

**STUDI KOMPARATIF PEMANFAATAN *VECTOR SPACE*
MODEL PADA PENERAPAN ALGORITMA NAZIEF
ADRIANI, *K-NEAREST NEIGHBOR* DAN FUNGSI JACCARD:
KASUS PROTOTIPE APLIKASI KATAGORISASI TEKS
BERBAHASA INDONESIA**

TESIS



Oleh:
Sukisno
1311601189

PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR

JAKARTA
2016

**STUDI KOMPARATIF PEMANFAATAN *VECTOR SPACE*
MODEL PADA PENERAPAN ALGORITMA NAZIEF
ADRIANI, *K-NEAREST NEIGHBOR* DAN FUNGSI JACCARD:
KASUS PROTOTIPE APLIKASI KATAGORISASI TEKS
BERBAHASA INDONESIA**

TESIS

Diajukan untuk memenuhi salah satu persyaratan
memperoleh gelar Magister Ilmu Komputer (MKOM)



Oleh:
Sukisno
1311601189

PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR

JAKARTA
2016



**PROGRAM STUDI: MAGISTER ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR**

LEMBAR PERNYATAAN

Nama Mahasiswa : Sukisno
Nomor Induk Mahasiswa : 1311601189
Program Studi : Magister Ilmu Komputer
Konsentrasi : Teknologi Sistem Informasi
Fakultas / Program : Magister Ilmu Komputer

menyatakan bahwa Tesis yang berjudul :
Studi Komparatif Pemanfaatan Vector Space Model Pada
Penerapan Algoritma Nazief Adriani, K-Nearest Neighbor
Dan Fungsi Jaccard : Kasus Prototipe Aplikasi Kategorisasi
Teks Berbahasa Indonesia

1. merupakan hasil karya tulis ilmiah sendiri dan bukan merupakan karya yang pernah diajukan untuk memperoleh gelar akademik oleh pihak lain,
2. saya ijin untuk dikelola oleh Universitas Budi Luhur sesuai dengan norma hukum dan etika yang berlaku.

Pernyataan ini saya buat dengan penuh tanggung jawab dan saya bersedia menerima konsekuensi apapun sesuai aturan yang berlaku apabila di kemudian hari pernyataan ini tidak benar.

Jakarta, 29 Januari 2016



Sukisno



**PROGRAM STUDI: MAGISTER ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR**

LEMBAR PENGESAHAN

Nama Mahasiswa : Sukisno
Nomor Induk Mahasiswa : 1311601189
Konsentrasi : Teknologi Sistem Informasi
Jenjang Studi : Strata 2 / S2
Judul Tesis : Studi Komparatif Pemanfaatan *Vector Space Model*
Pada Penerapan Algoritma Nazief Adriani, *K-Nearest Neighbor* dan Fungsi *Jaccard*: Kasus Prototipe
Aplikasi Katagorisasi Teks Berbahasa Indonesia

Jakarta, 29 Januari 2016

Tim Penguji:

Tanda Tangan:

Ketua,

(Prof. Dr. Moedjiono, M.Sc)

Anggota,

(Dr. Ir. Nazori Az, M.T)

Pembimbing,

(Dr. Krisna Adiyarta, M.Sc)

Ketua Program Studi

(Dr. Ir. Nazori Az, M.T)

