A Computational Model for Task-adapted Knowledge Organisation: Improving Learning through Concept Maps Extracted from Lecture Slides

A thesis submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy by

Atapattu Mudiyanselage Thushari Dilhani Atapattu

Supervised by

Associate Professor Katrina Falkner
Dr Nickolas Falkner
Dr Edward Palmer

School of Computer Science
University of Adelaide
November 2014
Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institute and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in future, be used in a submission for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give my consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University’s digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

..................................................................................

Atapattu Mudiyan selage Thushari Dilhani Atapattu
This thesis is dedicated to my beloved husband
Abstract

This thesis presents a framework for automatically generating concept maps from lecture slides. A concept map is recognised as a valuable educational visualisation tool, which assists students in organising, sharing and representing knowledge. Expert maps (also known as expert concept maps) are prepared by domain experts with the intention to serve as scaffolding to facilitate learning. Automated concept map generation provides an alternative solution to the labour-intensive and time-consuming process of manually constructing expert maps. Therefore, the main objective of this thesis is to develop techniques to extract maps from lecture slides, ensuring that auto-generated concept maps may be utilised as a positive alternative to expert maps. This process is known as concept map mining (CMM).

The particular interest of this thesis is on CMM from lecture slides, due to their wide usage within the teaching context and the poor support of sequentially-structured lecture slides in aiding learners in identifying relationships between information. In general, semantically and syntactically missing and ambiguous text in lecture slides make it undesirable for adopting previously developed algorithms for CMM.

Within this thesis, a set of Natural Language Processing (NLP) algorithms are developed to support concept-relation-concept triple extraction to form concept maps. To support knowledge extraction and to overcome the noise associated with text, this work utilises contextual features specific to lecture slides. The natural layout of the lecture slides is incorporated to organise the extracted triples in a hierarchy. Structural (e.g. co-occurrence, term frequency) and graph-based features (e.g. degree of centrality) are utilised to rank the triples according to their importance within the domain. A series of evaluation studies in this thesis identify promising results, with several case studies demonstrating a strong positive correlation between auto-generated concept maps and human generated maps. These results indicate that this research provides an effective and efficient alternative to expert maps.

Auto-generated concept maps can be utilised to provide scaffolding in the problem solving context, in particular supporting students who are lacking the required skills. Even though this application has been studied previously, these studies do not specifically focus on the relevance of information to learning. To fill this gap, this thesis investigates an approach to provide more relevant concept maps to a given problem. In pursuit of this goal, a framework capable of automatically extracting concept maps according to the given problems (named task-adapted concept maps) is developed, utilising auto-generated concept maps from lecture slides as domain knowledge. In order to investigate the effect of task-adapted concept maps as scaffolding for learning, an evaluation study was undertaken, with students in the task-adapted
concept map scaffolding group demonstrated statistically significant learning gain compared to the students who received lecture slides or full concept maps as scaffolding.
Acknowledgements

I would like to express my heartfelt gratitude to my supervisors Associate Professor Katrina Falkner, Dr Nickolas Falkner and Dr Edward Palmer for the immense support, guidance, valuable advice, encouragement and constructive feedback given throughout the entire time period to complete this research.

I would also like to thank the members of the Computer Science Education Research Group (CSER) for their generous support in making this research project a great success. I am grateful to Dr Amali Weerasinghe for assisting me with in-class experiments and Dr Rebecca Vivian for assisting me to obtain ethics approval for the research. I would also like to thank Mr. Hamid Tarmazadi for assisting me with developing the web-based prototype. I am also grateful to all the HDR colleagues and teaching assistants who helped me with the experiments and all the participants involved in the experiments.

Finally, I take this opportunity to express my sincere gratitude to my beloved parents and husband for their encouragement and support given during my studies.

This research is supported by a Divisional Scholarship of University of Adelaide and Research Training Scheme of Australian Government.
# Table of Contents

List of Tables ......................................................................................................................... iv

List of Figures ......................................................................................................................... vi

List of Publications .................................................................................................................. ix

Chapter 1 Introduction ............................................................................................................. 1

1.1 Concept Map Extraction from Lecture Slides ................................................................. 3

1.2 Task-adapted Concept Maps as Scaffolding ................................................................. 7

1.3 Research Objectives and Questions .............................................................................. 8

1.4 Guide to the Thesis ......................................................................................................... 9

Chapter 2 Concept Mapping: A Review .................................................................................. 11

2.1 Overview of Concept Mapping ...................................................................................... 11

