Case-driven Collaborative Classification

Megan Margaret Vazey

Bachelor of Electrical Engineering (BEng) - First Class Honours,
University of NSW (UNSW), 1993.

Master of Business Administration (MBA) - Executive,
Australian Graduate School of Management (AGSM),
University of NSW (UNSW) and University of Sydney (USyd), 1999.

This thesis is presented for the degree of:

Doctor of Philosophy (PhD)

Submitted January 27, 2007

Revised July 27, 2007

Department of Computing,
Division of Information and Communication Sciences,
Macquarie University, Sydney, Australia.

MACQUARIE
UNIVERSITY ~ SYDNEY

For my Family

and in Loving Memory of

Gladys Hopkins

1916-2005

TABLE OF CONTENTS

TABLE OF CONTENTS I

SYNOPSIS XI

STATEMENT OF CANDIDATE XII

ACKNOWLEDGEMENTS XIII

CHAPTER 1: INTRODUCTION 1

1.1 Chapter Outline 1

1.2 Context 1

1.3 Objectives and Motivation 3

1.4 Research Questions 4

1.5 Significance 5

1.6 Outcomes 7

1.7 Methodology 7

1.8 Thesis Organisation 8

1.9 Chapter Summary 12

CHAPTER 2: THE HTG SUPPORT CENTRE ENVIRONMENT 13

2.1 Chapter Outline 13

2.2 Business Context 14

2.3 Health Check Interviews 14

2.4 Observations 17

2.5 Support Centre Survey 22

2.6 Requirements Analysis 27

2.7 Chapter Summary 28

CHAPTER 3: OTHER KM APPROACHES 30

3.1 Chapter Outline 30

3.2 Data Mining.....31

3.3 Popular Search Technologies 32

3.4 Ontologies.....34

3.5 Knowledge Acquisition and Expert Systems 35

3.6 Chapter Summary 38

CHAPTER 4: RIPPLE DOWN RULES.....40

4.1 Chapter Outline 40

4.2 Introduction to RDR 40

4.3 SCRDR and MCRDR.....41

4.4 Variations on the RDR theme.....60

4.5 Chapter Summary 63

CHAPTER 5: COLLABORATION FEATURES AND TRENDS.....64

5.1 Chapter Outline 64

5.2 Collaboration and Conflict 66

5.3 Collaboratively Generated Anchor Text.....71

5.4 Folksomonies.....72

5.5 Wikipedia 74

5.6 The Semantic Web.....76

5.7 Chapter Summary 77

CHAPTER 6: ADAPTING MCRDR 79

6.1 Chapter Outline 79

6.2 Support Centre Challenges for MCRDR 81

6.3 Supporting Case Change 82

6.4 Supporting Cornerstone Substitution.....83

6.5 Tracking Case drop-throughs 84

6.6 Direct Edit Facility 86

6.7 Supporting Case Construction / Configuration 87

6.8 No Notification of changes to the Knowledge. 88

6.9 Lost links between General and more Specific Classifications.....89

6.10 Chapter Summary 90

CHAPTER 7: A MODEL OF KNOWLEDGE TRANSFER.....92

7.1 Chapter Outline.....92

7.2 An Analysis of Case-driven Knowledge Acquisition.....93

7.3 Review of Past Experimental Data 124

7.4 An Analysis of Rule-driven (RD) Knowledge Acquisition.....127

7.5 Discussion and Implications 128

7.6 Chapter Summary 129

CHAPTER 8: HYBRID CASE AND RULE-DRIVEN KA 133

8.1 Chapter Outline.....133

8.2 Does RDR Require Knowledge Engineering Expertise? 134

8.3 The Call for Hybrid Case And Rule Driven System 140

8.4 Chapter Summary 158

CHAPTER 9: 7CS TOP LEVEL DESIGN 161

9.1 Chapter Outline.....161

9.2 7Cs Condition Mesh 161

9.3 7Cs Data Flow Diagram 164

9.4 Legacy Problem Ticketing System and Knowledge Base 166

9.5 Applicability – Solution Space 166

9.6 Chapter Summary 168

CHAPTER 10: THE DIAL-HOME PROBLEM CONTEXT 169

10.1 Chapter Outline 169

10.2 Problem Scope 169

10.3 Dial-home System Overview 170

10.4 Chapter Summary 174

CHAPTER 11: THE 7CS AND FASTFIX DESIGN CORE 176

11.1 Chapter Outline 176

11.2 Users 179

11.3 RuleNodes and the Rule Mesh 183

11.4 Attributes 197

11.5 Cases 199

11.6 Change Histories 208

11.7 Case-RuleNode Associations 210

11.8 Chapter Summary 221

CHAPTER 12: RESULTS OF THE FASTFIX TRIAL 223

12.1 Chapter Outline 223

12.2 User Activity 223

12.3 Other Statistics 225

12.4 Combined KA Curves for All Users 226

12.5 Effectiveness of the Solution 227

12.6 Case Drop-throughs 228

12.7 Features Employed by Users 231

12.8 Experimental results 231

