

Daftar Isi

Lembar Pernyataan	i
Lembar Pengesahan	ii
Abstrak	iii
Kata Pengantar	iv
Daftar Isi	vi
Daftar Gambar	viii
Daftar Tabel	x
Daftar Istilah	xi
1 Pendahuluan	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Tujuan	2
1.4 Metodologi Penyelesaian Masalah	2
1.5 Sistematika Penulisan	3
2 Tinjauan Pustaka	4
2.1 Question Answering	4
2.1.1 Question-Comment Similarity	5
2.1.2 Question-Question Similarity	5
2.1.3 Question-External Comment Similarity	5
2.2 Penelitian Terkait	5
2.2.1 Penelitian Ilmuwan IBM	6
2.2.1.1 Metode Pembelajaran Pemberian Peringkat	6
2.2.2 Kelompok JAIST	7
2.2.2.1 Praproses	7
2.2.2.2 Fitur	7
2.3 Dataset	8
2.3.1 SemEval-2015 Task 3	8
2.3.1.1 English Data: CQA-QL Corpus	8
2.3.1.2 Arabic Data: Fatwa Corpus	8
2.3.2 SemEval-2016 Task 3	8
2.3.3 Stack Exchange	10
2.4 WordNet	11
2.5 Extensible Markup Language	11
2.6 Praproses	12
2.6.1 Text Extraction	12
2.6.2 Tokenization	12
2.6.3 Stopword Removal	12
2.6.4 Lemmatization	12
2.6.5 POS Tagging	13
2.7 Ekstraksi Fitur	13

2.7.1	Word Matching	13
2.7.1.1	Special Words	13
2.7.1.2	Question Author	14
2.7.1.3	Question–Answer Pair	14
2.7.2	Similarity	14
2.7.2.1	Cosine Similarity	14
2.7.2.2	Average Semantic Similarity	15
2.7.2.3	Align Disambiguate and Walk Semantic Similarity	16
2.8	Klasifikasi	18
2.8.1	Logistic Regression	18
2.8.2	Support Vector Machine	18
2.9	Evaluasi Performansi Sistem	19
2.9.1	Evaluasi Klasifikasi	19
2.9.1.1	Precision dan Recall	19
2.9.1.2	F1 Measure	19
2.9.2	Evaluasi Pemberian Peringkat	20
2.10	WEKA	20
3	Sistem yang dibangun pada Tugas Akhir	22
3.1	Gambaran Umum Sistem	22
3.1.1	Praproses	22
3.1.2	Ekstraksi Fitur	24
3.1.2.1	Special Words	24
3.1.2.2	Question Author	25
3.1.2.3	Question–Answer Pair	25
3.1.2.4	Cosine Similarity	26
3.1.2.5	Average Semantic Similarity	26
3.1.2.6	ADW Semantic Similarity	27
3.1.3	Klasifikasi dan Pemberian Peringkat Komentar	28
3.2	Evaluasi	28
3.3	Pengembangan	28
3.4	Analisis Kebutuhan Sistem	29
4	Evaluasi	30
4.1	Dataset	30
4.2	Baseline	30
4.3	Evaluation Metric	31
4.4	Pengujian	31
4.4.1	Tujuan Pengujian Sistem	31
4.4.2	Skenario Pengujian Sistem	32
4.5	Hasil Pengujian	32
4.5.1	Analisis Hasil Ekstraksi Fitur	32
4.5.1.1	Good Words	33
4.5.1.2	Bad Words	33
4.5.1.3	Emoticon	35
4.5.1.4	Question Author	37

4.5.1.5	Question-Answer Pair	39
4.5.1.6	Cosine Similarity	42
4.5.1.7	Average Semantic Similarity	47
4.5.1.8	ADW Semantic Similarity	51
4.5.2	Analisa Hasil Klasifikasi dan Pemberian Peringkat	56
4.5.2.1	Hasil Logistic Regression	56
4.5.2.2	Hasil Support Vector Machine	57
4.5.3	Analisis Kontribusi Fitur	57
5	Kesimpulan dan Saran	60
5.1	Kesimpulan	60
5.2	Saran	60
	Daftar Pustaka	60
	Lampiran 1: Contoh Dataset SemEval-2016 Task 3	63
	Lampiran 2: Contoh Hasil Ekstraksi Fitur	66
	Lampiran 3: Source Code	68