

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Improving Effectiveness of Dialogue in Learning Communities

A thesis presented in partial fulfilment of the requirements for the degree of

PhD in Information Systems

at Massey University, Palmerston North, New Zealand.

Jingyu Yang

2009

ABSTRACT

In a learning community, conventional discussion forums are integral to web-based interventions in traditional classrooms as well as on-line learning environments. Despite popular belief that they are a great success in fostering deep and meaningful discussions and support active learning; research has found that there are millions of messages posted by users to express such an opinion, but it is hard to be directly delivered to all users. Finally there are millions of postings in databases across the country stored away and never reused. This thesis introduces a PhD student's current research work. It proposes a distributed intelligent discussion forum system dedicated to supporting both students and teachers. The system is developed with the primary goal of reducing the number of problems associated with conventional discussion forum systems in web-based environments and improving the effectiveness of dialogue between students with each other and with teachers so that it can enhance each individual student's ability to share and learn knowledge.

ACKNOWLEDGEMENTS

In the last four years, the writing process of the thesis included both joyful and painful moment. Even there were times when I thought the work will never finish. Therefore I must be sincerely grateful to my supervisor, Dr. Kinshuk, Professor for his guidance, advice and inspiring wisdom throughout this project and for his lots of valuable support and supervision from the beginning of the research project, without which this thesis would not have been completed. He gave me such patient, effective support and encourages in such long time,

In addition, I would like to thank my whole family including my parents, my brother and my partner for their support and encouragement. And also thank my second supervisor, Dr. Lynn Jeffrey and all my friends who gave me such help and encourages during last four years.

Financial support towards this study from E-Learning Collaborative Development Fund (eCDF) (Ministry of Education, New Zealand), Massey University and IBM New Zealand is highly appreciated.

Your thought, help and kindness will encourage me to keep going!

Thanks You All!

Yours truly,
Jingyu Yang

CONTENTS

ABSTRACT	II
ACKNOWLEDGEMENTS	III
LIST OF TABLES.....	IX
LIST OF FIGURES.....	X
LIST OF ABBREVIATIONS	XII
CHAPTER 1 GENERAL INTRODUCTION	1
1.1 INTRODUCTION	1
1.2 DIALOGUE IN THE LEARNING COMMUNITIES.....	2
1.3 RESEARCH QUESTIONS, TREATMENT, METHODOLOGY AND SCOPE.....	3
1.3.1 Limitations of Discussion Forums in Educational Context	3
1.3.2 Research Questions.....	4
1.3.3 Treatment of Research Question.....	5
1.3.4 Methodology.....	6
1.3.5 Scope - eQuake Project.....	7
1.4 DISSERTATION STRUCTURE	8
CHAPTER 2 LITERATURE REVIEW.....	10
2.1 INTRODUCTION	10
2.2 ONLINE LEARNING COMMUNITIES.....	10
2.2.1 Community-based Learning.....	10
2.2.2 Online Learning Communities.....	13
2.2.3 C4P Model.....	14
2.2.4 Dialogue via Technologies	18
2.3 DISCUSSION FORUM SYSTEMS	21
2.3.1 Overview	21
2.3.2 The Use of Discussion Forums.....	22
2.3.3 General Introduction of Major Systems.....	24

2.3.4	Limitations of Current Discussion Forums	27
2.4	INFORMATION RETRIEVAL.....	29
2.4.1	Overview	29
2.4.2	Text Similarity.....	35
2.4.3	Relevance Feedback	37
2.4.4	Latent Semantic Indexing	42
2.4.5	Search Engine	48
2.5	RECOMMENDER SYSTEMS	52
2.6	CONCLUSION	56
CHAPTER 3 REQUIREMENTS ANALYSIS.....		57
3.1	INTRODUCTION	57
3.2	REQUIREMENTS GATHERING	57
3.2.1	Overview	57
3.2.2	Challenges in Discussion Forum	58
3.2.3	Requirement Analysis.....	61
3.2.4	Requirements Summary.....	63
CHAPTER 4 SOLUTION DESIGN		65
4.1	INTRODUCTION	65
4.2	EXISTING SOLUTIONS FOR TRADITIONAL DISCUSSION FORUMS.....	65
4.2.1	Overview	65
4.2.2	EQuake	66
4.2.3	Intelligent Discussion-Bot (ID-B).....	68
4.2.4	Limitations of Current Solutions	71
4.3	PROPOSED SOLUTION	72
4.3.1	Introduction.....	72
4.3.2	Requirements Recap	72
4.3.3	Solution Overview	73
4.3.4	High-Level System Architecture and Components Design	74
4.4	SUMMARY	78