ABSTRAK

Kemajuan teknologi yang sangat cepat mendorong manusia dalam memanfaatkan tumbuh kembangnya teknologi tersebut untuk melakukan pekerjaan yang dahulu dikerjakan secara manual. Terlebih lagi didorong dengan perkembangan ilmu pengetahuan yang sangat cepat. Sebagai contoh, dengan adanya teknologi komputer segala kegiatan dapat dilakukan dengan cepat dan meminimalkan resiko kesalahan. Teknik yang digunakan untuk memecahkan masalah adalah dengan menggunakan teknik *text mining* untuk pengkategorian dokumen penulisan ilmiah. Sedangkan untuk mencari nilai similaritas suatu dokumen dengan dokumen lainnya menggunakan kata kunci yang didapat dari hasil pengkategorian dokumen dan algoritma yang digunakan adalah algoritma TF-IDF (*Term Frequency – Inversed Document Frequency*), WIDF (*Weighted Inverse Document Frequency*) dan algoritma *Vector Space Model*. Selain itu salah satu metode kategorisasi teks yang lainnya adalah algoritma *K-Nearest Neighbour* (KNN). Untuk pengujian sistem adalah dengan *Recall & Precision*. Berdasarkan hal tersebut, penulis akan melakukan penelitian pengkategorisasian dokumen teks berformat *Portable Document Format* (PDF). Tujuan dari penelitian ini adalah untuk implementasi sistem klasifikasi dokumen teks berbahasa Indonesia dengan banyak kategori sehingga dapat mempermudah dalam pencarian dokumen teks berbahasa Indonesia dan mengetahui tingkat akurasi hasil klasifikasi dengan metode TF-IDF, WIDF dan *Vector Space Model* dalam mengklasifikasikan dokumen teks berbahasa Indonesia. Dengan menggunakan data set sebagai dokumen *learning* sebanyak 12 *data set* dengan 3 kategori menghasilkan nilai *precision* dan *recall* rata-rata sebesar 87.50% dengan pembobotan TF-IDF dan WIDF sebesar 83.33%. Dengan ini diambil kesimpulan bahwa TF-IDF menghasilkan *precision & recall* tertinggi sebesar 87.50% dibandingkan dengan WIDF yaitu 83.33% menggunakan similaritas fungsi *jaccard*. Hasil terbaik adalah pembobotan kata dengan TF-IDF.

Kata Kunci : Pengkategorian Dokumen, Similaritas, *Text Mining*, TF-IDF, WIDF, *Vector Space Model*, *K-Nearest Neighbor*, Fungsi *Jaccard*

ABSTRACT

Rapid technological advances that encourage people to take advantage of the growth of the technology to do the work formerly done manually. Moreover, driven by the development of knowledge very quickly. For example, the presence of computer technology all activities can be done quickly and minimize the risk of errors. The technique used to solve the problem is to use text mining techniques for categorizing documents scientific writing. To find the value of similarity of a document with other documents using keywords derived from the results of the categorization of documents and the algorithm used is algorithm TF-IDF (Term Frequency - Inversed Document Frequency), WIDF (Weighted Inverse Document Frequency) and algorithms Vector Space Model. In addition one of the other methods of text categorization algorithm is the K-Nearest Neighbor (KNN). For the testing of the system is to Recall and Precision. Based on this, the author will conduct research documents formatted text categorization Portable Document Format (PDF). The aim of this study was to implementation of the classification system in Indonesian language text documents with a lot of categories so as to facilitate the search for Indonesian language text documents and determine the level of accuracy of the classification with TF-IDF method, WIDF and Vector Space Model in classifying the Indonesian language text documents. By using the data set as a document of learning as much as 12 data sets with 3 categories produce precision and recall values average of 87.50% with a weighted TF-IDF and WIDF amounted to 83.33%. Hereby conclude that the TF-IDF produce the highest precision and recall of 87.50% compared with 83.33% WIDF ie using Jaccard similarity function. The best results are weighted with TF-IDF said.

