LIST OF TABLES

Table 2. 1 The Classification result of Fawaz et.al. research [13] 7
Table 2. 2 The result achieved by Dastgheib and Koleini research [12] 8
Table 2. 3 The Flat Measure result in Adams and Bedrick research [17] 9
Table 2. 4 The Hierarchical Measures result in Adams and Bedrick research [17] .. 9
Table 2. 5 Example of term-document matrix [20] 12
Table 2. 6 Example of matrix U after SVD calculation [20] 13
Table 2. 7 Example of Matrix S after SVD calculation [20] 13
Table 2. 8 The example of matrix V^T after SVD calculation [20] 13
Table 2. 9 The example of truncated matrix U [20] 14
Table 2. 10 The example of truncated matrix S [20] 15
Table 2. 11 The example of matrix V^T' [20] 15
Table 2. 12 The example of the truncated SVD document matrix [20] 15
Table 2. 13 Confusion Matrix [24] 18
Table 2. 14 The example of word vectors 28
Table 2. 15 The example of document vector calculation 29
Table 2. 16 The comparison of the number of terms processed 30
Table 3. 1 Sample of the article 20
Table 3. 2 Similarity measurement on sample data 20
Table 3. 3 The number of the same terms appears among categories 21
Table 3. 4 The percentage of the number of the same terms appears among categories 22
Table 3. 5 The dataset used 24
Table 3. 6 Dataset Distribution 32
Table 4. 1 Types of Stoplist 36
Table 4. 2 Category combination used in similarity measurement 38
Table 4. 3 Classification result on different stoplist 39
Table 4. 4 Classification performance on different dataset based on TFIDF 40

Table 4. 5 Classification performance on different dataset based on LSA with TFIDF 40
Table 4. 6 LSA performance with TFIDF feature.....41
Table 4. 7 LSA performance with word2vec feature 42
Table 4. 8 Classification Performance based on different feature representation ...43
Table 4. 9 The performance of the classifier 44
Table 4. 10 Articles used to measure the similarity between articles.....45
Table 4. 11 Sample of the similarity measurement between *aqidah* and *ibadah* article.....46
Table 4. 12 Classification performance on similar category using LSA method 50