

DAFTAR ISI

PERNYATAAN.....	i
DAFTAR ISI.....	ii
DAFTAR GAMBAR.....	iii
DAFTAR TABEL	iv
PRAKATA.....	v
INTISARI	vi
ABSTRACT.....	vii
BAB I PENDAHULUAN.....	1
1.1 Landasan Teori	1
1.2 Perumusan Masalah.....	2
1.3 Tujuan Penelitian.....	2
1.4 Batasan Penelitian	2
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	3
BAB II TINJAUAN PUSTAKA	5
BAB III LANDASAN TEORI	7
3.1 Pemrosesan Teks Informal	7
3.1.1 Normalization	7
3.1.2 Domain Adaptation	7
3.2 Noisy Channel Model.....	8
3.3 Language Model (n-gram)	13
3.4 <i>Library</i> langid.py.....	15
3.5 Evaluasi (Precision, Recall, Accuracy, F1 Score).....	15
3.6 <i>Library</i> NLTK	17
3.7 Algoritma Ratcliff/Obershelp.....	17
BAB IV ANALISIS DAN PERANCANGAN.....	19
4.1 Analisis Sistem	19
4.2 Analisis Kebutuhan	19
4.3 Perancangan Proses	19
4.4 Pembuatan Training Data.....	20
4.5 Data Flow Diagram	24
4.6 <i>Training</i>	26
4.6.1 Preprocessing	28
4.6.2 Pembuatan <i>Alignment</i> kata baku dan kata tidak bakunya	28
4.6.3 Pembuatan confusion matrix untuk $P(x/w)$	30
4.6.4 Penghitungan $P(w)$	30
4.7 Normalisasi.....	30
4.8 Pengujian.....	33
BAB V IMPLEMENTASI SISTEM	34
5.1 Pembuatan Training Data.....	34
5.2 Pembuatan Tampilan Utama	35
5.3 Pembuatan Fungsi <i>Preprocessing</i>	36
5.4 Inisialisasi data	37



5.5	Pembuatan Fungsi Penghitungan Alignment	38
5.6	Pembuatan Fungsi Training	40
5.7	Pembuatan Fungsi Normalisasi	40
5.8	Pembuatan <i>Testing</i>	41
BAB VI HASIL DAN PEMBAHASAN		43
6.1	Pengujian Fungsionalitas.....	43
6.2	Eksperimen.....	44
6.3	Pembahasan Normalisasi.....	45
6.4	Kemampuan Sistem Menormalisasi Kata Baru	48
6.5	Kekurangan Sistem Dalam Jumlah Training Data	48
6.6	Kekurangan Sistem Dalam Mengolah Kata Yang Tak Perlu Diubah	49
6.7	Kekurangan Sistem Dalam Menentukan Konteks Kata	50
6.8	Penambahan Kata dari Kamus untuk Training Data	51
BAB VII KESIMPULAN DAN SARAN.....		52
7.1	Kesimpulan.....	52
7.2	Saran.....	52
Daftar Pustaka		53
LAMPIRAN.....		54

DAFTAR GAMBAR

Gambar 3.1 Noisy Channel Model (Jurafksy dan Martin, 2009)	8
Gambar 3.2 Confusion matrix (Jurafksy dan Martin, 2009).....	11
Gambar 3.3 Perhitungan $P(x/w)$ (Kernighan et al, 1990)	11
Gambar 3.4 Probabilitas dari <i>bigram</i> (Jurafksy dan Martin, 2009).....	13
Gambar 3.5 Probabilitas dari kata dengan menggunakan <i>bigram</i> (Jurafksy dan Martin, 2009)	13
Gambar 3.6 Contoh perbandingan kata (Ilyankou, 2014).....	17
Gambar 3.7 <i>Substring</i> terpanjang (Ilyankou, 2014).....	17
Gambar 3.8 <i>Substring</i> terpanjang di sebelah kiri <i>anchor</i> (Ilyankou, 2014).....	18
Gambar 4.1 <i>Flowchart</i> penelitian	20
Gambar 4.2 Tokenisasi Sistem	21
Gambar 4.3 Contoh Data	21
Gambar 4.4 DFD Level 0 Sistem Normalisasi	24
Gambar 4.5 DFD Level 1 Proses Training	25
Gambar 4.6 DFD Level 1 Proses Normalisasi.....	26
Gambar 4.7 Urutan Training.....	27
Gambar 4.8 Proses <i>Preprocessing</i>	28
Gambar 4.9 Contoh Alignment Kata	28
Gambar 4.10 Proses Alignment	29
Gambar 4.11 Urutan Fungsi Normalisasi	32
Gambar 5.1 Kode pembuatan <i>training</i> data.....	34
Gambar 5.2 Perubahan TweetTokenizer.....	35
Gambar 5.3 Kode untuk menampilkan tampilan utama	36
Gambar 5.4 Kode untuk preprocess.....	37
Gambar 5.5 Kode untuk inisialisasi data	38



Gambar 5.6 Kode untuk pembuatan <i>alignment</i>	38
Gambar 5.7 Hasil dari fungsi <i>diffib.SequenceMatcher()</i>	39
Gambar 5.8 Kode untuk pembuatan Training	39
Gambar 5.9 Kode Sistem Normalisasi	41
Gambar 5.10 Kode Sistem Testing	42
Gambar 6.1 Percobaan normalisasi dengan data training	45
Gambar 6.2 Kesalahan pada normalisasi data training	46
Gambar 6.3 Kalimat pada data testing yang berhasil dinormalisasi	46
Gambar 6.4 Kalimat yang tidak dinormalisasi dengan sempurna	47
Gambar 6.5 Contoh nama orang yang dinormalisasi sistem	49

DAFTAR TABEL

Tabel 2.1 Perbandingan tinjauan pustaka	6
Tabel 3.1 Kandidat <i>Spell Checking</i> pada kata <i>acress</i> (Jurafksy dan Martin, 2009)	9
Tabel 3.2 Probabilitas dari kandidat (Jurafksy dan Martin, 2009)	10
Tabel 3.3 $P(x/w)$ yang didapatkan (Jurafksy dan Martin, 2009)	12
Tabel 3.4 Hasil perhitungan $P(w)$ dan $P(x/w)$ (Jurafksy dan Martin, 2009)	12
Tabel 4.1 Contoh perhitungan normalisasi	31
Tabel 4.2 Contoh evaluasi	33
Tabel 6.1 Pengujian Fungsionalitas	43
Tabel 6.2 Nilai Akurasi dan F1 Score	44
Tabel 6.3 Hasil-hasil normalisasi	50
Tabel 6.4 Nilai Akurasi dan F1 Score dengan kamus	51