Keywords: Document Categorization, Similarity, Text Mining, TF-IDF, WIDF, Vector Space Model, K-Nearest Neighbor, Jaccard Function

KATA PENGANTAR

Dengan mengucapkan Puji syukur ke hadirat Allah SWT yang telah mencurahkan berkah dan pertolongan-Nya penulis dapat menyelesaikan penelitian ini. Dan shalawat beriring salam semoga senantiasa tetap tercurah kepada junjungan Nabi Besar Muhammad SAW yang sangat menyayangi umatnya, sehingga penulis dapat menyelesaikan naskah penelitian ini yang berjudul **“Studi Komparatif Pemanfaatan Vector Space Model Pada Penerapan Algoritma Nazief Adriani, K-Nearest Neighbor dan Fungsi Jaccard: Kasus Prototipe Aplikasi Katagorisasi Teks Berbahasa Indonesia”**

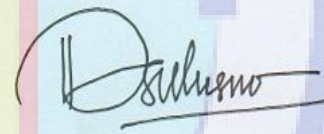
Naskah penelitian ini merupakan salah satu persyaratan yang harus dipenuhi di Magister Ilmu Komputer Universitas Budi Luhur Jakarta. Dengan selesainya naskah penelitian ini tidak terlepas dari bantuan banyak pihak yang telah memberikan masukan – masukan kepada penulis. Untuk itu penulis mengucapkan terimakasih kepada :

1. Bapak Prof. Dr. Moedjiono, M.Sc, selaku Direktur Program Pascasarjana sekaligus selaku Dosen yang telah banyak memberikan pengarahan dan saran mengenai tata cara penulisan penelitian ini.
2. Bapak Dr. Ir. Nazori Az, M.T, selaku Kaprodi Magister Komputer yang memberikan kesempatan penulis menyusun penelitian ini.
3. Bapak Dr. Krisna Adiyarta M.Sc, selaku Dosen pembimbing penelitian yang telah membimbing dan memotivasi penulis dalam mengerjakan penelitian dengan penuh kesabaran.
4. Ibu Sulichah selaku orang tua penulis tercinta yang terus memberikan motivasi didalam penyusunan penelitian ini.
5. Aris Ardika selaku Adik penulis tercinta yang terus memberikan motivasi didalam penyusunan penelitian ini.
6. Teman – teman Mahasiswa Magister Ilmu Komputer Universitas Budi Luhur kelompok XC pada Semester 1, kelompok XL pada Semester 2 dan kelompok MI (Teknologi Sistem Informasi) pada Semester 3, terima kasih atas kebersamaan, kerja keras dan dukungan serta motivasinya.

7. Kelompok belajar Taufik Hidayat, Diki Susandi, Bambang Priambodo & Hasanusi selaku teman teman satu *team* dalam persiapan penyusunan penelitian ini.
8. Bapak dan Ibu Dosen pengampu mata kuliah di Program Studi Magister Ilmu Komputer Universitas Budi Luhur yang telah dengan sabar memberikan ilmu pengetahuan, pencerahan dan bimbingan dalam belajar.
9. Semua pihak yang secara langsung maupun tidak langsung telah membantu penulis dalam menyelesaikan penelitian yang tidak disebutkan satu persatu.

Penulis menyadari masih banyak kekurangan dari penelitian ini, oleh karena itu penulis mengharapkan kritik dan saran yang bersifat membangun demi kesempurnaan penelitian nantinya. Semoga penelitian ini masih dapat memberikan manfaat dari keterbatasannya. *Aamiin.*

Jakarta, 29 Januari 2016



Sukisno

LEMBAR PERSEMBAHAN

Puji syukur kepada Allah SWT atas segala rahmat dan hidayahnya yang telah memberikan kekuatan, kesehatan dan kesabaran untuk ku dalam mengerjakan naskah akhir tesis ini.

Aku persembahkan cinta dan sayangku kepada Orang tua ku, Bapak Sarpiyo (alm) & Ibu Sulichah yang telah menjadi motivasi dan inspirasi dan tiada henti memberikan dukungan do'anya buatku.

Terimakasih yang tak terhingga buat dosen-dosen ku, terutama pembimbingku Bapak Krisna Adiyarta yang tak pernah lelah dan sabar memberikan bimbingan dan arahan kepada ku

Terimakasihku juga ku persembahkan kepada sayangku Selly Sriwulan yang senantiasa menjadi penyemangat dan menemani disetiap hariku.