12.9 Individual KA Curves.....233

12.10 Managing the Complexity for Users.....237

12.11 Collaboration and Conflict 238

12.12 Inadvertent Gross Errors.....239

12.13 Knowledge Engineering Effort.....240

12.14 Chapter Summary 240

CHAPTER 13: DESIGN ENHANCEMENTS.....242

13.1 Chapter Outline.....242

13.2 Reciprocity between the Case and RuleNode Views 242

13.3 Rolling Up RuleNodes.....244

13.4 Shared Child RuleNodes 245

13.5 Enhancing Robustness through Inference 256

13.6 Using Classification Labels 258

13.7 MCRDR and Ontology.....261

13.8 Prudence 263

13.9 Chapter Summary 265

CHAPTER 14: SUMMARY AND CONCLUSIONS.....267

14.1 Chapter Outline.....267

14.2 Thesis Summary 267

14.3 Claims and Caveats 270

14.4 Future Work and Thesis Conclusion 275

GLOSSARY 277

PUBLICATIONS 281

PATENT FILED.....283

AWARDS 283

REFERENCES 284

APPENDIX A – TROUBLE-SHOOTING SURVEY 305

Survey Part A – The Trouble-shooting Environment.....306

Q. Who is solving problems, and how are they doing it?.....306

Q. Is trouble-shooting a team effort, or an individual effort?.....315

Q. What type of resources are used to solve problems?.....325

Q. How successful is HTG at solving customer problems? 334

Assessing CaseDB's Fitness for Purpose 353

Assessing SolutionDB's Fitness for Purpose 358

Survey Part B - First Impressions of FastFIX 377

APPENDIX B – SAMPLE CASEDB CASE 387

APPENDIX C – SYSTEM REQUIREMENTS 390

C.1 Key Performance Indicators 390

C.2 Search and Retrieval 390

C.3 Knowledge Acquisition (KA).....391

C.4 Customer Satisfaction 392

C.5 Employee Satisfaction 392

C.6 Usability.....393

C.7 Local Concordance 393

C.8 Workflow Integration 394

C.9 Feedback and Collaboration 394

C.10 Organisational Learning and Continuous Improvement 394

C.11 Other Requirements 396

APPENDIX D – CONCEPTS FROM THE LITERATURE 398

D.1 Information Retrieval (IR) 398

D.2 Decision Support Systems (DSS) 398

D.3 Knowledge Discovery in Databases (KDD) 399

D.4 Artificial Intelligence (AI) 399

D.5 Machine Learning (ML) 399

D.6 Data Mining (DM) 400

D.7 Popular AI, ML and DM Techniques 401

D.8 Problems with AI, ML and DM Techniques 403

APPENDIX E – TACIT KNOWLEDGE 404

APPENDIX F – ONTOLOGIES 405

APPENDIX G – TRANSCRIPT OF PKS INTERVIEW 408

G.1 Scope 408

G.2 New Clients 408

G.3 History 409

G.4 Size of the knowledge bases 409

G.5 Level of reuse 409

G.6 Comment on IP 410

G.7 Performance of the knowledge bases 410

G.8 Pricing Model 410

G.9 Training 410

G.10 Collaborative Knowledge Building 411

G.11 Quality Standards 411

G.12 Technical features 411

APPENDIX H – ABOUT ANCHOR TEXT 414

APPENDIX I – FOLKSOMONIES 415

I.1 del.icio.us 415

I.2 flickr.com 415

APPENDIX J – WIKIPEDIA VERSUS BRITANNICA 417

APPENDIX K – SEMANTIC WEB LANGUAGES 419

K.1 Resource Description Framework (RDF) 419

K.2 Web Ontology Language (OWL) 419

APPENDIX L – DARPA KNOWLEDGE SHARING 421

L.1 An Interlingua for Knowledge Interchange 421

L.2 The Knowledge Representation System Specification (KRSS) 422

L.3 The Knowledge Query and Manipulation Language (KQML) 422

APPENDIX M – THE FASTFIX PROTOTYPE SHELL 424

M.1 Introducing FastFIX 424

M.2 Account Management 426

M.3 Login 427

M.4 Join / Register 428

M.5 Password Reminder 428

M.6 Home Page 429

M.7 Menu Items 430

M.8 Password Changes 431

M.9 Logout 432

M.10 “Check RuleNodes” 433

M.11 “Check All Cases” 434

M.12 “Check My Cases” 435

M.13 “Write Rule Tree” 435

M.14 “Solutions Preview” 436

M.15 “Upload Files” and “View Uploaded Files” 437

M.16 Available Tests.....438

M.17 Summary 439

APPENDIX N – ACQUIRED RULENODES 440

APPENDIX O – 7CS IMPLEMENTATION ENHANCEMENTS 451

O.1 Different User Profiles.....451

O.2 Improved Case-RuleNode Associations 452

O.3 Entity Relationship Expiry 453

O.4 Case Tracking 454

