

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	iii
DAFTAR ISI.....	vi
DAFTAR TABEL.....	vii
DAFTAR GAMBAR	viii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Penelitian	1
1.2 Rumusan Masalah	5
1.3 Tujuan.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	6
1.6 Sistematika Penulisan.....	6
BAB II KAJIAN PUSTAKA	8
2.1. Tata Bahasa Baku Bahasa Indonesia.....	8
2.2. Kalimat	8
2.1.1. Unsur dan Pola Kalimat dalam Sintaksis.....	10
2.1.2. Kalimat Rancu.....	14
2.1.3. Kalimat Rancu pada Sisi Bentuk atau Struktur.....	14
2.1.4. Kalimat Rancu pada Sisi Makna.....	19
2.2. <i>Natural Language Processing (NLP)</i>	20
2.2.1. Part of Speech (PoS) Tagging.....	21
2.3. <i>Grammatical Error Detection</i>	25
2.4. Klasifikasi Teks	29
2.4.1. Feature Extraction	30
2.4.2. Classifier	30
2.4.3. Evaluasi Classifier.....	32
2.5. <i>Sequential Pattern Mining</i>	33
2.5.1. Minimum Support	35
2.5.2. Class Sequential Rules	37
2.5.3. Labeled Sequence Patterns (LSP)	38
2.5.4. Pembuatan Sequence Database	39
2.5.5. Penambangan (Mining) LSP	42

MIRA NURHAYATI, 2018

DETEKSI KALIMAT BAHASA INDONESIA BERSTRUKTUR RANCU MENGGUNAKAN LABELED SEQUENTIAL PATTERN

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

2.5.6.	Konversi LSP untuk Klasifikasi Teks	45
BAB III METODE PENELITIAN.....		46
3.1.	Tahapan dan Prosedur Penelitian	46
3.1.1.	Tahap Pemrosesan Data	46
3.1.2.	Tahap Perancangan	48
3.1.3.	Tahap Pembangunan	50
3.1.4.	Tahap Pengujian (Komponen & Sistem)	50
3.2.	Alat dan Bahan Penelitian	52
3.2.1.	Alat Penelitian	52
3.2.2.	Bahan Penelitian.....	53
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....		54
4.1.	Data Penelitian	54
4.2.	Pengembangan Sistem.....	57
4.2.1.	POS Tagger Bahasa Indonesia	58
4.2.2.	Labeled Sequential Patterns (LSP) Mining	69
4.3.	Data Latih dan Data Uji.....	73
4.4.	Aplikasi LSP pada <i>Classifier</i>	73
4.5.	Pengujian Sistem dan Hasil Penelitian	74
4.5.1.	Pengujian Sistem	74
4.3.	Hasil Penelitian dan Analisis.....	76
4.3.1.	Mengganti Train Corpus	77
4.3.2.	Menambahkan Kosa Kata pada Train Corpus	79
4.3.3.	Menambahkan Tag untuk Kata Kerja Transitif (VBT).....	84
BAB V KESIMPULAN DAN SARAN.....		89
5.1.	Kesimpulan.....	89
5.2.	Rekomendasi dan Saran	90
DAFTAR PUSTAKA		91

DAFTAR TABEL

Tabel 2.1	Keterangan dan contoh setiap kelas kata.	11
Tabel 2.2.	Daftar imbuhan beserta masing-masing fungsi dan ketentuannya.....	16
Tabel 2.3	Confusion matrix untuk kasus klasifikasi kelas biner.....	32
Tabel 2.4.	Contoh sequence database yang dibuat dari data teks	36
Tabel 2.5.	Sequential pattern yang terekstrak pada support $\geq 25\%$	37
Tabel 2.6.	Contoh sequence database untuk menambang CSR.	38
Tabel 2.7.	Contoh sequence database untuk menambang LSP.....	39
Tabel 2.8.	Contoh sequence database yang akan di-mining PrefixSpan.....	43
Tabel 2.9.	Contoh data kalimat yang sudah diberi POS tag	45
Tabel 2.10.	Contoh LSP yang ditemukan pada data sequence kalimat rancu.	45
Tabel 2.11.	Contoh arsitektur feature pada classifier.....	45
Tabel 4.1	Daftar cek untuk pengelompokkan data kalimat.....	54
Tabel 4.2	Data kalimat yang terkumpul.....	56
Tabel 4.3	Tagset yang digunakan pada training corpus.....	59
Tabel 4.4	Parameter HMM untuk kalimat “ <i>Ibu sedang memasak sayur.</i> ”	60
Tabel 4.5	Data uji untuk komponen tagger.....	62
Tabel 4.6	Confusion matrix untuk data yang diproses oleh tagger.....	64
Tabel 4.7	Rincian <i>confusion measure</i> untuk masing-masing <i>tag</i>	65
Tabel 4.8	Nilai kinerja tagger.....	66
Tabel 4.9	Daftar perhitungan HMM pada data 508	67
Tabel 4.10	Hasil ekstraksi LSP pada data di Tabel 4.6.....	70
Tabel 4.11	Skenario pengujian sistem	75
Tabel 4.12	Hasil eksperimen (bigram dan trigram)	77
Tabel 4.13	Tagset corpus Bahasa Indonesia dari PANL10N1.....	77
Tabel 4.14	Hasil penelitian setelah mengganti train corpus	79
Tabel 4.15	Hasil pengujian setelah menambah kosa kata pada train corpus.	80
Tabel 4.16	Confusion matrix untuk skenario pengujian nomor 5.....	80
Tabel 4.17	Sampel data kalimat yang salah diprediksi oleh classifier.....	82
Tabel 4.18	Hasil pengujian setelah penambahan tag VBT pada train corpus.....	84
Tabel 4.19	Persentase kesalahan prediksi untuk setiap kategori kerancuan	85
Tabel 4.20	Data yang salah diprediksi pada eksperimen 9	86

DAFTAR GAMBAR

Gambar 2.1 Penguraian kalimat menjadi konstituen-konstituennya.....	10
Gambar 2.2. Contoh POS tagging pada dua kalimat	22
Gambar 2.3. Contoh corpus yang sudah diberi tag secara manual	24
Gambar 2.4. Konsep klasifikasi	31
Gambar 2.5. Konsep classifier	31
Gambar 2.6 Cara kerja algoritma PrefixSpan	44
Gambar 3.1. Desain penelitian	47
Gambar 4.1 Hasil analisis bentuk kesalahan struktur pada kalimat (1)	55
Gambar 4.2 Hasil analisis bentuk kesalahan struktur pada kalimat (2)	55
Gambar 4.3 Hasil analisis kesalahan struktur pada kalimat (3)	56
Gambar 4.4 Skema mekanisme kerja POS tagger.	59
Gambar 4.5 Flowchart algoritma Viterbi Backpointer	61
Gambar 4.6 Cara menghitung confusion measure untuk tag Z.....	65
Gambar 4.7 Bentuk data latih yang menjadi masukan pada model <i>classifier</i>	74
Gambar 4.8 Bentuk data uji yang menjadi masukan pada model classifier	75