Aku belajar, aku tegar, dan aku bersabar hingga aku berhasil. Terimakasih untuk Semua ^_^

DAFTAR ISI

	Halaman
LEMBAR JUDUL	i
LEMBAR PERNYATAAN	ii
LEMBAR PENGESAHAN TESIS	iii
ABSTRAK	iv
ABSTRACT.....	v
KATA PENGANTAR	vi
LEMBAR PERSEMBAHAN	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Masalah Penelitian	4
1.2.1 Identifikasi Masalah	4
1.2.2 Pembatasan Masalah	4
1.2.3 Rumusan Masalah	5
1.3 Tujuan dan Manfaat Penelitian	5
1.3.1 Tujuan Penelitian	5
1.3.2 Manfaat Penelitian	6
1.4 Tata-Urut Penulisan	6
1.5 Daftar Pengertian	7
BAB II LANDASAN TEORI DAN KERANGKA KONSEP	10
2.1 Tinjauan Pustaka	10
2.1.1 Data Mining	10
2.1.2 Information Retrieval (IR)	14
2.1.2.1 Tujuan dan Fungsi IR	17
2.1.2.2 Arsitektur IR	19
2.1.3 Klasifikasi Teks	20
2.1.3.1 Pengertian Klasifikasi Teks	20
2.1.3.2 Metode Pengklasifikasian Teks	22
2.1.4 Vector Space Model	24
2.1.5 Model Representasi Dokumen	25
2.1.6 Algoritma K-Nearest Neighbor	28
2.1.7 Stemming dan Algoritma Nazief Adriani	29
2.1.7.1 Stemming	29
2.1.7.2 Stemming Bahasa Indonesia Nazief Adriani .	32
2.1.8 Morfologi Bahasa Indonesia	39
2.1.8.1 Morfologi	39
2.1.8.2 Proses Morfologi	39
2.1.9 Tokenisasi	42
2.1.10 Filtering	43
2.1.11 Stopword	43
2.1.12 Pembobotan Kata (Term Weighting).....	44

	2.1.12.1 TF-IDF	44
	2.1.12.2 WIDF	45
	2.1.13 Kesamaan (<i>Similarity</i>)	46
	2.2 Tinjauan Studi	47
	2.3 Tinjauan Obyek Penelitian	52
	2.4 Kerangka Konsep/Pola Pikir Pemecahan Masalah	53
	2.5 Hipotesis	55
BAB III	METODOLOGI DAN RANCANGAN/DESAIN PENELITIAN	56
	3.1 Metode Penelitian	56
	3.2 Metode Pemilihan Sampel	58
	3.3 Instrumentasi	59
	3.4 Teknik Analisis, <i>Prototype Model</i> dan Pengujian.....	59
	3.4.1 Teknik Analisis	59
	3.4.2 <i>Prototype Model</i>	60
	3.4.3 Teknik Pengujian	62
	3.4.3.1 <i>Precision</i>	65
	3.4.3.2 <i>Recall</i>	65
	3.5 Langkah-langkah Penelitian	65
	3.6 Jadwal Penelitian	68
BAB IV	PEMBAHASAN HASIL PENELITIAN	71
	4.1 Data Set	71
	4.2 <i>Prototype Model</i>	72
	4.2.1. Tampilan <i>Home</i>	72
	4.2.2. Tampilan Halaman Menu <i>Login</i>	72
	4.2.3. Halaman <i>Setting Akun</i>	73
	4.2.4. Halama <i>Input Learning Document</i>	73
	4.2.5. Tampilan <i>Menu Admin</i>	74
	4.2.6. Tampilan Halaman Klasifikasi Dokumen	74
	4.2.7. Tampilan Input Kategori	75
	4.3 <i>Learning Document</i> dengan TF-IDF dan WIDF	76
	4.4 Hasil Pengujian	88
	4.4.1 Hasil <i>Similarity</i> Dokumen dengan TF-IDF.....	88
	4.4.1.1 Jika D1 Sebagai Q1.....	88
	4.4.1.2 Jika D2 Sebagai Q2	89
	4.4.1.3 Jika D3 Sebagai Q3.....	89
	4.4.1.4 Jika D4 Sebagai Q4	89
	4.4.1.5 Jika D5 Sebagai Q5	90
	4.4.1.6 Jika D6 Sebagai Q6	90
	4.4.1.7 Jika D7 Sebagai Q7	90
	4.4.1.8 Jika D8 Sebagai Q8	91
	4.4.1.9 Jika D9 Sebagai Q9	91
	4.4.1.10 Jika D9 Sebagai Q10	91
	4.4.1.11 Jika D9 Sebagai Q11	92
	4.4.1.12 Jika D9 Sebagai Q12	92
	4.4.2 Hasil <i>Similarity</i> Dokumen dengan WIDF	92
	4.4.2.1 Jika D1 Sebagai Q1	92

