

DAFTAR ISI

Halaman Judul	ii
Halaman Pengesahan	ii
Halaman Pernyataan	iv
Halaman Persembahan	v
Halaman Motto	vi
PRAKATA	vii
DAFTAR ISI	ix
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
INTISARI	xvii
ABSTRACT	xviii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
II TINJAUAN PUSTAKA	7
III DASAR TEORI	13
3.1 <i>Rule-based System</i>	13
3.1.1 Pembuatan Aturan	13
3.1.2 Pencarian Solusi	13

3.1.3	<i>Inference Engine</i> dalam <i>Rule-based System</i>	14
3.2	<i>Lexical Analysis</i> (Analisis Leksikal)	16
3.3	Tokenisasi	16
3.4	Kutipan	17
3.4.1	Definisi	17
3.4.2	Pola Kalimat Kutipan pada Penelitian Lain	19
3.5	XPATH	21
3.5.1	Definisi	21
3.5.2	Contoh XPATH	22
3.6	JSON	23
3.7	<i>Named Entity Recognition</i> (Pengenal Entitas Bernama)	24
3.8	<i>Markov Model</i>	25
3.9	<i>Hidden Markov Model</i>	26
3.9.1	Peluang Awal	26
3.9.2	Peluang Transisi	27
3.9.3	Peluang Bersyarat Observasi	27
3.10	Algoritma Viterbi	27
3.11	<i>Regular Expression</i>	28
3.12	Pengukuran Kinerja Sistem	28
IV	ANALISIS DAN PERANCANGAN SISTEM	30
4.1	Deskripsi Umum	30
4.2	Analisis dan Kebutuhan Sistem	30
4.3	Pengumpulan Data	31
4.4	Penyimpanan Data	34
4.5	Pra-pemrosesan Data	37
4.6	Pembuatan Sistem	42
4.6.1	Pengenal Entitas Bernama	42
4.6.2	Pembuatan Sistem Ekstraksi Kalimat Pernyataan Berbasis Aturan	46
4.7	Rancangan Antarmuka Pengguna	53
4.7.1	Deskripsi	53
4.7.2	<i>Wireframe</i> rancangan antarmuka	53
4.8	Evaluasi Kinerja Sistem	61

V	IMPLEMENTASI SISTEM	63
5.1	Spesifikasi <i>Hardware</i> dan <i>Software</i>	63
5.2	Proses Pengumpulan Data	63
5.2.1	Mengumpulkan Daftar Tautan Berita	63
5.2.2	Mendapatkan Data Teks Berita	64
5.2.3	Melakukan <i>Grabbing</i> Teks Berita dari Situs Kompas, Tempo dan Tribun	66
5.3	Penyimpanan Data	73
5.4	Persiapan Data	78
5.4.1	Pelabelan Data untuk Sistem Ekstraksi Kalimat Pernyataan . .	79
5.4.2	Pelabelan Data untuk Sistem Pengenalan Entitas Bernama . .	80
5.4.3	Pembersihan Data	81
5.5	Tokenisasi	83
5.6	Implementasi Sistem Pengenalan Entitas Bernama	86
5.6.1	Pembuatan Model	86
5.6.2	Prediksi Label Entitas	90
5.7	Implementasi Proses Pembuatan Aturan untuk Ekstraksi Kalimat Pernyataan	92
5.7.1	Evaluasi Kalimat	93
5.7.2	Pembuatan Aturan	98
5.7.3	Pencocokan Aturan	100
VI	PENGUJIAN DAN PEMBAHASAN SISTEM	106
6.1	Pengujian Sistem Pengenalan Entitas Bernama	106
6.1.1	Data Latih	106
6.1.2	Data Uji	108
6.1.3	Hasil Pengujian	108
6.2	Pengujian Sistem Ekstraksi Kalimat Pernyataan	110
6.2.1	Data Basis Pengetahuan	110
6.2.2	Data Uji	110
6.2.3	Hasil Pengujian	110
6.2.4	Analisis Kesalahan Deteksi	115
VII	PENUTUP	120
7.1	Kesimpulan	120

7.2	Saran	121
-----	-----------------	-----

DAFTAR PUSTAKA	122
----------------	-----

A	BERKAS JSON SEBAGAI KAMUS <i>REPORTING VERB</i> PADA SISTEM BERBASIS ATURAN UNTUK DETEKSI PERNYATAAN	126
---	---	-----

DAFTAR TABEL

2.1	Perbandingan penelitian yang pernah dilakukan sebelumnya	12
3.1	Contoh kode XPATH dan keterangannya	23
3.2	<i>Confusion matrix</i>	29
4.1	Keterangan pelabelan data untuk sistem deteksi kutipan	38
4.2	Keterangan dan contoh pelabelan data untuk sistem pengenalan entitas bernama	41
5.1	Tautan ke daftar berita dan XPATH yang menunjukkan bagian daftar berita pada halaman	68
5.2	XPATH untuk masing-masing bagian berita	69
5.3	Kelas aturan dan pola yang digunakan	100
6.1	Rincian jumlah koleksi data latih untuk sistem pengenalan entitas bernama	107
6.2	Perbandingan hasil tokenisasi pada penelitian Fachri (2014) dan penelitian ini	107
6.3	Rincian jumlah koleksi data uji untuk sistem pengenalan entitas bernama	108
6.4	Model perhitungan kinerja sistem pengenalan entitas bernama	109
6.5	Hasil perhitungan kinerja sistem pengenalan entitas bernama	109
6.6	Model perhitungan kinerja sistem ekstraksi kalimat pernyataan	111
6.7	Hasil perhitungan kinerja sistem ekstraksi kalimat pernyataan dengan menggunakan data uji	111
6.8	Hasil perhitungan kinerja masing-masing kelas aturan dengan menggunakan data uji	113
6.9	Hasil perhitungan kinerja masing-masing kelas aturan dengan menggunakan data pada basis pengetahuan	114
6.10	Hasil perhitungan kinerja sistem ekstraksi kalimat pernyataan dengan menggunakan data basis pengetahuan	114

