

ABSTRAK

Ardian Asman

Skripsi

Pengembangan Sistem Pengecekan Kemiripan Topik Skripsi Mahasiswa untuk Website
Pra Skripsi Teknik Informatika Menggunakan Metode TF-IDF dan Teknik Phrase Recognition

Dalam proses penyusunan skripsi teknik informatika penting untuk dipastikan bahwa topik skripsi yang ingin diajukan oleh mahasiswa tidak memiliki kemiripan dengan skripsi yang telah lulus. Namun hingga sekarang, sistem pengecekan pada website yang sudah ada masih belum dapat memberikan jawaban yang akurat. Hal ini dikarenakan sistem pengecekan hanya menggunakan judul skripsi saja. Hal ini dapat berdampak pada topik yang diajukan memiliki judul yang kurang lebih sama namun memiliki isi pembahasan yang berbeda.

Oleh karena itu dibutuhkan sebuah sistem yang dapat melakukan pengecekan kemiripan secara efisien. Sistem pengecekan akan dibuat menggunakan metode TF-IDF dan Phrase Recognition, dimana pengecekan akan dilakukan pada 5 kriteria yaitu judul, kata kunci, abstrak, latar belakang, serta kesimpulan skripsi.

Dari pengujian data, program yang telah dibuat dapat berjalan dengan baik dan membantu untuk pengecekan kemiripan skripsi. Penggunaan metode TF-IDF dan Phrase Recognition dapat memberikan jawaban yang valid serta memiliki nilai rata rata perbedaan hasil program dan perhitungan manual sebesar 2.19% untuk nilai terendah dan 18.91% untuk nilai tertinggi.

Keywords : Pengecekan Kemiripan, Skripsi, *Term Frequency - Inverse Document Frequency*, *Phrase Recognition*

ABSTRACT

Ardian Asman

Undergraduate Thesis

Pengembangan Sistem Pengecekan Kemiripan Topik Skripsi Mahasiswa untuk Website
Pra Skripsi Teknik Informatika Menggunakan Metode TF-IDF dan Teknik Phrase Recognition

In the process of preparing an informatics engineering thesis, it is important to ensure that the thesis topic that the student wishes to submit does not have any similarities to the thesis that has been passed. However, until now, the checking system on the existing website is still unable to provide accurate answers. This is because the checking system only uses the title of the thesis. This can have an impact on the proposed topics having more or less the same title but having different discussion content.

Therefore, we need a system that can check similarities efficiently. The checking system will be created using the TF-IDF and Phrase Recognition methods, where checking will be carried out on 5 criteria, namely title, keywords, abstract, background and conclusion of the thesis.

From data testing, the program that has been created can run well and helps to check the similarity of the thesis. Using the TF-IDF and Phrase Recognition methods can provide valid answers and has an average difference between program results and manual calculations of 2.19% for the lowest value and 18.91% for the highest value.

Keywords : Pengecekan Kemiripan, Skripsi, *Term Frequency - Inverse Document Frequency*, *Phrase Recognition*

DAFTAR ISI

LEMBARAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	iii
KATA PENGANTAR.....	iv
ABSTRAK.....	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xv
DAFTAR SEGMENT PROGRAM.....	xvii
1. PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan.....	2
1.4 Ruang Lingkup.....	2
1.5 Manfaat Penelitian.....	4
1.6 Metodologi Penelitian.....	4
1.7 Sistematika Penulisan.....	5
2. LANDASAN TEORI.....	7
2.1 Tinjauan Pustaka.....	7
2.1.1 TF-IDF (Term Frequency - Inverse Document Frequency).....	7
2.1.2 Phrase Recognition.....	7
2.1.3 Cosine Similarity.....	8
2.2 Tinjauan Studi.....	8
2.2.1 Assessing Short Answers in Indonesia Using Semantic Text Similarity Method and Dynamic Corpus (Hasanah et al, 2020).....	8