2.2 Theoretical Basis of Concept Mapping .......................................................................... 11

2.3 Applications of Concept Mapping ................................................................................ 13

2.4 Knowledge Organisation versus Text Representations .................................................. 17

2.5 Concept Maps as Scaffolding ....................................................................................... 21

2.6 Summary ....................................................................................................................... 28

Chapter 3 Concept Map Mining: A Review ........................................................................... 29

3.1 Overview of Concept Map Mining ................................................................................ 29

3.2 Concept Map Mining Systems and Methods ................................................................. 30

3.3 Evaluation of Concept Map Mining ................................................................................ 42

3.4 Concept Map Mining Applications in the Educational Context .................................... 45

3.5 Summary ....................................................................................................................... 49

Chapter 4 Concept Map Mining from Lecture Slides ............................................................. 50

4.1 Design of the Concept Map Mining Framework ............................................................. 51

4.1.1 Structured Data Analysis: Design ........................................................................... 54

4.1.2 Knowledge Acquisition: Design ............................................................................. 58

4.1.3 Ranking: Design ..................................................................................................... 61

4.2 Development of the Concept Map Mining Framework .................................................. 64

4.2.1 Development Environment ..................................................................................... 64

4.2.2 Natural Language Annotation ................................................................................ 64
6.1.7 Evaluation of Concept Map Extraction ........................................... 136
6.1.8 Summary ......................................................................................... 143
6.2 Evaluation of the Task-Adapted Scaffolding Framework ...................... 144
  6.2.1 Experimental Design ....................................................................... 144
  6.2.2 Data Analysis .................................................................................. 150
  6.2.3 Results and Discussion ................................................................... 151
  6.2.4 Summary ....................................................................................... 166
Chapter 7 Conclusion ................................................................................. 167
  7.1 Main Contributions ............................................................................. 167
  7.2 Future Directions ............................................................................... 170
  7.3 Concluding Remarks .......................................................................... 172
Appendix A ................................................................................................. 173
Appendix B ................................................................................................. 177
Appendix C ................................................................................................. 180
Appendix D ................................................................................................. 195
References ................................................................................................. 197
List of Tables

Table 2.1: Comparison between concept mapping and PowerPoint (Kinchin, et al., 2008) ............ 20
Table 2.2: Types of instructional scaffolding (Alibali, 2006) .................................................. 22
Table 2.3: Bloom Taxonomy of educational objectives (Bloom, et al., 1956) ............................. 26
Table 2.4: Question classification by (Graesser, et al., 1992) ...................................................... 27
Table 3.1: Comparison of concept map mining systems ............................................................ 40
Table 3.2: Knowledge organisation from lecture slides ............................................................. 41
Table 4.1: Issues associated with the corpus .............................................................................. 53
Table 4.2: Performance of pronoun resolution by Leskovec et al. (2004) .................................... 56
Table 4.3: Examples of demonstrative determiners ................................................................. 57
Table 4.4: Part-of-speech tags (word-level) used in Penn Treebank project .............................. 65
Table 4.5: NLP Annotations of an example sentence ............................................................... 66
Table 4.6: Features of co-occurrence calculation for noise detection in slide-level .................. 68
Table 4.7: Features for pronoun resolution .............................................................................. 69
Table 4.8: Phrase- and clause-level Penn TreeBank tags .......................................................... 72
Table 4.9: Commonly used link types in the Link Grammar parser ........................................ 79
Table 4.10: Features for triple extraction .................................................................................... 81
Table 4.11: Regular expression patterns to identify concepts .................................................... 82
Table 4.12: Hierarchy levels assigned for slide data ................................................................. 84
Table 4.13: Concept importance determined by location ......................................................... 86
Table 4.14: Concept importance determined by grammatical structure ................................... 87
Table 4.15: Elements and attributes of CXL ............................................................................ 90
Table 5.1: Comparison between the components of typical QA system and TASF ................... 100
Table 5.2: Question types defined in QA systems (Dali, et al., 2009; Kolomiyets & Moens, 2011; Olney, et al., 2012) .......................................................... 100
Table 5.3: Question types supported by TASF ....................................................................... 102
Table 5.4: NLP annotations of an example question ............................................................... 106
Table 5.5: Penn Treebank tags for question texts ................................................................. 107
Table 5.6: Important features for task-adapted concept map extraction ................................. 112
Table 6.1: Contingency table .................................................................................................. 123
Table 6.2: Agreement between two evaluators ............................................................... 123
Table 6.3: Evaluation results of concept extraction .......................................................... 126
Table 6.4: Comparison between F-measure and inter-rater agreement ......................... 127
Table 6.5: Evaluation results of noise detection ............................................................... 129
Table 6.6: Statistics of pronouns discovered in the corpus ............................................ 131
Table 6.7: Evaluation results of pronoun resolution ....................................................... 131
Table 6.8: Comparison between F-measure and inter-rater agreement ......................... 134
Table 6.9: Spearman’s ranking correlation ($r_s$) of baseline model ............................... 138
Table 6.10: Spearman’s ranking correlation ($r_s$) of linguistic feature model ............... 139
Table 6.11: Best fit parameter values for structural features ....................................... 140
Table 6.12: Spearman’s ranking correlation ($r_s$) of structural feature model ............... 140
Table 6.13: Visual annotations used by observers ......................................................... 149
Table 6.14: Statistics of student groups ........................................................................ 151
Table 6.15: Students’ prior experience on knowledge organisations techniques .............. 151
Table 6.16: Descriptive statistics of pre-test scores ......................................................... 152
Table 6.17: Descriptive statistics of learning gain ........................................................... 153
Table 6.18: Test of homogeneity of variances (learning gain) ....................................... 154
Table 6.19: Tests of normality ....................................................................................... 154
Table 6.20: Summary results of ANOVA ...................................................................... 155
Table 6.21: Summary of Tukey HSD post-hoc test ......................................................... 156
Table 6.22: Pearson correlation between the time spent on scaffolding and learning gain .... 159
Table 6.23: Pros and cons about concept maps as scaffolding ..................................... 165
Table 6.24: Suggestions for improvements of the system ............................................. 165
Table D.1: What do you think about concept maps used in this study to answer questions? ... 195
Table D.2: What are the issues related to concept maps in this study? ........................... 195
Table D.3: Do you think a tool to automatically generate concept maps from lecture slides will help answering questions? ............................................................. 195
Table D.4: Which form of resources do you prefer if our tool made available through CS Forums in future? ........................................................................... 195
Table D.5: Would you like a tool which can extract partial concept maps to assist answering questions? ........................................................................... 196
Table D.6: What type of courses do you think this kind of tool will be more useful? .......... 196
List of Figures

