

DAFTAR ISI

KATA PENGANTAR	i
ABSTRAK	iii
DAFTAR ISI	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL	xii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan	4
1.5 Manfaat	4
1.6 Sistematika Penulisan	4
BAB II KAJIAN PUSTAKA	7
2.1 Klasifikasi Dokumen	9
2.2 Berita	10
2.3 Data Mining	10
2.4 <i>Text Mining</i>	11
2.5 Tahap <i>Preprocessing Text</i>	11
2.5.1 <i>Case Folding</i>	11
2.5.2 <i>Tokenizing</i>	12
2.5.3 <i>Filtering</i>	12
2.5.4 <i>Stemming</i>	12
2.5.5 Ontologi	20
2.5.6 WordNet	21
2.5.7 RiTa.WordNet	22
2.5.8 <i>Dictionary Construction</i>	22
2.5.9 <i>Feature Selection</i>	23

2.5.10 <i>Feature Weight</i> (Pembobotan)	24
2.6 <i>Classifier Construction</i>	25
2.6.1 Model Ruang Vektor (<i>Vector Space Model</i>)	26
2.6.2 Algoritma <i>Bubble Sort</i>	27
2.6.3 Algoritma <i>K-Nearest Neighbor</i> (KNN)	28
2.7 Evaluasi	28

BAB III METODOLOGI DAN PERANCANGAN 31

3.1 Deskripsi Data	32
3.2 Perancangan Sistem	33
3.2.1. Diagram Blok Sistem	33
3.2.2. Diagram UML (<i>Unified Model Language</i>)	35
3.3 Perancangan Algoritma	37
3.3.1 <i>Preprocessing</i> Dokumen Latih	39
1. <i>Case folding</i>	40
2. <i>Tokenizing</i>	40
3. <i>Filtering</i>	41
4. <i>Stemming</i>	42
5. <i>Dictionary Construction</i>	43
6. <i>Feature Selection</i>	44
7. <i>Ontology Extraction</i>	47
8. Pembobotan Dokumen Latih	49
3.3.2 <i>Preprocessing</i> Dokumen Uji	50
3.3.3 <i>Classifier Construction</i>	53
3.3.4 <i>Document Categorization</i>	56
3.4. Perhitungan Manual	58
3.5. Rancangan Antar Muka	80
3.6. Rancangan Uji Coba	83

BAB IV IMPLEMENTASI 85

4.1 Lingkungan Implementasi	85
4.1.1 Lingkungan Implementasi Perangkat Keras	85



4.1.2 Lingkungan Implementasi Perangkat Lunak	85
4.2 Implementasi Program	86
4.2.1 Kelas dan Fungsi	86
4.2.2 Tahapan Pemrosesan	91
4.3. Implementasi Antar Muka	109
BAB V ANALISA HASIL DAN PEMBAHASAN.....	119
5.1 Skenario Pengujian.....	119
5.1.1 Skenario Uji Jumlah “k” dan Nilai <i>Document Frequency Threshold</i>	119
5.1.2 Skenario Perbandingan Jumlah Dokumen Uji dengan Dokumen Latih	120
5.2 Hasil Pengujian	121
5.2.1 Hasil Pengujian Jumlah “k” dan Nilai <i>Document Frequency Threshold</i>	121
5.2.1.1 Pengujian dengan Menggunakan Data Uji yang Diambil dari Data Latih	121
5.2.1.2 Pengujian dengan Menggunakan Data Uji yang Berbeda dari Data Latih	124
5.2.2 Hasil Pengujian Perbandingan Jumlah Dokumen Uji dengan Dokumen Latih	135
5.3. Analisa Hasil Pengujian.....	143
BAB VI PENUTUP.....	147
6.1 Kesimpulan	147
6.2 Saran.....	148
DAFTAR PUSTAKA	149
LAMPIRAN	151
Lampiran 1 Daftar <i>Stopwords</i>	147
Lampiran 2 <i>TFIDF Term</i> Dokumen Latih	148

