

**CLUSTERING USERS FROM USER PROFILES  
IN THE DIGITAL LIBRARY**

by

**AO IEONG U**

**Master of Science in Software Engineering**

**2002**



**Faculty of Science and Technology  
University of Macau**

## TABLE OF CONTENTS

LIST OF FIGURES .....	iv
LIST OF TABLES .....	v
Chapter 1 Introduction .....	1
Chapter 2 Related Work in This Area .....	3
2.1 From User Access Patterns to Dynamic Hypertext Linking .....	3
2.1.1 Preprocessing .....	4
2.1.2 Clustering .....	5
2.1.3 Dynamic Link Generation .....	5
2.2 The View of Web Users .....	5
2.2.1 User's interest .....	6
2.2.2 Distinction rules .....	6
2.3 Discussion .....	6
Chapter 3 System Architecture .....	8
3.1 User Identification .....	8
3.2 User profile .....	10
3.3 User Cluster .....	11
3.4 Match new document .....	12
3.5 Dissemination (E-Mail) .....	12
3.6 Create Cluster with Session .....	12
Chapter 4 User Profile .....	15
4.1 User Profile modeling .....	15
4.1.1 User Profile Introduction .....	15
4.1.2 The Data Categories of a User Profile .....	16
4.1.3 The User Profile Schema .....	17
4.2 Profile Representation .....	19
4.2.1 Defining a Vector Space .....	19
Chapter 5 Clustering .....	22
5.1 Introduction .....	23
5.2 K-Means Algorithm .....	27
5.3 K-Means applied to user profiles .....	28
5.4 Enhanced Clustering Algorithm .....	39
Chapter 6 Match new document (Classification) .....	43
6.1 Introduction .....	43

6.1.1 Document Abstract Representation .....	43
6.2 Classification Algorithm .....	46
6.3 Classification algorithm applied to new document. ....	47
Chapter 7 Dissemination (E-Mail) .....	50
7.1 Strategies of dissemination new document .....	50
7.2 Ways of dissemination new document to users.....	50
7.3 Object of E-Mail .....	51
Chapter 8 Implementation.....	52
8.1 Table structure .....	54
8.2 Initialization .....	59
8.3 User clustering .....	60
8.3.1 Initial Clusters.....	60
8.3.2 Improving the cluster .....	61
8.3.3 Create cluster with session.....	62
8.4 Match new document (Classification).....	63
8.4.1 Initiation.....	63
8.4.2 Improve the new document classification result .....	64
8.5 Dissemination (E-Mail).....	65
Chapter 9 Conclusion.....	67
9.1 Conclusion of this paper .....	67
9.2 Future work.....	68
Appendix A Programs.....	69
A.1 Create Vector .....	69
A.1.1 Create_Vector.h .....	70
A.1.2 Create_Vector.cpp .....	71
A.2 Create session vector.....	73
A.2.1 Create_session_vector.h.....	74
A.2.2 Create_session_vector.cpp.....	75
A.3 Create Cluster.....	77
A.3.1 Create_Cluster.h.....	78
A.3.2 Create_Cluster.cpp.....	80
A.3.1 Create_Cluster.h.....	88
A.3.2 Create_Cluster.cpp.....	90
A.4 Create session cluster .....	98
A.4.1 Create_session_cluster.h.....	99
A.4.2 Create_session_cluster.cpp .....	101
A.5 Classification document .....	109
A.5.1 Class_doc.h.....	110

A.5.2 Class\_doc.cpp ..... 112

A.6 Send E\_Mail..... 119

    A.6.1 Send\_EMail.h..... 120

    A.6.1 Send\_EMail.cpp..... 121

APPENDIX B: Reference ..... 123

VITA..... 125