2.2.2	Sentence Similarity Measurement for Bengali Abstractive Text Summarization (Masum, 2019)	9
3.	ANALISIS DAN DESAIN SISTEM	10
3.1	Analisis Permasalahan	10
3.2	Analisis Kebutuhan	10
3.3	Alur Kerja Sistem	11
3.3.1	Alur Kerja Text Preprocessing	12
3.3.2	Alur Kerja Proses Perhitungan TF (Term Frequency)	13
3.3.3	Alur Kerja Proses Perhitungan IDF (Inverse Document Frequency).....	15
3.3.4	Alur Kerja Proses Perhitungan Term Frequency - Inverse Document Frequency	17
3.3.5	Alur Kerja Proses Perhitungan Cosine Similarity	19
3.4	Database.....	20
3.4.1	Tabel Dosen.....	20
3.4.2	Tabel Skripsi.....	21
3.4.3	Tabel Jurnal	21
3.4.4	Tabel Proposal.....	22
3.4.5	Tabel Frequently Asked Questions.....	22
3.5	Desain UI.....	23
3.5.1	Judul Proposal Skripsi	23
3.5.2	Pop Up Belum Melakukan Pengecekan Kemiripan	24
3.5.3	Hasil Pengecekan Kemiripan	25
3.5.4	Daftar Pengajuan Proposal Skripsi Mahasiswa (Sisi Dosen).....	26
3.5.5	Detail Proposal Mahasiswa (Sisi Dosen).....	27
4.	IMPLEMENTASI SISTEM.....	28
4.1	Aplikasi Pemrograman	28
4.2	Implementasi Program	28
4.2.1	Koneksi Database	28

4.2.2	Fungsi Back End Pengambilan Data Skripsi	29
4.2.3	Fungsi Back End Pengambilan Data Jurnal	29
4.2.4	Fungsi Back End Upload Skripsi Mahasiswa	30
4.2.5	Fungsi Back End Pengambilan Data Submit Skripsi.....	31
4.2.6	Fungsi Back End Submit Skripsi Mahasiswa	31
4.2.7	Fungsi Back End Add FAQ.....	32
4.2.8	Fungsi Back End Update FAQ	33
4.2.9	Fungsi Back End Delete FAQ.....	34
4.2.10	Fungsi Back End Get FAQ	34
4.2.11	Fungsi Back End Askbot.....	34
4.2.12	Fungsi Front End Fetch Data Awal.....	35
4.2.13	Fungsi Pengambilan Isi Data Skripsi	37
4.2.14	Fungsi Pengambilan Isi Data Skripsi File.....	37
4.2.15	Fungsi Pengambilan Isi Data Skripsi Mahasiswa	42
4.2.16	Fungsi Fetch Data Jurnal.....	42
4.2.17	Fungsi Pengambilan Isi Data Jurnal	43
4.2.18	Fungsi Pengambilan Seluruh Data Jurnal	44
4.2.19	Fungsi Pengecekan Kemiripan.....	47
4.2.20	Fungsi Pemrosesan Text.....	49
4.2.21	Fungsi Perhitungan Kemiripan Skripsi	51
4.2.22	Fungsi Perhitungan Kemiripan Jurnal.....	58
4.2.23	Fungsi Perhitungan <i>Term Frequency</i>	65
4.2.24	Fungsi Perhitungan Inverse Document Frequency	67
4.2.25	Fungsi Perhitungan Term Frequency-Inverse Document Frequency.....	67
4.2.26	Fungsi Perhitungan <i>Cosine Similarity</i>	67
4.2.27	Fungsi Perhitungan Weighted Cosine Similarity	69
4.2.28	Fungsi Submit Skripsi.....	70

5. PENGUJIAN SISTEM	72
5.1 Aplikasi	72
5.1.1 Tampilan Input Data Skripsi	72
5.1.2 Tampilan Hasil Cek Kemiripan	72
5.1.3 Tampilan Askbot FAQ	74
5.1.4 Tampilan Upload Skripsi	74
5.1.5 Tampilan Login Dosen	75
5.1.6 Tampilan Register Dosen	76
5.1.7 Tampilan Halaman Dosen	78
5.1.8 Tampilan Detail Skripsi	78
5.1.9 Tampilan Halaman FAQ	79
5.2 Pengujian Sistem	81
5.2.1 Program Testing 1	81
5.2.1.1 Skenario 1	81
5.2.1.2 Skenario 2	82
5.2.1.3 Skenario 3	84
5.2.1.4 Skenario 4	85
5.2.1.5 Skenario 5	86
5.2.2 Program Testing 2	88
5.2.2.1 Skenario 1	88
5.2.2.2 Skenario 2	89
5.2.2.3 Skenario 3	91
5.2.2.4 Skenario 4	92
5.2.2.5 Skenario 5	94
5.2.3 Uji Coba Bobot	95
5.2.4 Uji Coba Ruang Lingkup	98
5.2.5 Pengujian N-gram	105

