

1950  
BOND

# Contents

---

1950

## LIST OF MAIN CONTENTS

SL.NO	TITLE	PAGE NO
1	INTRODUCTION	1
2	REVIEW OF LITERATURE	25
3	METHODOLOGY	45
4	RESULTS AND DISCUSSION	60
5	SUMMARY AND CONCLUSION	74
	BIBLIOGRAPHY	

# CONTENTS

CHAPTER	TITLE	PAGE No.
<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
1.1	Web mining – An Overview	2
1.2	General Framework of Web Mining Process	3
1.3	Paradigm of Web Mining	4
1.4	Web Usage Mining (WUM)	6
1.5	Web Usage Data	9
1.6	General Web Log Mining System	12
1.7	Clustering and Classification	19
1.8	Motivation and Objectives	20
1.9	Chapter Formulation	23
1.10	Chapter Summary	24
<b>2</b>	<b>REVIEW OF LITERATURE</b>	<b>25</b>
2.1	Web Usage Mining	25
2.2	User Navigation Pattern Mining	31
2.3	Classification and Clustering Techniques	35
2.4	Graph Partition Clustering	39
2.5	Ant-Based Clustering	41
2.6	Log File Analysis for Web Usage Mining	43
2.7	Chapter Summary	44
<b>3</b>	<b>METHODOLOGY</b>	<b>45</b>
3.1	Web Usage Mining	45
3.2	Log Files	46
3.3	Preprocessing	48

3.4	Pattern Discovery	49
3.5	Pattern Analysis	50
3.6	Ant-Based Clustering (Offline Phase)	51
3.7	Longest Common Subsequence (LCS) problem	55
3.8	Online Phase - LCS approach	58
3.9	Chapter Summary	59
<b>4</b>	<b>RESULTS AND DISCUSSION</b>	<b>60</b>
4.1	Data environment	60
4.2	Preprocessing result	61
4.3	Clustering Results	63
4.4	Classification Results	70
4.5	Chapter Summary	73
<b>5</b>	<b>SUMMARY AND CONCLUSION</b>	<b>74</b>
	<b>BIBLIOGRAPHY</b>	

## LIST OF TABLES

<b>TABLE NO</b>	<b>PARTICULARS</b>	<b>PAGE NO</b>
<b>1</b>	<b>Web log file format</b>	<b>47</b>
<b>2</b>	<b>Web page and its numerical code</b>	<b>68</b>
<b>3</b>	<b>Extracted Navigation Patterns</b>	<b>69</b>
<b>4</b>	<b>Sample Pattern</b>	<b>71</b>
<b>5</b>	<b>Common Sequence and Weight</b>	<b>71</b>

## LIST OF FIGURES

<b>FIGURE NO</b>	<b>PARTICULARS</b>	<b>PAGE NO</b>
1	General Framework of Web Mining	3
2	Paradigm of Web Mining	4
3	Applications Areas	8
4	Potential Data Sources	10
5	Sample Web Server Log	11
6	Knowledge Discovery	12
7	Format of Web Log File	13
8	Sample Web Log File	13
9	Proposed System	23
10	Web Usage Mining	46
11	Ant based algorithm	53
12	Demo sequence of ant-based clustering of user sequences	54
13	Offline Phase	55
14	Online Phase	55

<b>15</b>	<b>Raw Web Log File</b>	<b>62</b>
<b>16</b>	<b>After Cleaning Step</b>	<b>62</b>
<b>17</b>	<b>Effect of cleaning step on raw web log file</b>	<b>64</b>
<b>18</b>	<b>User Identification</b>	<b>64</b>
<b>19</b>	<b>User Session Identification</b>	<b>65</b>
<b>20</b>	<b>Effect of user and session identification on web log data</b>	<b>65</b>
<b>21</b>	<b>Unique URLs identified</b>	<b>66</b>
<b>22</b>	<b>Numerical code assigned to each URL</b>	<b>66</b>
<b>23</b>	<b>Formatted web log file after preprocessing</b>	<b>67</b>
<b>24</b>	<b>Cluster Groups</b>	<b>67</b>
<b>25</b>	<b>Webpage Traffic</b>	<b>69</b>
<b>26</b>	<b>Referral Pattern</b>	<b>72</b>
<b>27</b>	<b>Comparison in terms of Accuracy</b>	<b>73</b>