DAFTAR GAMBAR

2.1	Contoh kalimat tidak bertanda	9
3.1	Contoh kumpulan aturan	15
3.2	Contoh dokumen XML	23
3.3	Contoh penulisan objek pada JSON	24
3.4	Contoh penulisan array pada JSON	24
3.5	Markov chain pada penelitian Fachri (2014)	25
4.1	Diagram alir cara mendapatkan data teks berita <i>online</i>	32
4.2	<i>Pseudocode</i> untuk mendapatkan data teks berita	33
4.3	<i>Pseudocode</i> untuk menguraikan data	34
4.4	Rancangan struktur basis data	35
4.5	Contoh data yang telah diberi label sebagai basis pengetahuan pembuatan aturan	38
4.6	Contoh data yang telah diberi label untuk sistem pengenalan entitas bernama	42
4.7	Diagram alir pembuatan model sistem pengenalan entitas bernama	44
4.8	<i>Pseudocode</i> perhitungan peluang awal	45
4.9	<i>Pseudocode</i> perhitungan peluang transisi	45
4.10	<i>Pseudocode</i> perhitungan peluang bersyarat observasi fitur 1	45
4.11	<i>Pseudocode</i> perhitungan peluang bersyarat observasi fitur 2	45
4.12	<i>Pseudocode</i> perhitungan peluang bersyarat observasi fitur 3	46
4.13	Diagram alir pembuatan aturan untuk sistem ekstraksi kalimat pernyataan	47
4.14	Contoh kutipan langsung dalam berita	48
4.15	Diagram alir pencocokan aturan untuk sistem ekstraksi kalimat pernyataan	51
4.16	<i>Pseudocode</i> sistem untuk mencari aturan yang cocok sesuai kalimat pernyataan	52
4.17	Contoh hasil keluaran saat menjalankan aturan	52
4.18	<i>Wireframe</i> halaman utama sistem	54
4.19	<i>Wireframe</i> halaman pelabelan	55
4.20	<i>Wireframe</i> halaman hasil pelabelan	55
4.21	<i>Wireframe</i> halaman pelatihan sistem pengenalan entitas bernama	56

4.22	<i>Wireframe</i> halaman prediksi entitas bernama	57
4.23	<i>Wireframe</i> halaman evaluasi kalimat	57
4.24	<i>Wireframe</i> halaman daftar aturan	58
4.25	<i>Wireframe</i> halaman evaluasi sistem ekstraksi kalimat pernyataan berbasis aturan	59
4.26	<i>Wireframe</i> halaman daftar berita	60
4.27	<i>Wireframe</i> halaman ekstraksi kalimat pernyataan	60
4.28	Diagram alir evaluasi kinerja sistem	62
5.1	Fungsi <i>grab</i> untuk membuka dan memilih bagian daftar berita	64
5.2	Kode untuk mendapatkan bagian konten dari suatu website dengan XPath	65
5.3	Kode untuk membersihkan teks dari kode html	67
5.4	Halaman indeks berita kompas.com	70
5.5	Daftar berita kompas.com	71
5.6	Mendapatkan daftar tautan berita pada kompas.com	72
5.7	Penguraian tanggal, judul, isi dan kategori berita pada kompas.com	72
5.8	Contoh keluaran fungsi kompas_news	73
5.9	Kode untuk melakukan <i>grabbing</i> pada situs kompas.com	74
5.10	Kode untuk melakukan <i>grabbing</i> pada situs tribunnews.com	75
5.11	Kode untuk melakukan <i>grabbing</i> pada situs tempo.co	76
5.12	Kode untuk melakukan <i>grabbing</i> melalui terminal	77
5.13	Cara menjalankan pengumpulan data pada terminal	78
5.14	Tampilan sistem <i>tagger</i> untuk pelabelan data sistem ekstraksi kalimat pernyataan	80
5.15	Tampilan sistem <i>tagger</i> untuk pelabelan data sistem pengenalan entitas bernama	81
5.16	<i>Query SQL</i> untuk melakukan perubahan karakter pada data	82
5.17	Fungsi untuk memecah teks menjadi sekumpulan kalimat	85
5.18	Hasil <i>regular expression</i> dalam pemecahan kalimat	86
5.19	Fungsi untuk melakukan pelatihan dan membuat model pada sistem pengenalan entitas bernama	88
5.20	Struktur <i>json</i> untuk penyimpanan model pada sistem pengenalan entitas bernama	89
5.21	Cuplikan kode untuk menghitung probabilitas label entitas bernama	91

5.22	Cuplikan kode untuk mencari probabilitas terbesar	92
5.23	<i>Interface</i> untuk setiap aturan	93
5.24	Tampilan evaluasi kalimat	94
5.25	Kode untuk mengekstraksi kalimat pernyataan menggunakan kelas aturan yang telah dibuat	103
5.26	Contoh hasil prediksi	104
5.27	Contoh tampilan hasil prediksi	105
6.1	Cuplikan berita tempo.co yang memiliki kalimat kutipan langsung tanpa mencantumkan pembicara	119