CHAPTER 5 DIALOGUE MECHANISM DESIGN	79
5.1 INTRODUCTION	79
5.2 RELATED APPROACHES	79
5.2.1 Syntactically Enhanced Latent Semantic Analysis (LSA)	81
5.2.2 Assessment Model to Analyzing Student’s Performance	83
5.2.3 Ontology-based Information Retrieval System	88
5.2.4 Dialogue Manager	90
5.2.5 Summary	91
5.3 DIALOGUE MECHANISM DESIGN	93
5.3.1 Introduction	93
5.3.2 Cosine Methodology Based on LSA	93
5.3.3 Framework Design	96
5.3.4 Concept Classification & Information Retrieval	97
5.3.5 Relevance Feedback and Concept Reclassification	101
5.3.6 Summary	104
CHAPTER 6 DEVELOPMENT PLAN	106
6.1 INTRODUCTION	106
6.2 DEVELOPMENT OVERVIEW	106
6.3 FORUM SELECTION	109
6.3.1 Requirement Analysis	109
6.3.2 Survey of Existing Open Source Forum Systems	110
6.3.3 Base forum selection	114
6.4 DEVELOPMENT APPROACHES AND TECHNOLOGIES USED	117
6.4.1 Web Services	117
6.4.2 Agents	118
6.4.3 Object Relational Mapping (ORM) Tools	118
6.4.4 Development Approaches	119
6.4.5 Development Environment and Tools	121
CHAPTER 7 ARCHITECTURE AND IMPLEMENTATION	123

7.1 INTRODUCTION	123
7.2 ARCHITECTURE OVERVIEW	123
7.2.1 Plug-in Component	124
7.2.2 Dialogue Mechanism Component.....	126
7.2.3 Dispatch Component	131
7.2.4 Forum Display Component.....	132
7.2.5 Data Architecture.....	133
7.2.6 Deployment Architecture.....	136
7.3 IMPLEMENTATION	139
7.3.1 Overview	139
7.3.2 General Features	139
7.3.3 New Features	141
7.4 SUMMARY	145
CHAPTER 8 RELATED EVALUATION APPROACHES	146
8.1 INTRODUCTION	146
8.2 EVALUATION METHODOLOGIES	146
8.2.1 Evaluation for IR Performance	147
8.2.2 Evaluation for User Performance and Satisfaction	149
8.3 RELATED EVALUATION APPROACHES	150
8.3.1 Overview	150
8.3.2 User Performance versus Precision Measures for Simple Search Tasks	151
8.3.3 User-centred Evaluation of an E-learning Repository	151
8.3.4 Evaluating Effectiveness of a Context-based IR System	152
8.3.5 Evaluating an Intelligent Discussion-Bot	153
8.3.6 Assessing Distance Learning Student Performance.....	153
8.4 SUMMARY	154
CHAPTER 9 EVALUATION APPROACH	155
9.1 INTRODUCTION	155
9.2 EVALUATION GOALS	155

9.2.1 Research Objectives.....	156
9.2.2 Evaluation Goals.....	156
9.3 EVALUATION APPROACHES.....	158
9.3.1 Baseline Systems.....	158
9.3.2 Evaluation Approach.....	161
9.4 EVALUATION EXPERIMENTS.....	163
9.4.1 The First Experiment.....	164
9.4.2 The Second Experiment.....	166
9.5 SUMMARY.....	169
CHAPTER 10 EVALUATION RESULTS AND ANALYSIS.....	170
10.1 INTRODUCTION.....	170
10.2 THE FIRST EXPERIMENT.....	170
10.2.1 Experiment Procedure.....	170
10.2.2 Results and Analysis.....	173
10.3 THE SECOND EXPERIMENT.....	178
10.3.1 Experiment Procedure.....	178
10.3.2 Experiment Results and Analysis.....	182
10.3.3 User Satisfaction.....	187
10.4 DISCUSSION AND CONCLUSION.....	192
10.4.1 Further Discussion.....	192
10.4.2 Conclusion.....	195
CHAPTER 11 CONCLUSIONS AND FUTURE WORK.....	197
11.1 DISSERTATION SUMMARY.....	197
11.2 CONCLUSION: CONTRIBUTIONS, LIMITATIONS, AND FUTURE RESEARCH.....	200
REFERENCES.....	205
APPENDIX.....	228
APPENDIX I – STOP-WORD LIST.....	228
APPENDIX II – QUESTIONNAIRE.....	230

