LIST OF TABLES

Table 1.1  Transaction Database T  4
Table 1.2  Frequent Item Sets of T (minimum support=50%)  4
Table 2.1  Comparison of the F-measure (Beil F., Ester M. and Xu X., 2002)  22
Table 2.2  F-measure of the Clustering Algorithm (Yanjun Li and Chung, 2005)  24
Table 3.1  Word Sequences Associated with the Nodes in Figure 3.5 (Yanjun Li et al., 2008 b)  57
Table 4.1  Size Reduction of Improved Stemming  80
Table 4.2  Words Reduction of Improved Stemming  81
Table 4.3  Size Reduction using Frequent Term Set Generation Algorithm  92
Table 4.4  Words Reduction using Frequent Term Set Generation Algorithm  93
Table 5.1  Matching Percentage of Books Sample with Various Minimum Support  114
Table 5.2  Matching Percentage of Reuters-21578 Mails with Various Minimum Support  116
Table 6.1  Representation of Text Database (D) as Frequent Terms  122
Table 6.2  Representation of Text Database D as Frequent Vector Space Model  
Table 6.3  F-measure Value for Reuters Transcribed Subset  
Table 6.4  Paired t-test between Algorithms based on F-measure for RTS Sample with respect to Minimum Support  
Table 6.5  F-measure Value for Mini-News Group – 700 Documents  
Table 6.6  F-measure Value for Mini-News Group – 1200 Documents  
Table 6.7  Purity Value for Reuters Transcribed Subset  
Table 6.8  Purity Value for Mini-News Group – 700 Documents  
Table 7.1  Test Data Sets  
Table 8.1  Memory Reduction of Stemming for Different Cases  
Table 8.2  Dimension Reduction of Stemming for Different Cases  
Table 8.3  Memory Reduction using Frequent Term Set  
Table 8.4  Dimension Reduction using Frequent Term Set  
Table 8.5  Performance for Reuters Transcribed Subset  
Table 8.6  Performance for Mini-News Group  
Table 8.7  Performance of Proposed Algorithms for Different Data Sets  
Table 8.8  Paired t-test between Algorithms based on Execution Time for RTS Sample with respect to Minimum Support