

Contents

Statement of Originality	ii
Supervisory Agreement	iii
Contents	vi
List of Figures	vii
List of Tables	viii
Glossary	ix
Abstract	xi
1 Introduction	1
1.1 Background	1
1.2 Problem Statement	2
1.3 Research Objectives	2
1.4 Delimitations	3
1.5 Research Methodology	3
1.6 Book Structure	4
2 Literature Review	5
2.1 The Holy Book Quran	5
2.2 Semantic Relation	5
2.2.1 Synonymy	6
2.2.2 Hyponymy	6
2.2.3 Antonymy	6
2.2.4 Polysemy	6
2.2.5 Meronymy	7
2.3 Lexical Semantic Textual Similarity	7
2.4 WordNet	7
2.5 Similarity Measurement by Yuhua Li et. al.	9
2.5.1 Strategy 1	9

2.5.2	Strategy 2	9
2.5.3	Strategy 3	10
2.5.4	Strategy 4	10
2.5.5	Strategy 5	10
2.6	Ontology of Quranic Concepts	11
2.7	Ontology Merging	11
2.8	Gold Standard	11
2.8.1	Existing Semantic Similarity Gold Standard	12
2.8.2	Inter-annotator Agreement	13
2.9	Correlation	13
2.9.1	Pearson’s Correlation	14
2.9.2	Spearman’s Correlation	14
3	System Design	15
3.1	Data Analysis	16
3.2	Gold Standard Development	18
3.2.1	Manual Selection of Quranic Concepts	18
3.2.2	Concepts Substitution	19
3.2.3	Manual Selection of Pairs of Quranic Concepts	22
3.2.4	Questionnaire Structure and Design	23
3.3	System Development	24
3.3.1	Preprocessing	24
3.3.2	Calibrating Constant α and β	24
3.4	Testing Scenario	25
3.5	Measuring Correlation	25
3.6	System Implementation Specification	25
4	Evaluation	27
4.1	Gold Standard	27
4.1.1	Dissimilarity is More Agreeable	28
4.1.2	Inter-annotator Agreement	28
4.2	Automatic System	29
4.2.1	Yuhua Li’s Strategies	29
4.2.2	The Best Yuhua Li’s Strategy	34
4.2.3	Reviewing The Correlation Result	35
4.2.4	Low Ratings in General	36
4.2.5	Similarity Between <i>Entity</i> and Any Noun Concepts	37

5 Conclusion and Recommendation	38
5.1 Conclusions	38
5.2 Suggestions	39
Bibliography	44
Appendices	45
Appendix A. Unrecognized Tokens in Saheeh International Dataset by WordNet	45
Appendix B. Questionnaire Instruction (Written in Indonesian)	46
Appendix C. Sample of Questionnaire Questions (Written in Indonesian) .	47
Appendix D. The Exact Value of WordNet Features	48
Appendix E. Gold Standard Dataset and System's Output Scores	53