6.KESIMPULAN DAN SARAN	113
6.1 Kesimpulan	113
6.2 Saran.....	113
DAFTAR REFERENSI	114

DAFTAR TABEL

Tabel 3.1 Tabel Dosen	21
Tabel 3.2 Tabel Skripsi.....	21
Tabel 3.3 Tabel Jurnal.....	22
Tabel 3.4 Tabel Proposal	22
Tabel 3.5 Tabel Frequently Asked Questions	23
Tabel 5.1 Data Query Skenario 1.....	81
Tabel 5.2 Daftar Hasil Judul Skenario 1	82
Tabel 5.3 Hasil Perbandingan Skenario 1	82
Tabel 5.4 Data Query Skenario 2.....	82
Tabel 5.5 Daftar Hasil Judul Skenario 2	83
Tabel 5.6 Hasil Perbandingan Skenario 2	83
Tabel 5.7 Data Query Skenario 3.....	84
Tabel 5.8 Daftar Hasil Judul Skenario 3	84
Tabel 5.9 Hasil Perbandingan Skenario 3	84
Tabel 5.10 Data Query Skenario 4.....	85
Tabel 5.11 Daftar Hasil Judul Skenario 4	86
Tabel 5.12 Hasil Perbandingan Skenario 4	86
Tabel 5.13 Data Query Skenario 5.....	87
Tabel 5.14 Daftar Hasil Judul Skenario 5	87
Tabel 5.15 Hasil Perbandingan Skenario 5	87
Tabel 5.16 Data Query Skenario 1.....	88
Tabel 5.17 Daftar Hasil Judul Skenario 1	89
Tabel 5.18 Hasil Perbandingan Skenario 1	89
Tabel 5.19 Data Query Skenario 2.....	90
Tabel 5.20 Daftar Hasil Judul Skenario 2	90

Tabel 5.21 Hasil Perbandingan Skenario 2	90
Tabel 5.22 Data Query Skenario 3.....	91
Tabel 5.23 Daftar Hasil Judul Skenario 3	91
Tabel 5.24 Hasil Perbandingan Skenario 3	92
Tabel 5.25 Data Query Skenario 4.....	92
Tabel 5.26 Daftar Hasil Judul Skenario 4	93
Tabel 5.27 Hasil Perbandingan Skenario 4.....	93
Tabel 5.28 Data Query Skenario 5.....	94
Tabel 5.29 Daftar Hasil Judul Skenario 5	94
Tabel 5.30 Hasil Perbandingan Skenario 5	94
Tabel 5.31 Tabel Komposisi Uji Coba Bobot.....	95
Tabel 5.32 Hasil Perhitungan Komposisi 1	96
Tabel 5.33 Hasil Perhitungan Komposisi 2	96
Tabel 5.34 Hasil Perhitungan Komposisi 3	97
Tabel 5.35 Hasil Perhitungan Komposisi 4	97
Tabel 5.36 Hasil Perhitungan Komposisi 5	98
Tabel 5.37 Hasil Perbandingan Ruang Lingkup	99
Tabel 5.38 Tabel Query Pengajuan Threshold	106
Tabel 5.39 Tabel Pencarian Threshold	107
Tabel 5.40 Tabel Query Pengujian N Gram	108
Tabel 5.41 Nilai Kemiripan Tertinggi Tiap Kombinasi N Gram	111