4.4.2.2	Jika D2 Sebagai Q2	93
4.4.2.3	Jika D3 Sebagai Q3	93
4.4.2.4	Jika D4 Sebagai Q4	93
4.4.2.5	Jika D5 Sebagai Q5	94
4.4.2.6	Jika D6 Sebagai Q6	94
4.4.2.7	Jika D7 Sebagai Q7	95
4.4.2.8	Jika D8 Sebagai Q8	95
4.4.2.9	Jika D9 Sebagai Q9	95
4.4.2.10	Jika D9 Sebagai Q10	96
4.4.2.11	Jika D9 Sebagai Q11	96
4.4.2.12	Jika D9 Sebagai Q12	96
4.5	Hasil Perangkingan KNN	97
4.5.1	Hasil Perangkingan KNN Pembobotan TF-IDF	97
4.5.1.1	Hasil Perangkingan TF-IDF Q1	97
4.5.1.2	Hasil Perangkingan TF-IDF Q2	98
4.5.1.3	Hasil Perangkingan TF-IDF Q3	99
4.5.1.4	Hasil Perangkingan TF-IDF Q4	99
4.5.1.5	Hasil Perangkingan TF-IDF Q5	100
4.5.1.6	Hasil Perangkingan TF-IDF Q6	101
4.5.1.7	Hasil Perangkingan TF-IDF Q7	101
4.5.1.8	Hasil Perangkingan TF-IDF Q8	102
4.5.1.9	Hasil Perangkingan TF-IDF Q9	103
4.5.1.10	Hasil Perangkingan TF-IDF Q10	103
4.5.1.11	Hasil Perangkingan TF-IDF Q11	104
4.5.1.12	Hasil Perangkingan TF-IDF Q12	105
4.5.2	Hasil Perangkingan KNN Pembobotan WIDF	105
4.5.2.1	Hasil Perangkingan WIDF Q1	106
4.5.2.2	Hasil Perangkingan WIDF Q2	106
4.5.2.3	Hasil Perangkingan WIDF Q3	107
4.5.2.4	Hasil Perangkingan WIDF Q4	108
4.5.2.5	Hasil Perangkingan WIDF Q5	108
4.5.2.6	Hasil Perangkingan WIDF Q6	109
4.5.2.7	Hasil Perangkingan WIDF Q7	110
4.5.2.8	Hasil Perangkingan WIDF Q8	110
4.5.2.9	Hasil Perangkingan WIDF Q9	111
4.5.2.10	Hasil Perangkingan WIDF Q10	112
4.5.2.11	Hasil Perangkingan WIDF Q11	112
4.5.2.12	Hasil Perangkingan WIDF Q12	113
4.6	Hasil <i>Precision</i> dan <i>Recall</i>	114
4.7	Hasil dan Pembahasan	114
4.8	Implikasi Penelitian	115
4.8.1	Aspek Sistem	115
4.8.2	Aspek Manajerial	117
4.8.3	Aspek Penelitian Lanjut	118
4.9	Rencana Implementasi	119