O.5 RuleNode Approval 456

O.6 Read Only Users 462

O.7 Autonomous Handling of Cases 462

O.8 Undo and Rollback 463

O.9 Lessons from OLAP 463

O.10 Production Subsystems.....464

O.11 Alternate Embodiments 465

O.12 Discussion Forums.....466

O.13 Summary.....466

APPENDIX P – LIMITATIONS OF MCRDR – AN EXAMPLE 468

APPENDIX Q – ROUGH KNOWLEDGE 473

Q.1 Discussion of Transfer Probability 475

Q.2 Probability of Teaching and Learning 477

Q.3 Generalising Case-driven KA 478

Q.4 Location Dependent Errors 479

APPENDIX R – RULENODE RELATIONSHIPS 480

R.1 Containment Relationships 480

R.2 Abstraction Relationships 481

R.3 Semantic Relationships 481

R.4 State-based Relationships 483

R.5 Temporal Relationships 483

R.6 Associated Conclusions 484

R.7 An Example RuleNode 485

SYNOPSIS

The Multiple Classification Ripple Down Rules (MCRDR) knowledge acquisition approach is explored in an environment where knowledge is continuously changing and where multiple agents need to contribute to and learn from the resultant expert system. A collaborative approach to expert system development is proposed in which private individual views and public shared views of the knowledge can coexist, permitting changes in the knowledge to be highlighted, the impact notified to interested parties, and conflicts between agents to be exposed and resolved. By tracking the public and private historical case-RuleNode associations, the proposed approach allows consensus to be built; hence the brittleness of the acquired knowledge can be more rapidly reduced.

The support centre of a major international corporation in the information and communications industry forms a case study in which the nature of trouble-shooting (problem solving) is studied. The research finds that troubleshooting comprises a case-configuration-classification-conclusion cycle. An analogy between collaborative MCRDR and collaborative tagging systems is developed and a stochastic model is derived showing that the trajectories in much of the past machine learning case-driven knowledge acquisition studies can be predicted by the acquisition of classification knowledge on the basis of a random set of incoming repetitive cases, irrespective of the specific case-driven knowledge acquisition approach employed. Further to this, the research highlights the significant amount of domain dependent Knowledge Engineering expertise that can be required by various Ripple Down Rules (RDR) implementations, and proposes that rather than primacy being given to the classification as in conventional rule based systems, or alternatively to the case as in existing RDR systems, primacy should be given to building consensus between collaborating agents in as far as consensus is required to achieve a mature knowledge base.

A collaborative hybrid Case And Rule Driven (CARD) approach to knowledge acquisition known as 7Cs is proposed. The 7Cs approach supports the **C**ollaborative **C**onfiguration and **C**lassification of a stream of incoming problem **C**ases via a set of **C**onditionNodes linked to their **C**lasses and associated **C**onclusions. The approach is trialled through a software prototype system known as *FastFIX*. After less than a day of collective effort the test team had acquired enough knowledge for *FastFIX* to automatically identify and locate solutions to approximately 90% of problems in the selected sub-domain.

STATEMENT OF CANDIDATE

This thesis is submitted in fulfilment of the requirements of the Doctor of Philosophy (PhD) degree at Macquarie University and has not been submitted for a higher degree to any other university or institution. This thesis represents my original work and contributions. Part of the included top-level design and detailed design appears in U.S. Patent application PCT/AU2005/001087.

