

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PENGESAHAN PEMBIMBING.....	iii
HALAMAN PENGESAHAN PENGUJI.....	iv
PERNYATAAN BEBAS PLAGIAT	v
SURAT PERNYATAAN KARYA ASLI TUGAS AKHIR.....	vi
HALAMAN PERSEMBAHAN.....	vii
ABSTRAK.....	viii
ABSTRACT.....	ix
KATA PENGANTAR	x
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xv
DAFTAR GAMBAR.....	xvi
DAFTAR MODUL PROGRAM.....	xviii
DAFTAR LAMPIRAN.....	xix
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	2
1.3. Batasan Masalah	2
1.4. Tujuan Penelitian	3
1.5. Manfaat Penelitian	3
1.6. Metodologi Penelitian dan Pengembangan Sistem.....	3
1.6.1. Metode Penelitian.....	3
1.6.2. Metode Pengembangan Sistem	3
1.7. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1. Information Retrieval.....	5
2.2. <i>Search Engine</i>	5
2.3. Flowchart	5

2.4.	Text Mining	6
2.5.	Text Similarity	7
2.5.1.	String-based Similarity.....	8
2.5.2.	Corpus-based Similarity.....	9
2.5.3.	Knowledge-based Similarity	10
2.5.4.	Hybrid Similarity.....	11
2.6.	Preprocessing	12
2.6.1.	Tokenization.....	12
2.6.2.	Filtering	12
2.6.3.	Stemming	12
2.7.	Word Embedding.....	12
2.7.1.	Word2Vec	12
2.7.2.	FastText.....	15
2.8.	Vector Space Modeling.....	16
2.8.1.	Cosine Similarity.....	16
2.9.	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	17
2.10.	Root Mean Square (RMSE)	17
2.11.	Mean Reciprocal Rank (MRR)	17
2.12.	State of The Art.....	18
BAB III METODOLOGI PENELITIAN DAN PERANCANGAN SISTEM.....		19
3.1.	Metodologi Penelitian.....	19
3.1.1.	Variabel Penelitian	20
3.1.2.	Alat Penelitian	20
3.2.	Sumber Data.....	21
3.2.1.	Studi Pustaka	21
3.2.2.	Data Sekunder	21
3.2.3.	Populasi	21
3.2.4.	Sampel.....	21
3.3.	Teknik Pengumpulan Data.....	22
3.4.	Metodologi Pengembangan Software	22
3.5.	Arsitektur Sistem	23

3.6.	Data Preprocessing.....	24
3.6.1.	Cleansing.....	24
3.6.2.	Case Folding.....	26
3.6.3.	Tokenizing.....	27
3.6.4.	Stopword Removal.....	29
3.6.5.	Stemming	29
3.7.	Pembuatan Index TF-IDF	30
3.7.1.	Menyusun Bag of Words.....	31
3.7.2.	Perhitungan Term Frequency.....	32
3.7.3.	Perhitungan Inverse Document Frequency	32
3.7.4.	Perhitungan TF-IDF	33
3.7.5.	Proses Similarity dengan TF-IDF	33
3.8.	Pembuatan Index <i>FastText</i>	35
3.8.1.	Proses Training.....	35
3.8.2.	Pembuatan Set of Document untuk Proses Similarity	36
3.8.3.	Proses Similarity dengan <i>FastText</i>	37
3.9.	Perancangan Pengujian	38
BAB IV HASIL, PENGUJIAN DAN PEMBAHASAN		40
4.1.	Hasil Penelitian	40
4.1.1.	Pengambilan Data Eprint	40
4.1.2.	Preprocessing	40
4.1.3.	Pembuatan Index Algoritma TF-IDF	43
4.1.4.	Proses Pencarian dengan Algoritma TF-IDF	44
4.1.5.	Pelatihan Model <i>Fasttext</i>	45
4.1.6.	Pembuatan Index <i>Fasttext</i>	46
4.1.7.	Proses Pencarian dengan Algoritma <i>Fasttext</i>	46
4.2.	Pengujian Penelitian.....	50
4.2.1.	Pembuatan Kasus Uji	50
4.2.2.	Pengujian Manual.....	51
4.2.3.	Pengujian Hasil TF-IDF	52
4.2.4.	Pengujian Hasil <i>Fasttext</i>	57

4.2.5. Perbandingan Hasil Pengujian	62
4.3. Pembahasan.....	72
BAB V KESIMPULAN DAN SARAN	78
5.1. Kesimpulan	78
5.2. Saran	78
DAFTAR PUSTAKA	79
LAMPIRAN.....	83