LIST OF TABLES

TABLE 2-1: FIVE COMPONENTS OF A COMMUNITY-BASED LEARNING ENVIRONMENT (KANG AND BYUN, 2001).....	12
TABLE 2-2: A QUERY STRING “COMPUTER-BASED INFORMATION LOOK-UP”	43
TABLE 3-1: ADDRESSING THE FIVE CHALLENGES IN DISCUSSION FORUMS.....	62
TABLE 4-1: CHARACTERISTICS OF A TRADITIONAL FORUM, EQUAKE & ID-B	71
TABLE 5-1: THE SUMMARY OF RELATED APPROACHES	92
TABLE 6-1: NEW SYSTEM DEVELOPMENT PLAN.....	108
TABLE 6-2: A LIST OF OPEN SOURCE FORUMS.....	110
TABLE 10-1: THE GENERAL INFORMATION OF PARTICIPANTS	172
TABLE 10-2: RELEVANCE RESULTS IN THE FIRST EVALUATION EXPERIMENT.....	175
TABLE 10-3: RATIO OF RELEVANCE RESULTS IN THE FIRST EVALUATION EXPERIMENT	175
TABLE 10-4: AVERAGE NUMBER OF RELEVANT POSTS FOUND FOR EACH SUB-TASK.....	183
TABLE 10-5: AVERAGE NUMBER OF RELEVANT POSTS FOUND FOR EACH SUB-TASK.....	184
TABLE 10-6: AVERAGE NUMBER OF RELEVANT POSTS FOUND FOR EACH SUB-TASK.....	185
TABLE 10-7: SUMMARY OF EVALUATORS.....	187
TABLE 10-8: STATISTICAL DATA FOR THE SECTION 1	188
TABLE 10-9: GENERAL INFORMATION OF EVALUATORS	189

LIST OF FIGURES

FIGURE 1-1: OUTLINE OF THE THESIS.....	9
FIGURE 2-1: THE PROCESS OF COMMUNITY-BASED LEARNING (PALLOFF & PRATT, 1999)	11
FIGURE 2-2: MODEL OF AN ELECTRONIC LEARNING COMMUNITY (SHEARD, 2004)	21
FIGURE 2-3: A WEBSITE USED FOR A COURSE (HTTP:IS157730.MASSEY.AC.NZ)	24
FIGURE 2-4: DIFFERENT TOPIC FOLDERS IN THE DISCUSSION FORUM.....	25
FIGURE 2-5: DISCUSSING AREA OF THE DISCUSSION FORUM	25
FIGURE 2-6: SCREENSHOT OF EQUAKE WEB FORUM (FROM HTTP://WWW.EQUAKE.ORG.NZ/M1).....	27
FIGURE 2-7: GENERAL MODEL OF INFORMATION RETRIEVAL (SALTON, 1983)	30
FIGURE 2-8: SEARCH ENGINE ARCHITECTURE IN GENERAL (ARASU, ET AL., 2001).....	50
FIGURE 4-1: USE OF TRADITIONAL FORUMS VS. EQUAKE FORUMS (SMESTAD, 2006)	66
FIGURE 4-2: EQUAKE SYSTEM ARCHITECTURE (SMESTAD, 2006)	67
FIGURE 4-3: SCREENSHOT OF EQUAKE WEB FORUM (SMESTAD, 2006)	68
FIGURE 4-4: DISCUSSION-BOT ARCHITECTURE (FENG ET AL., 2006)	69
FIGURE 4-5: USE OF PROPOSED SYSTEM FOR DIFFERENT INSTITUTIONS	74
FIGURE 4-6: SYSTEM LAYER DIAGRAM	75
FIGURE 4-7: PLUG-IN COMPONENT	76
FIGURE 4-8: DIALOGUE MECHANISM COMPONENT	77
FIGURE 5-1: VISUALIZATION OF FILTERED CONCEPT SPACE (ZHANG ET AL., 2002).....	85
FIGURE 5-2: AN EXAMPLE OF KEYWORD DENSITY	86
FIGURE 5-3: AN ONTOLOGY-BASED DOCUMENT RETRIEVAL (PARALIC AND KOSTIAL, 1999).....	89
FIGURE 5-5: AN EXAMPLE OF CONCEPT CLASSIFICATION IN A DISCUSSION FORUM.....	98
FIGURE 5-6: AN EXAMPLE OF RELATIONSHIPS BETWEEN SUB-CONCEPTS AND CONCEPTS	99
FIGURE 5-7: AN EXAMPLE OF CONCEPT RECLASSIFICATION – BEFORE USER EVALUATION	103
FIGURE 5-8: AN EXAMPLE OF CONCEPT RECLASSIFICATION – AFTER USER FEEDBACK.....	104
FIGURE 6-1: HIGH LEVEL SYSTEM OVERVIEW	107
FIGURE 6-2: A SAMPLE SCREENSHOT OF A BEEHIVE FORUM WEBSITE.....	112
FIGURE 6-3: A SAMPLE SCREENSHOT OF PHPBB FORUM	113
FIGURE 6-4: A SAMPLE SCREENSHOT OF PUNBB FORUM	113