BAB V	PENUTUP	122
5.1	Kesimpulan	122
5.2	Saran	122
DAFTAR PUSTAKA		125
LAMPIRAN-LAMPIRAN		130
RIWAYAT HIDUP SINGKAT		198



DAFTAR GAMBAR

Gambar	Halaman
II-1 Langkah-langkah Dalam Proses <i>Data Mining</i>	10
II-2 Bidang Ilmu <i>Data Mining</i>	14
II-3 Illustrasi <i>Information Retrieval System</i>	15
II-4 Skema <i>Information Retrieval System</i>	16
II-5 Proses <i>Information Retrieval System</i>	19
II-6 Pembelajaran Mesin Dengan Klasifikasi Terpandu	21
II-7 Diagram Proses Klasifikasi Teks	23
II-8 Representasi Dokumen dan Vektor	26
II-9 Representasi Matriks kata-dokumen.....	27
II-10 <i>Flowchart Stemming</i> Bahasa Indonesia	31
II-11 Kerangka Pikir	54
III-1 <i>Recall & Precision</i> Pada Contoh Hasil Temu Kembali.....	62
III-2 <i>Trade-off</i> antara <i>Precision & Recall</i>	64
III-3 Langkah-langkah Penelitian	66
IV-1 Tampilan Halaman <i>Home</i>	72
IV-2 Tampilan Menu <i>Login</i>	72
IV-3 Tampilan <i>Setting Akun Admin</i>	73
IV-4 Tampilan <i>Input Learning Document</i>	73
IV-5 Tampilan Halaman Admin	74
IV-6 Tampilan Halaman Klasifikasi Document	74
IV-7 Tampilan Halaman <i>Input Kategori</i>	75
IV-8 <i>Document Classification Form TF-IDF Q1</i>	98
IV-9 <i>Document Classification Form TF-IDF Q2</i>	98
IV-10 <i>Document Classification Form TF-IDF Q3</i>	99
IV-11 <i>Document Classification Form TF-IDF Q4</i>	100
IV-12 <i>Document Classification Form TF-IDF Q5</i>	100
IV-13 <i>Document Classification Form TF-IDF Q6</i>	101
IV-14 <i>Document Classification Form TF-IDF Q7</i>	102
IV-15 <i>Document Classification Form TF-IDF Q8</i>	102
IV-16 <i>Document Classification Form TF-IDF Q9</i>	103
IV-17 <i>Document Classification Form TF-IDF Q10</i>	104
IV-18 <i>Document Classification Form TF-IDF Q11</i>	104
IV-19 <i>Document Classification Form TF-IDF Q12</i>	105
IV-20 <i>Document Classification Form WIDF Q1</i>	106
IV-21 <i>Document Classification Form WIDF Q2</i>	107
IV-22 <i>Document Classification Form WIDF Q3</i>	107
IV-23 <i>Document Classification Form WIDF Q4</i>	108
IV-24 <i>Document Classification Form WIDF Q5</i>	109
IV-25 <i>Document Classification Form WIDF Q6</i>	109
IV-26 <i>Document Classification Form WIDF Q7</i>	110
IV-27 <i>Document Classification Form WIDF Q8</i>	111
IV-28 <i>Document Classification Form WIDF Q9</i>	111
IV-29 <i>Document Classification Form WIDF Q10</i>	112

IV-30	<i>Document Classification Form WIDF Q11</i>	113
IV-31	<i>Document Classification Form WIDF Q12</i>	113



