

## DAFTAR ISI

PRAKATA.....	v
ARTI LAMBANG DAN SINGKATAN .....	vii
ABSTRACT.....	viii
INTISARI.....	ix
DAFTAR ISI .....	x
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah .....	3
1.3 Keaslian Penelitian.....	3
1.4 Tujuan Penelitian .....	8
1.5 Manfaat Penelitian .....	8
1.6 Batasan Penelitian.....	9
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	1
2.1 Tinjauan Pustaka .....	1
2.2 Landasan Teori.....	2
2.2.1 Pemetaan Ilmu Pengetahuan .....	2
2.2.2 Semantik Leksikal.....	3
2.2.3 Word Embedding.....	4
2.2.4 BERT .....	5
2.2.5 Clustering.....	8
2.2.6 Dimensional Reduction.....	10
2.3 Pertanyaan Penelitian.....	11
BAB III METODOLOGI PENELITIAN.....	12
3.1 Alat dan Bahan.....	12
3.1.1 Alat.....	12
3.1.2 Bahan.....	12
3.2 Jalannya Penelitian.....	13
3.2.1 Tahap 1 – Identifikasi Masalah dan Studi Literatur .....	14
3.2.2 Tahap 2 - Pengumpulan dan Pembuatan Dataset .....	15
3.2.3 Tahap 3 – Persiapan dan Instalasi Alat.....	18
3.2.4 Tahap 4 – Percobaan dengan <i>BERT</i> .....	19
3.2.5 Tahap 5 – Percobaan Clustering.....	21
3.2.6 Tahap 6 - Pelabelan Kelompok .....	22
3.2.7 Tahap 7 - Pemetaan Bidang Ilmu .....	23
3.2.8 Tahap 8 – Analisis Hasil.....	23
3.3 Perancangan Sistem .....	24
3.4 Cara Analisis .....	24
3.4.1 Evaluasi Clustering .....	24
3.4.2 Analisis Hasil Pelabelan Kelompok dan Pemetaan Bidang Ilmu ...	25
BAB IV HASIL DAN PEMBAHASAN.....	27

4.1	Hasil Percobaan dengan BERT .....	27
4.1.1	Preprocessing .....	27
4.1.2	Processing .....	29
4.1.3	Post Processing .....	31
4.1.4	Fine Tuning .....	35
4.2	Hasil Clustering .....	37
4.3	Hasil Pelabelan Kelompok .....	43
4.4	Hasil Pemetaan Bidang Ilmu .....	47
BAB V KESIMPULAN DAN SARAN .....		53
5.1	Kesimpulan .....	53
5.2	Saran .....	54
DAFTAR PUSTAKA .....		55
LAMPIRAN .....		1
Lampiran 1. Hasil uji coba menggunakan <i>dimensional reduction</i> (n=5) .....		1

## DAFTAR GAMBAR

Gambar 2.1 Peta bidang ilmu.....	3
Gambar 2.2 Representasi input dari BERT .....	6
Gambar 2.3 Arsitektur Pre-training pada BERT .....	7
Gambar 2.4 <i>Overlapping clustering</i> dengan FCM.....	9
Gambar 2.5 Metode Pendekatan <i>Dimensional Reduction</i> .....	10
Gambar 3.1 Struktur korpus data ETD UGM .....	13
Gambar 3.2 Struktur dataset <i>context-gloss pairs</i> .....	13
Gambar 3.3 Tahapan Penelitian .....	14
Gambar 3.4 Ilustrasi distribusi abstrak penelitian dengan BERT .....	19
Gambar 3.5 Contoh hasil pelabelan kelompok .....	23
Gambar 3.6 Contoh Hubungan Bidang Ilmu .....	23
Gambar 3.7 Arsitektur pemetaan bidang ilmu .....	24
Gambar 4.1 Hasil proses tokenisasi teks abstrak penelitian .....	27
Gambar 4.2 Hasil transformasi dari token WordPiece .....	29
Gambar 4.3 <i>Word embedding</i> dari sebuah abstrak penelitian .....	30
Gambar 4.4 Distribusi token abstrak penelitian menggunakan BERT .....	31
Gambar 4.5 Vektor representasi sebuah abstrak penelitian.....	31
Gambar 4.6 Distribusi abstrak penelitian menggunakan BERT <sub>BASE</sub> .....	32
Gambar 4.7 Grafik akurasi <i>fine-tuning</i> BERT dengan dataset DS_SKRIPSI.....	35
Gambar 4.8 Distribusi abstrak penelitian setelah dilakukan <i>fine-tuning</i> .....	36
Gambar 4.9 Grafik Skor F1 .....	41
Gambar 4.10 Grafik Koefisien Jaccard (JC) .....	42
Gambar 4.11 Grafik Fowlkes-Mallows (FM) .....	42
Gambar 4.12 Grafik Indeks Rand (RI).....	43
Gambar 4.13 Hasil pemetaan bidang ilmu .....	47
Gambar 4.14 Peta hubungan antara MIPA-Teknik .....	48
Gambar 4.15 Peta hubungan antara Ilmu Budaya-ISIPOL .....	48
Gambar 4.16 Peta hubungan antara ISIPOL-Ekonomi .....	49
Gambar 4.17 Peta hubungan Peternakan-Pertanian dan Pertanian-Farmasi .....	49

## DAFTAR TABEL

Tabel 1.1	Komparasi penelitian terkait analisis teks menggunakan <i>word embedding</i> dan <i>clustering</i> .....	5
Tabel 3.1	Struktur dataset latih DS_SKRIPSI-18 .....	15
Tabel 3.2	Struktur dataset latih DS_SKRIPSI-4 .....	16
Tabel 3.3	Struktur dataset validasi DS_TESIS-18 .....	17
Tabel 3.4	Komposisi jumlah abstrak penelitian dataset DS_TESIS-18 .....	17
Tabel 4.1	Interpretasi kelompok berdasarkan hasil pemetaan abstrak penelitian .....	33
Tabel 4.2	Performa <i>clustering</i> abstrak penelitian menggunakan BERT <sub>BASE</sub> dan hasil fine-tuning.....	38
Tabel 4.3	Performa <i>clustering</i> BERT <sub>BASE</sub> +DS_SKRIPSI-18 dengan metode EMBCLS dan EMBRerata .....	39
Tabel 4.4	Interpretasi istilah berdasarkan bidang ilmu tertentu untuk pelabelan kelompok .....	44
Tabel 4.5	Distribusi istilah di antara bidang ilmu yang saling berhubungan .....	50