DAFTAR GAMBAR

Gambar 3.1 Flowchart Sistem	11
Gambar 3.2 Flowchart Text Preprocessing	12
Gambar 3.3 Flowchart Proses Perhitungan Term Frequency	13
Gambar 3.4 Flowchart Proses Perhitungan Inverse Document Frequency	16
Gambar 3.5 Flowchart Proses Perhitungan Term Frequency - Inverse Document Frequency...	17
Gambar 3.6 Flowchart Proses Perhitungan Cosine Similarity	19
Gambar 3.7 Entity Relationship Diagram	20
Gambar 3.8 Desain Halaman Judul Proposal Skripsi	23
Gambar 3.9 Desain Pop Up Belum Melakukan Pengecekan Kemiripan	24
Gambar 3.10 Desain Halaman Hasil Pengecekan Kemiripan	25
Gambar 3.11 Desain Halaman Daftar Pengajuan Proposal Skripsi Mahasiswa (Sisi Dosen).....	26
Gambar 3.12 Desain Halaman Detail Proposal Mahasiswa (Sisi Dosen).....	27
Gambar 5.1 Tampilan Halaman Input Data Skripsi	72
Gambar 5.2 Tampilan Hasil Cek Kemiripan List Jurnal	73
Gambar 5.3 Tampilan Hasil Cek Kemiripan List Skripsi	73
Gambar 5.4 Tampilan Askbot FAQ	74
Gambar 5.5 Tampilan Upload Skripsi	75
Gambar 5.6 Tampilan Login Dosen	76
Gambar 5.7 Tampilan Register Dosen.....	77
Gambar 5.8 Tampilan Register Dosen Berhasil	77
Gambar 5.9 Tampilan Halaman Dosen.....	78
Gambar 5.10 Tampilan Detail Skripsi	79
Gambar 5.11 Tampilan Halaman FAQ.....	80
Gambar 5.12 Tampilan Penambahan Data FAQ	80
Gambar 5.13 Ruang Lingkup No 1 Kemiripan Tinggi	104

Gambar 5.14 Ruang Lingkup No 1 Kemiripan Rendah	105
---	-----

DAFTAR SEGMENT PROGRAM

Segment Program 4.1 Koneksi Database.....	28
Segment Program 4.2 Pengambilan Data Skripsi	29
Segment Program 4.3 Pengambilan Data Jurnal	30
Segment Program 4.4 Upload Skripsi Mahasiswa	30
Segment Program 4.5 Pengambilan Data Submit Skripsi	31
Segment Program 4.6 Upload Submitted Mahasiswa	32
Segment Program 4.7 Add FAQ.....	32
Segment Program 4.8 Update FAQ	33
Segment Program 4.9 Delete FAQ.....	34
Segment Program 4.10 Get FAQ.....	34
Segment Program 4.11 Askbot	35
Segment Program 4.12 Fetch Data Awal.....	35
Segment Program 4.13 Fungsi Fetch Data Skripsi	36
Segment Program 4.14 Pengambilan Isi Data Skripsi	37
Segment Program 4.15 Pengambilan Isi Data Skripsi File	37
Segment Program 4.16 Pengambilan Isi Data Skripsi Mahasiswa	42
Segment Program 4.17 Fungsi Fetch Data Jurnal.....	43
Segment Program 4.18 Pengambilan Isi Data Jurnal.....	43
Segment Program 4.19 Pengambilan Seluruh Data Jurnal	44
Segment Program 4.20 Pengecekan Kemiripan	47
Segment Program 4.21 Lowercase	49
Segment Program 4.22 Removal	50
Segment Program 4.23 Stopwords.....	50
Segment Program 4.24 Perhitungan Kemiripan Judul Skripsi	51
Segment Program 4.25 Perhitungan Kemiripan Kata Kunci Skripsi	52

Segmen Program 4.26 Perhitungan Kemiripan Abstrak Skripsi	54
Segmen Program 4.27 Perhitungan Kemiripan Latar Belakang Skripsi.....	55
Segmen Program 4.28 Perhitungan Kemiripan Kesimpulan Skripsi.....	56
Segmen Program 4.29 Perhitungan Kemiripan Judul Jurnal	58
Segmen Program 4.30 Perhitungan Kemiripan Kata Kunci Jurnal	59
Segmen Program 4.31 Perhitungan Kemiripan Abstrak Jurnal	61
Segmen Program 4.32 Perhitungan Kemiripan Latar Belakang Jurnal.....	62
Segmen Program 4.33 Perhitungan Kemiripan Kesimpulan Jurnal.....	63
Segmen Program 4.34 Perhitungan <i>Term Frequency</i>	65
Segmen Program 4.35 Perhitungan <i>Inverse Document Frequency</i>	67
Segmen Program 4.36 Perhitungan <i>Term Frequency-Inverse Document Frequency</i>	67
Segmen Program 4.37 Rumus <i>Cosine Similarity</i>	68
Segmen Program 4.38 Perhitungan <i>Cosine Similarity</i>	68
Segmen Program 4.39 Perhitungan <i>Weighted Cosine Similarity</i>	70
Segmen Program 4.40 Submit Skripsi	71