DAFTAR TABEL

Tabel	Halaman
II-1 Kombinasi Awalan Akhiran Yang Tidak Diijinkan	33
II-2 Cara Menentukan Tipe Awalan Untuk Awalan “te-“	34
II-3 Jenis Awalan Berdasarkan Tipe Awalannya	34
II-4 <i>Example of Collection</i>	45
II-5 Ringkasan Tinjauan Studi	49
III-1 Jadwal Penelitian	68
IV-1 <i>Learning Document TF-IDF</i>	76
IV-2 <i>Learning Document WIDF</i>	82
IV-3 <i>Similarity</i> Pembobotan TF-IDF Q1	88
IV-4 <i>Similarity</i> Pembobotan TF-IDF Q2	89
IV-5 <i>Similarity</i> Pembobotan TF-IDF Q3	89
IV-6 <i>Similarity</i> Pembobotan TF-IDF Q4	89
IV-7 <i>Similarity</i> Pembobotan TF-IDF Q5	90
IV-8 <i>Similarity</i> Pembobotan TF-IDF Q6	90
IV-9 <i>Similarity</i> Pembobotan TF-IDF Q7	90
IV-10 <i>Similarity</i> Pembobotan TF-IDF Q8	91
IV-11 <i>Similarity</i> Pembobotan TF-IDF Q9	91
IV-12 <i>Similarity</i> Pembobotan TF-IDF Q10	91
IV-13 <i>Similarity</i> Pembobotan TF-IDF Q11	92
IV-14 <i>Similarity</i> Pembobotan TF-IDF Q12	92
IV-15 <i>Similarity</i> Pembobotan WIDF Q1	92
IV-16 <i>Similarity</i> Pembobotan WIDF Q2	93
IV-17 <i>Similarity</i> Pembobotan WIDF Q3	93
IV-18 <i>Similarity</i> Pembobotan WIDF Q4	93
IV-19 <i>Similarity</i> Pembobotan WIDF Q5	94
IV-20 <i>Similarity</i> Pembobotan WIDF Q6	94
IV-21 <i>Similarity</i> Pembobotan WIDF Q7	95
IV-22 <i>Similarity</i> Pembobotan WIDF Q8	95
IV-23 <i>Similarity</i> Pembobotan WIDF Q9	95
IV-24 <i>Similarity</i> Pembobotan WIDF Q10	96
IV-25 <i>Similarity</i> Pembobotan WIDF Q11	96
IV-26 <i>Similarity</i> Pembobotan WIDF Q12	96
IV-27 Hasil perangkian TF.IDF Q1 dengan <i>data set</i>	97
IV-28 Hasil perangkian TF.IDF Q2 dengan <i>data set</i>	98
IV-29 Hasil perangkian TF.IDF Q3 dengan <i>data set</i>	99
IV-30 Hasil perangkian TF.IDF Q4 dengan <i>data set</i>	99
IV-31 Hasil perangkian TF.IDF Q5 dengan <i>data set</i>	100
IV-32 Hasil perangkian TF.IDF Q6 dengan <i>data set</i>	101
IV-33 Hasil perangkian TF.IDF Q7 dengan <i>data set</i>	101
IV-34 Hasil perangkian TF.IDF Q8 dengan <i>data set</i>	102
IV-35 Hasil perangkian TF.IDF Q9 dengan <i>data set</i>	103
IV-36 Hasil perangkian TF.IDF Q10 dengan <i>data set</i>	103
IV-37 Hasil perangkian TF.IDF Q11 dengan <i>data set</i>	104

