

DAFTAR ISI

PERNYATAAN	iii
PRAKATA	iv
ARTI LAMBANG DAN SINGKATAN	vi
ABSTRACT	viii
INTISARI	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
BABI PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan masalah.....	3
1.3 Keaslian penelitian.....	3
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	9
2.1 Tinjauan Pustaka	9
2.1.1 Named Entity Recognition dalam Information Extraction.....	10
2.1.2 NER berbasis machine learning.....	10
2.1.3 Semi-supervised Learning NER.....	11
2.2 Landasan Teori.....	12
2.2.1 Natural Language Processing.....	12
2.2.2 Developing Named Entity Recognition	15
2.2.3 Hidden Markov Model PoS Tagger	18
2.2.4 Semi Supervised Learning.....	22
2.2.5 Co-Training	23
2.2.6 SVM	26
2.2.7 Labeled and Unlabeled data	34
2.2.8 Evaluasi performa NER.....	35
2.3 Pertanyaan Penelitian.....	36
BAB III METODOLOGI	38
3.1 Alat dan Bahan.....	38
3.1.1 Alat.....	38
3.1.2 Bahan.....	38
3.2 Jalannya Penelitian.....	39
3.3 Perancangan Sistem	42
3.3.1 Data Preparation.....	42

3.3.2	Pemrosesan Teks	44
3.3.3	Semi-supervised Learning Co-Training	51
3.3.4	NER	54
3.4	Evaluasi Sistem	54
BAB IV HASIL DAN PEMBAHASAN		57
4.1	Dataset	57
4.2	Pemrosesan Datasets	58
4.3	Pembelajaran Co-training dan Klasifikasi NER	59
4.4	Pembahasan	61
BAB V KESIMPULAN DAN SARAN		64
5.1	Kesimpulan	64
5.2	Saran	65
DAFTAR PUSTAKA		66

DAFTAR GAMBAR

Gambar 2.1 Skema Pemrosesan <i>Information Extraction</i> [1].....	16
Gambar 2.2 Hidden Markov Model.....	19
Gambar 2.3 Skema Kerja Algoritme <i>Co-Training</i> [5].....	24
Gambar 2.4 Mencari Optimal <i>Hyperplane</i> dengan Fungsi Linear	27
Gambar 2.5 Penggunaan Slack pada SVM.....	29
Gambar 2.6 Contoh Data Tidak Linear	31
Gambar 3.1 Diagram Alur Tahapan Penelitian	40
Gambar 3.2 Diagram Sistem Yang Dikembangkan.....	40
Gambar 3.3 Alur Pemrosesan Sentence Segmentation.....	44
Gambar 3.4 Skema Pemrosesan Part Of Speech Tagging	47
Gambar 3.5 Ekstraksi Data Tanpa Label	50
Gambar 3.6 Proposed Semi-supervised Learning System.....	51
Gambar 3.7 Co-Training Process.....	53
Gambar 3.8 Pseudo Code Co-Training Process.....	53
Gambar 4.1 Distribusi data berlabel dari DBpedia Indonesia	58

DAFTAR TABEL

Table 1.1 Penelitian klasifikasi entitas yang telah ada	7
Tabel 2.1.1 Tagset PoS Tag [17]	17
Tabel 2.1.2 Tagset PoS Tag (Lanjutan)	18
Tabel 2.2 Contoh Pemrosesan PoS Tagger Pada Sebuah Kalimat	21
Tabel 2.3 Contoh <i>instance</i> dengan x_1 dan x_2 untuk pembelajaran <i>co-training</i> ...	25
Tabel 3.1 Struktur Data Penyimpanan Artikel Berita.....	43
Tabel 3.2 Probabilitas Transisi $P_{ti ti-1} = \lambda \cdot P_{ti ti-1} + 1 - \lambda \cdot P_{ti}$	48
Tabel 3.3 Probabilitas Emisi $P_{wi ti} = \lambda \cdot P_{wi ti} + 1 - \lambda \cdot 1_{nwi}$	48
Tabel 3.4 Viterbi Decoding Iterasi Ke-1	48
Tabel 3.5 Viterbi Decoding Iterasi Ke-2...n	49
Tabel 3.6 Contoh Ekstraksi view-1 & view-2 Sebagai Data Tanpa Label	51
Tabel 3.6 Input Data Pemrosesan Co-Training	54
Tabel 3.7 Output Data Pemrosesan Co-Training.....	54
Tabel 4.1 Contoh Data Berlabel	57
Tabel 4.2 Contoh Data Tanpa Label.....	59
Table 4.3 Hasil Pengujian Sistem NER (dalam 100%)	60
Tabel 4.4 Rentang skala nilai precision, recall dan f1 score.....	61