

Contents

CONTENTS

Chapter. No	Title	Page. No
	LIST OF TABLES	i
	LIST OF FIGURES	ii
1.	INTRODUCTION	1-10
1.1	WEB MINING	1
1.2	TYPES OF WEB MINING	3
1.2.1	Web Content Mining	3
1.2.2	Web Structure Mining	3
1.2.3	Web Usage Mining	4
1.3	SESSION IDENTIFICATION	7
1.4	IMPORTANCE OF NAVIGATION	8
1.4.1	Benefits of Navigation Pattern	9
1.5	PROBLEM JUSTIFICATION	9
1.6	PROBLEM STATEMENT	10
1.7	OBJECTIVE OF THESIS	10
1.8	ORGANIZATION OF THESIS	10
1.9	CHAPTER SUMMARY	10

2.	REVIEW OF LITERATURE	11-19
2.1	INTRODUCTION	11
2.2	TECHNIQUES APPLIED IN WEB PAGE NAVIGATION	11
2.2.1	Cluster Based Techniques	11
2.2.2	FP-Growth Based Techniques	12
2.2.3	Flame Algorithm Based Techniques	14
2.3	SURVEY ON SESSION IDENTIFICATION	16
2.4	SURVEY ON WEB PERSONALIZATION	17
2.5	CHAPTER SUMMARY	19
3.	RESEARCH METHODOLOGY	20-30
3.1	OVERVIEW OF THE METHODOLOGY	20
3.2	PREPROCESSING	21
3.3	DBSCAN ALGORITHM	22
3.3.1	Flowchart	23
3.3.2	Steps	24
3.4	FP-GROWTH ALGORITHM	24
3.4.1	Steps	25
3.4.2	Process	26
3.5	FLAME ALGORITHM	27
3.6	PERFORMANCE EVALUATION METRICS	30

3.7	CHAPTER SUMMARY	30
4.	RESULTS AND DISCUSSION	31-40
4.1	DATA SET	31
4.1.1	Description of data set	31
4.2	EXPERIMENTAL SET UP	32
4.3	MATLAB	32
4.3.1	MATLAB System	33
4.4	PERFORMANCE EVALUATION	35
4.5	OBSERVED RESULTS	36
4.5.1	Average number of patterns and sessions identified	36
4.5.2	Precision and Recall	37
4.5.3	Accuracy and F-measure	38
4.6	CHAPTER SUMMARY	40
5.	CONCLUSION AND SCOPE FOR FUTURE ENHANCEMENT	41
	REFERENCES	42
	PUBLICATIONS	45
	ANNEXURES	46
I.	DATASET	46
II.	PROGRAM CODE	48
III.	SCREEN SHOTS	58

LIST OF TABLES

Table. No	Title	Page. No
4.1	Performance comparison of DBscan, FP-growth and Flame algorithm	36
4.2	Precision and Recall	38
4.3	Accuracy and F-Measure	39

LIST OF FIGURES

Figure. No	Title	Page. No
1.1	Web mining types	3
1.2	Log files in web proxy sever	6
3.1	Flowchart showing overall Research plan	20
3.2	Preprocessing Procedure	22
3.3	Flowchart for DBscan Algorithm	23
3.4	Pseudocode for DBscan Algorithm	24
3.5	Pseudocode for FP-Growth Algorithm	25
3.6	Flow chart for FP-growth Algorithm	27
3.7	Pseudocode for Flame Algorithm	29
4.1	Average number of patterns comparison	36
4.2	Number of Sessions identified	37
4.3	Precision and Recall metrics	38
4.4	Accuracy of Patterns Predicted	39
4.5	F-Measure of Patterns Predicted	40