosine Similarity	57
Gambar 4. 16 Hasil Perhitungan Cosine Similarity dan WTS.....	59
Gambar 4. 17 Tampilan Awal.....	213
Gambar 4. 18 Hasil Pencarian Skripsi	214
Gambar 4. 19 Halaman Detail Skripsi	214
Gambar 4. 20 Repository UPN "Veteran" Jawa Timur	215