# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declaration</td>
<td>vi</td>
</tr>
<tr>
<td>Certificate</td>
<td>vii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>viii</td>
</tr>
<tr>
<td>List of Publications</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xiv</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>xv</td>
</tr>
<tr>
<td>Chapter 1: Introduction</td>
<td>1-10</td>
</tr>
<tr>
<td>1.1 General</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Research Motivation</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Problem statement</td>
<td>6</td>
</tr>
<tr>
<td>1.4 Objectives of the Thesis</td>
<td>8</td>
</tr>
<tr>
<td>1.5 Organization of the thesis</td>
<td>9</td>
</tr>
<tr>
<td>Chapter 2: Literature Review</td>
<td>11-52</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>11</td>
</tr>
<tr>
<td>2.2 Web Crawler Architectures</td>
<td>11</td>
</tr>
<tr>
<td>2.3 Meta Search Engines</td>
<td>23</td>
</tr>
<tr>
<td>2.4 Automatic Text Summarizers</td>
<td>32</td>
</tr>
<tr>
<td>2.5 Search Engine Evaluation</td>
<td>42</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>53</td>
</tr>
<tr>
<td>3.2 Web Directories</td>
<td>53</td>
</tr>
<tr>
<td>3.2.1 Issues and Challenges with Web Directories</td>
<td>54</td>
</tr>
<tr>
<td>3.3 Search Engine</td>
<td>55</td>
</tr>
<tr>
<td>3.3.1 Search Engine Architecture</td>
<td>57</td>
</tr>
</tbody>
</table>
3.3.2 Issues and Challenges with Search Engines 61
3.4 Meta Search Engine 62
  3.4.1 Architecture of Meta Search Engine 62
  3.4.2 Issues and Challenges with Meta Search Engines 63
3.5 Crawling the Web 67
  3.5.1 Basic Component of Web Crawler 69
  3.5.2 Features of a Web Crawler 71
3.6 Web Crawler: Issues and Challenges 73
  3.6.1 Basic Crawler 73
    3.6.1.1 Issues and Challenges with Basic Crawler 73
  3.6.2 Focused Crawler 74
    3.6.2.1 Focused Crawling Techniques 75
    3.6.2.2 Issues and Challenges with Focused Crawler 77
  3.6.3 Parallel Crawler 78
    3.6.3.1 Features of Parallel Web Crawler 79
    3.6.3.2 Issues and Challenges with Parallel Web Crawler 82
  3.6.4 Distributed Crawler 83
    3.6.4.1 Architecture of Distributed Crawler 83
    3.6.4.2 Issues and Challenges with Distributed Crawler 84
3.7 Crawling Policies 85
  3.7.1 Similarity Based Crawling Policy 85
  3.7.2 HITS Based Crawling Policy 86
  3.7.3 DOM Based Crawling Policy 87
  3.7.4 Page Rank Algorithm 87
  3.7.5 Location Based Policy 88
  3.7.6 Link Based Approach 89
3.8 Crawling Algorithms 89
  3.8.1 Depth First Search Algorithm 89
  3.8.2 Breadth First Search Algorithm 91
Chapter 5: Efficient Meta Search Engine  135-160
    5.1 Introduction  135
    5.2 Problems with Traditional and Existing Meta Search Engines  136
    5.3 The Architecture of the Proposed MSE  138
        5.3.1 Working of Proposed MSE  145
        5.3.2 Experimental Setup  149
        5.3.3 Results and Discussion  149
        5.3.4 Parameters Affecting the Efficiency of the MSE  157
    5.4 Summary  159

Chapter 6: Automatic Multiple Document Text Summarizer  161 - 181
    6.1 Introduction  161
    6.2 Problems with Traditional and Existing ATS  162
    6.3 Proposed Automatic Multiple Document Text Summarizer  163
        6.3.1 Working of Proposed Summarizer  169
        6.3.2 Experimental Setup  172
        6.3.3 Results and Discussion  175
    6.4 Applications of the Proposed Automatic Text Summarizer  178
    6.5 Summary  180

Chapter 7: A Framework to Find Specific Search Engine for a Domain  182-205
    7.1 Introduction  182
    7.2 Problems with Existing Evaluation Techniques of SE  183
    7.3 Proposed Framework for SEE  185
        7.3.1 Selection of SE, Domain, Query and Criterion  185
        7.3.2 Ordering of Results Using Relevancy  189
        7.3.3 Relevancy Score and Relevancy Calculation  190
        7.3.4 The Working of Proposed SEE Framework  191
        7.3.5 Experimental Setup  194
        7.3.6 Results and Discussion  194
    7.4 Summary  204
Chapter 8: Conclusion and Future Scope 206-208
8.1 Conclusion 206
8.2 Future Scope 208

References 209-235

Appendix A: List of URLs Used for Web Page Change Detection 236-239