IV-38	Hasil perangkingan TF.IDF Q12 dengan <i>data set</i>	105
IV-39	Hasil perangkingan WIDF Q1 dengan <i>data set</i>	106
IV-40	Hasil perangkingan WIDF Q2 dengan <i>data set</i>	106
IV-41	Hasil perangkingan WIDF Q3 dengan <i>data set</i>	107
IV-42	Hasil perangkingan WIDF Q4 dengan <i>data set</i>	108
IV-43	Hasil perangkingan WIDF Q5 dengan <i>data set</i>	108
IV-44	Hasil perangkingan WIDF Q6 dengan <i>data set</i>	109
IV-45	Hasil perangkingan WIDF Q7 dengan <i>data set</i>	110
IV-46	Hasil perangkingan WIDF Q8 dengan <i>data set</i>	110
IV-47	Hasil perangkingan WIDF Q9 dengan <i>data set</i>	111
IV-48	Hasil perangkingan WIDF Q10 dengan <i>data set</i>	112
IV-49	Hasil perangkingan WIDF Q11 dengan <i>data set</i>	112
IV-50	Hasil perangkingan WIDF Q12 dengan <i>data set</i>	113
IV-51	Hasil pengujian <i>precision</i> dan <i>recall</i>	114
IV-52	Hasil perhitungan rata-rata <i>precision</i> dan <i>recall</i>	114
IV-53	Standar Minimal Spesifikasi	116
IV-54	Rencana Implementasi Sistem	120

DAFTAR LAMPIRAN

Lampiran	Halaman
1 Perhitungan TF-IDF	133
2 Perhitungan WIDF	149
3 Dokumen <i>Online</i> D1	173
4 Dokumen <i>Online</i> D2	174
5 Dokumen <i>Online</i> D3	175
6 Dokumen <i>Online</i> D4	176
7 Dokumen <i>Online</i> D5	177
8 Dokumen <i>Online</i> D6	179
9 Dokumen <i>Online</i> D7	180
10 Dokumen <i>Online</i> D8	182
11 Dokumen <i>Online</i> D9	184
12 Dokumen <i>Online</i> D10	185
13 Dokumen <i>Online</i> D11	186
14 Dokumen <i>Online</i> D12	187
15 <i>Listing Program</i>	188
16 Kartu Bimbingan	199
17 Daftar Riwayat Hidup	200

DAFTAR PUSTAKA

- [Agusta 2009] Agusta, L. Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief dan Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia. Universitas Kristen Satya Wacana.Bali.2009.
- [Andreas 2012] Andreas Handojo, Charistian Widjaja dan Adi Wibowo. Pembuatan Aplikasi Pencarian Dokumen Berbasis Generalized Vector Space Model dan Semantic Relatedness. Surabaya: Universitas Kristen Petra.2012.
- [Ari 2015] Andika, Ari. Perancangan Aplikasi Pengukuran Similaritas Pada Dokumen Dengan Metode Semantic. STMIK Budi Darma: Medan, 2015
- [Attardi 2004] Attardi, G., Text Categorization, Roma: Pisa University Pr.2004
- [Baeza 1999] Baeza-Yates & Ribeiro-Neto, Modern Information Retrieval, Harlow, Addison-Wesley. 1999.
- [Cahyo 2012] Cahyo Darujati, Agustinus BG.Pemanfaatan Teknik Supervised untuk Klasifikasi Teks Bahasa Indonesia. Universitas Narotama, 2012.
- [Christopher 2009] Christopher, D. Manning, Prabhakar Raghavan, Hinrich Schütze. 2009]. An Introduction to Information Retrieval. Cambrigde: Cambridge UP, 2009.
- [Cios 2007] Cios, Krzysztof J. Etc. Data Mining A Knowledge Discovery Approach,Springer.(http://www.4shared.com/document/FyVdn5pm/Data_Mining_A_Knowledge_Discov.html , 2007.
- [Diah 2010] Diah Pudi Langgeni, ZK.Abdurahman Baizal dan Yanuar Firdaus A.W. Clustering Artikel Berita Berbahasa Indonesia Menggunakan Unsupervised Feature Selection. Institut Teknologi Telkom : Bandung.2010.
- [Dik 1997] Dik, L. LEE., Huei, C. & Kent E.S. Document Ranking and the Vector Space Model, 67-75.1997.
- [Fatmawati 2013] Fatmawati, Studi Komparatif Kecepatan Temu Kembali Informasi di Depo Arsip Koran Suara Merdeka Antara Sistem Simpan Manual dengan Foto Repro. Diponegoro University, 2013.