The sources of information used in the construction of this thesis include references to the literature, observation, interview, survey, and user-driven software-trials. I have also drawn on 15 years of industry experience as an ICT design engineer, software developer, systems integrator, tester and trouble-shooter. Encouraged by Margaret Vazey (a biologist, zoologist and ecologist), my software development work to support the classification of botanical species for the Friends of Malabar Headland[1] (since 1999) provided the initial inspiration for this work.

I certify that to the best of my knowledge, all sources used and assistance received in the preparation of this thesis have been duly acknowledged.

The support centre survey presented in this thesis has been conducted with the approval of Macquarie University Ethics Review Committee HE25FEB2005-D03895.

[1] http://www.malabarheadland.org.au/

ACKNOWLEDGEMENTS

Unending thanks to my husband Terje for love, support and insight. Without your help, this thesis would not have been possible. Thanks also for the terrific discussions regarding the model of knowledge transfer for case-driven knowledge acquisition (KA) and your derivations of Equations (1), (2) and (5).

Many thanks to our three beautiful children Oden, Zac and Jasmin for your patience and pride. You are a gift. Your love is perfect. Thanks to Mum (Margaret) and Dad (Brian) for your inspiration and much appreciated support. Thanks also to Gerd and Kurt for lots and lots of help as and when it was needed. Much thanks to Wendy for much needed help when moving house mid-stream. Thanks to Heidi, Tony, Nicole, Margo, Dawn, Helen, Donna, Rosie and Robyn for cuddling Jassy and caring for Oden and Zac through the preparations; also to Dr. Bose, Dr. Hector Karunajeewa and Dr. John Hopkins for being terrific role models; and to Mr. Andrew Parker for recognition and encouragement.

Thanks to my industry partner for their financial support, for participating in the observation, interview and survey of trouble-shooters in their support centre; and for supporting and participating in the 7Cs[2] / FastFIX[3] software trial. Thanks especially to Stephen Wright, Colin Berrell, Chris Bunting, and Kieran McGee for their receptiveness to new ideas, their nurture of this technology, and their perseverance within the HTG[4] corporate framework. Thanks also to David Law, John Bach, and Bernard Bravo for their participation and welcome feedback. Much thanks also to Julie Gibson for support and encouragement.

Thanks to Warren Bailey and Sean Blasdall at Access Macquarie, and Ben Smith in the Division of Information and Communication Sciences at Macquarie University for their support in pursuing commercialisation opportunities for the 7Cs system. In addition, I very

[2] This research proposes a knowledge representation and knowledge acquisition technique known as 7Cs that supports the **C**ollaborative **C**onfiguration and **C**lassification of a stream of incoming problem **C**ases via a set of **C**onditionNodes linked to their **C**lasses and associated **C**onclusions.

[3] FastFIX is the name of the prototype software developed in this thesis in order to trial the 7Cs design concept.

[4] To protect its confidentiality, the sponsoring company is referred to as High Tech Global (HTG). HTG has sales channels in more than 50 countries and annual revenues of ~$US 10 billion. It is one of the world's 10 largest ICT companies as measured by market capitalisation, and it employs more than 25,000 staff globally.

much appreciated receiving the 2005 Postgraduate Innovation Award from Macquarie University for the collaborative classification software solution presented herein.

Thanks to all the anonymous reviewers for the conferences and journals where my publications were submitted. Your considered review and feedback has helped me make the work so much stronger.

Thanks to Lee Flax at Macquarie University for academic supervision, encouragement, and helpful thesis review. Your "keep going" motto, technical review, and advice with thesis style and presentation have been very much appreciated.

Finally, thanks to Debbie Richards for academic supervision; for supporting my initial APA(I) scholarship application and subsequent upgrade to a PhD and RAACE scholarship; for the loan of various books and references; for assistance, advice, and encouragement with publications; for much appreciated work on the co-authored papers; for supporting my attendance at relevant conferences; for review of the ethics application for the HTG survey; for compilation of the invention disclosure for Access MQ; for review of the PKS patent and editorial support in the creation of the FastFIX / 7Cs patent; for suggestions regarding thesis compilation; and for repeated technical review of various discussions herein. Your assistance and competence in your final review of the thesis prior to examination was very much appreciated.

The first two years of this work was funded via the Australian Research Council Linkage Grant LP0347995 APA(I) scholarship. The final year was funded by Macquarie University via a RAACE scholarship, and by an HTG industry top-up scholarship.