FIGURE 6-5: MODEL-VIEW-CONTROLLER FRAMEWORK (GAMMA ET AL., 1995)	116
FIGURE 7-1: HIGH-LEVEL SYSTEM ARCHITECTURE	124
FIGURE 7-2: SYSTEM ARCHITECTURE FOR PLUG-IN COMPONENT	125
FIGURE 7-3: HIGH-LEVEL SYSTEM ARCHITECTURE FOR DIALOGUE MECHANISM	127
FIGURE 7-4: ARCHITECTURE FOR CONCEPT FORMULATION AGENT	128
FIGURE 7-5: ARCHITECTURE FOR QUERY MONITORING AGENT	129
FIGURE 7-6: WORK FLOW CHART IN DIALOGUE MECHANISM	130
FIGURE 7-7: ARCHITECTURE FOR DISPATCH COMPONENT	132
FIGURE 7-8: SYSTEM ARCHITECTURE FOR FORUM DISPLAY COMPONENT	133
FIGURE 7-9: INFORMATION STRUCTURE AND STORAGE STRUCTURE	134
FIGURE 7-10: ENTITY RELATIONSHIP DIAGRAM	136
FIGURE 7-11: DEPLOYMENT ARCHITECTURE FOR PROTOTYPE SYSTEM	137
FIGURE 7-12: LAYERS OF PROTOTYPE SYSTEM	138
FIGURE 7-13: TOPIC CREATION IN PROTOTYPE DISCUSSION FORUM	140
FIGURE 7-14: VIEWING ALL TOPICS IN PROTOTYPE DISCUSSION FORUM	140
FIGURE 7-15: VIEW MESSAGES IN PROTOTYPE DISCUSSION FORUM	141
FIGURE 7-16: POST A REPLY IN PROTOTYPE DISCUSSION FORUM	141
FIGURE 7-17: ARCHITECTURE FOR INTEGRATION WITH PREVIEW FUNCTION	142
FIGURE 7-18: MESSAGE REPLY PAGE	143
FIGURE 7-19: MESSAGE PREVIEW PAGE	143
FIGURE 7-20: USER EVALUATION FOR RETRIEVED RELEVANT MESSAGES	144
FIGURE 7-21: USER FEEDBACK SUBMITTING	144
FIGURE 10-1: STATISTICS OF THREAD LENGTH	172
FIGURE 10-2: DISTRIBUTION OF TESTING QUESTION LENGTH	173
FIGURE 10-3: THE RELEVANCE RESULTS IN THE FIRST EVALUATION EXPERIMENT	176
FIGURE 10-4: AVERAGE LENGTH OF SUBMITTED POST FOR EACH TASK	182
FIGURE 10-5: AVERAGE RELEVANT POSTS FOUND IN THREE FORUMS FOR THREE TASKS	186
FIGURE 10-6: RELATIONSHIP BETWEEN EFFECTIVENESS RATIO AND AVERAGE LENGTH	193

LIST OF ABBREVIATIONS

ACM	-	Association of Computing Machinery
AI	-	Artificial Intelligence
AWBES	-	Adaptive Web-based Educational Systems
BSD	-	Berkeley Software Distribution License
C4P	-	Content, Conversation, Connections, (information) Context, and Purpose
Equake	-	Electronic Question and Answer Knowledge Environment
ER	-	Entity Relationship
GPL	-	General Public License
HTTP	-	Hypertext Transfer Protocol Overview
IDE	-	Integrated Development Environment
IDF	-	Inverse Document Frequency
ID-B	-	Intelligent Discussion-Bot
IEEE	-	Institute of Electrical & Electronics Engineers
IIS	-	Information Interactive Systems
IR	-	Information Retrieval
IRS	-	Information Retrieval Systems
LSA	-	Latent Semantic Analysis
LSI	-	Latent Semantic Indexing
MVC	-	Model-View-Controller
NPL	-	Natural Language Process
SOAP	-	Simple Object Access Protocol
SVD	-	Single-Value Decomposition

TF	-	Term Frequency
URL	-	Uniform Resource Locator
W3C	-	World Wide Web Consortium
WSDL	-	Web Services Description Language
XML	-	Extensible Markup Language