

Daftar Isi

LEMBAR PERNYATAAN	ii
LEMBAR PENGESAHAN	iii
ABSTRAK	iv
ABSTRACT	v
LEMBAR PERSEMBAHAN	vi
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR ISTILAH	xiv
1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan masalah	1
1.3 Tujuan	2
1.4 Metodologi penyelesaian masalah	2
2. LANDASAN TEORI	3
2.1 <i>Text Mining</i>	3
2.2 Ekstraksi Kata Kunci	3
2.3 Frase	3
2.4 Preprocessing Data	4
2.4.1 <i>Case Folding</i>	4
2.4.2 <i>Cleaning</i>	4
2.4.3 <i>Tokenisasi</i>	4
2.4.4 <i>Stopwords Removal</i>	4
2.4.5 <i>POS Tagging</i>	5
2.5 <i>Part-of-Speech Tagging (POS Tagging)</i>	5
2.6 <i>Hidden Markov Model (HMM)</i>	6
2.6.1 Elemen-Elemen dalam HMM	7
2.6.2 Permasalahan dasar HMM	7
2.6.3 Solusi Permasalahan-Permasalahan Dasar HMM	7
2.6.4 HMM – Hedge	8
2.6.5 Menghitung Probabilitas Emisi (Model Translasi)	9
2.6.6 Menghitung Probabilitas Transisi (Model Bahasa)	9
2.6.7 <i>Jelinec – Mercer Smoothing</i>	10

2.6.8	<i>ViterbiDecoding</i>	10
2.7	Parameter Pengukuran	11
2.7.1	Precision, Recall, F-Measure	11
3.	DESKRIPSI SISTEM	13
3.1	Tahapan Preprocessing	14
3.1.1	<i>Case Folding</i>	16
3.1.2	<i>Tokenisasi</i>	16
3.1.3	<i>Cleaning dan StopwordRemoval</i>	16
3.1.4	<i>POS Tagging</i>	17
3.1.5	Penandaan Frase.....	17
3.1.6	<i>MappingKeywords</i>	18
3.2	Tahapan Training Hidden Markov Model (HMM).....	19
3.3	Tahapan Testing Hidden Markov Model (HMM)	20
3.4	Analisis dan Kebutuhan Sistem	22
3.4.1	Kebutuhan Fungsionalitas	22
3.4.2	Spesifikasi Perangkat Keras	22
3.4.3	Spesifikasi Perangkat Lunak	22
3.5	DFD.....	22
3.6	Struktur Database.....	25
3.7	Kamus Data.....	25
3.8	<i>Process Specification</i>	26
4.	IMPLEMENTASI DAN HASIL PENGUJIAN	30
4.1	Pengujian Sistem.....	30
4.1.1	<i>Dataset</i>	30
4.1.2	Tujuan Pengujian	30
4.2	Skenario Pengujian	30
4.2.1	Mengetahui pengaruh parameter α terhadap akurasi.....	31
4.2.2	Mengetahui komposisi perbandingan data <i>training</i> dan data <i>testing</i>	31
4.2.3	Mengetahui pengaruh penggunaan frase dan tanpa penggunaan frase	31
4.2.4	Mengetahui pengaruh homogenitas/heterogenitas data yang digunakan	31
4.2.5	Melihat tag yang paling dominan menjadi kata kunci berdasarkan kata kunci yang diekstrak oleh sistem.....	32
4.3	Hasil Pengujian	32
4.3.1	Hasil Skenario Pengaruh Nilai <i>Alpha</i>	32
4.3.2	Hasil Skenario Perbandingan Data <i>Train</i> dan Data <i>Test</i>	33
4.3.3	Hasil Skenario Penggunaan Frase dan Tanpa Frase.....	33
4.3.4	Hasil Skenario Pengaruh Homogenitas/Heterogenitas Data	34

4.3.5	Hasil Skenario Mengetahui Tag / Jenis Kata dari <i>Keywords</i> yang dihasilkan	34
4.4	Analisa Pengujian	35
4.4.1	Analisa Skenario Pengaruh Nilai <i>Alpha</i>	35
4.4.2	Analisa Skenario Perbandingan Data <i>Train</i> dan Data <i>Test</i>	36
4.4.3	Analisa Skenario Perbandingan Penggunaan Frase dan Tanpa Penggunaan Frase	37
4.4.4	Analisa Skenario Pengaruh Homogenitas/Heterogenitas Data	38
4.4.5	Analisa Skenario Tag / Jenis Kata yang menjadi Kata Kunci.....	39
4.4.6	Analisis Keseluruhan Skenario	39
5.	KESIMPULAN.....	40
5.1	Kesimpulan	40
5.2	Saran	40
6.	DAFTAR PUSTAKA.....	41
	Lampiran A : Contoh <i>Dataset</i> Abstrak <i>Paper</i> Tiap Kelas.....	43
	Lampiran B : Daftar <i>StopwordList</i>	54