

DAFTAR ISI

PERNYATAAN BEBAS PLAGIASI.....	iii
PRAKATA.....	iv
ARTI LAMBANG DAN SINGKATAN	vi
ABSTRACT.....	vii
INTISARI.....	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan masalah.....	8
1.3 Keaslian penelitian.....	9
1.4 Tujuan Penelitian	10
1.5 Manfaat Penelitian	11
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	12
2.1 Tinjauan Pustaka	12
2.2 Landasan Teori	14
2.2.1 Information Retrieval System.....	14
2.2.2 Question-Answering Sistem	15
2.2.3 Inverted Index	16
2.2.4 <i>Boolean Query</i>	16
2.2.5 Coarse-Fine Taxonomy.....	17
2.2.6 Stanford NLP	18
2.2.7 Pembobotan TFIDF	18
2.2.8 <i>Vector Space Model</i>	19
2.2.9 <i>Cosine Similarity</i>	20
2.2.10 <i>Query Expansion</i>	21
2.3 Hipotesis	24
BAB III METODOLOGI.....	25
3.1 Alat dan Bahan.....	25
3.1.1 Alat.....	25
3.1.1.1 Perangkat keras.....	25
3.1.1.2 Perangkat Lunak.....	25
3.1.2 Bahan/Dataset	25
3.2 Alur Penelitian	26
3.3 Perancangan Sistem	27
3.3.1 <i>Preprocessing</i>	30
3.3.1.1 <i>Lowercasing</i>	30

3.3.1.2	<i>Stemming</i>	30
3.3.1.3	<i>Inverted Index</i>	30
3.3.1.4	Pembuatan Vektor TFIDF Dokumen	31
3.3.2	<i>Question Analysis</i>	32
3.3.2.1	Ekstraksi Entitas	32
3.3.2.2	Klasifikasi Pertanyaan ke dalam Fine Taxonomy	34
3.3.2.3	Ekspansi Kueri	36
3.3.3	Document Retrieval	41
3.3.3.1	<i>Boolean Retrieval</i>	41
3.3.3.2	<i>Similarity Count</i>	43
3.4	Evaluasi	45
3.4.1	<i>Recall</i>	46
3.4.2	<i>Mean Reciprocal Rank</i>	47
BAB IV HASIL DAN PEMBAHASAN		49
4.1	Dataset	49
4.1.1	Dokumen	49
4.1.2	Pertanyaan	49
4.1.3	Dokumen Sumber Jawaban	50
4.2	Eksperimen	50
4.2.1	Eksperimen Untuk Mencari Threshold Jumlah Dokumen Yang Ditemu Kembalikan	53
4.2.1	Eksperimen Menggunakan Semua Dataset	54
4.2.2	Pengujian pada Threshold Top-N Jumlah Dokumen Yang Ditemu Kembalikan	56
4.2.3	Analisis Efektivitas Kueri Hasil Ekspansi dalam meningkatkan bobot similaritas	57
4.2.4	Analisis Efektivitas penggunaan Boolean “OR” dalam mengambil dokumen relevan	60
4.2.5	Analisis Pengaruh Jumlah Dokumen Relevan Terhadap Lama Waktu Eksekusi	60
4.3	Diskusi	62
4.3.1	Dokumen berada di luar dari posisi 20 teratas	62
4.3.2	Pemberian Kelas yang salah pada tahap Klasifikasi kelas pertanyaan	64
4.3.3	Jumlah Dokumen Relevan yang banyak	66
4.3.4	Kueri Ekspansi Yang Tidak Berhasil Meningkatkan Bobot Similaritas	69
4.4	Kelebihan dan Kekurangan Sistem	70
4.4.1	Kelebihan Sistem	70
4.4.2	Kekurangan Sistem	70
BAB V KESIMPULAN DAN SARAN		71
5.1	Kesimpulan	71
5.2	Saran	72
DAFTAR PUSTAKA		74
LAMPIRAN		L-1