# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF SYMBOLS AND ABBREVIATIONS</td>
<td>vi</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>ix</td>
</tr>
</tbody>
</table>

## CHAPTERS

1. **INTRODUCTION**

   1.1 INTRODUCTION
   1.2 ASSOCIATION MINING
   1.3 CLUSTER ANALYSIS
   1.4 TEXT CLUSTERING
   1.5 MOTIVATION
   1.6 OBJECTIVES
   1.7 BASIC ISSUES
   1.8 THESIS ORGANISATION

2. **STUDY ON TEXT CLUSTERING**

   2.1 INTRODUCTION
   2.2 HIERARCHICAL BASED CLUSTERING
   2.3 FREQUENT ITEM SET BASED CLUSTERING
   2.4 HEAVY FREQUENT VECTOR BASED CLUSTERING
   2.5 FREQUENT WORD SEQUENCE CLUSTERING
   2.6 FREQUENT TERM BASED CLUSTERING
   2.7 CONCEPT BASED CLUSTERING
   2.8 SEMANTIC APPROACH BASED CLUSTERING
   2.9 SUMMARY
7.5 APPLICATIONS

7.5.1 Document Organisation 156
7.5.2 Text Filtering 156
7.5.3 Hierarchical Categorization of Web Pages 157
7.5.4 Word Sense Disambiguation 158
7.5.5 Automated Survey Coding 158
7.5.6 Automated Authorship Attributes and Genre Classification 159

7.6 DISCUSSION 161

8 RESULTS AND DISCUSSION 162

8.1 INTRODUCTION 162
8.2 DOCUMENT STEMMING 162
8.3 FREQUENT TERM SET GENERATIONS 166
8.4 DOCUMENT CLUSTERING 169
8.5 SUMMARY 175

9 CONCLUSION AND FUTURE ENHANCEMENT 176

9.1 CONCLUSIONS 176
9.2 LIMITATIONS 178
9.3 RECOMMENDATION FOR FURTHER RESEARCH 179

REFERENCES 181

APPENDICES

A. GLOSSARY OF TERMS 189
B. USER INTERFACE SCREENS 191
C. PROTOTYPE 195

LIST OF PUBLICATIONS 187