Figure 1.1: An example concept map ............................................................. 1
Figure 1.2: A sample slide set ........................................................................ 6
Figure 2.1: Distinction between rote and meaningful learning (Novak & Canas, 2006) .... 12
Figure 2.2: Concept map structures a) spoke b) chain c) net (Kinchin, et al., 2000) .......... 14
Figure 2.3: Overview of transformation cycle (Kinchin, et al., 2008) ......................... 18
Figure 2.4: An example of a transformation cycle (Bradley, et al., 2006; Kinchin, et al., 2008) 19
Figure 2.5: Illustrative model of scaffolding ..................................................... 22
Figure 2.6: Concept mapping tasks model (Ruiz-Primo, 2004) .................................. 25
Figure 3.1: Concept map mining process (Villalon & Calvo, 2008) ............................. 29
Figure 3.2: An example for generalisation/specialisation (is-a) relation ....................... 33
Figure 3.3: Concept map activity of Guru (Person, et al., 2012) ............................... 45
Figure 4.1: High-level architecture of the CMMF .............................................. 53
Figure 4.2: Sample slide to illustrate incompleteness of sentences ......................... 55
Figure 4.3: Sample slide to illustrate semantically ambiguous sentences ..................... 55
Figure 4.4: Contextual features embedded in presentation framework ....................... 58
Figure 4.5: (a) Sample lecture slide (b) corresponding concept map generated using CMMF ... 60
Figure 4.6: Term frequency using ‘Operating System’ slide ...................................... 62
Figure 4.7: Degree of co-occurrence using ‘Operating System’ slide ......................... 62
Figure 4.8: An illustration of degree centrality .................................................... 63
Figure 4.9: Typographic information in a sample lecture slide ................................... 63
Figure 4.10: Parse tree of an example sentence .................................................. 66
Figure 4.11: The structure of the Apache POI for Microsoft PowerPoint documents .......... 67
Figure 4.12: An example slide with demonstrative determiners ............................... 71
Figure 4.13: A parse tree of an example NP ‘Generic software process models include’ .... 72
Figure 4.14: Parse tree of an example VP ‘should start with well-understood requirements’ ... 72
Figure 4.15: Sample lecture slide from Software Engineering course .......................... 73
Figure 4.16: Decision tree of CMMF ............................................................... 75
Figure 4.17: Overview of SVO extraction from simple sentences ............................. 75
Figure 4.18: Parse tree for an example sentence with highlighted SVO ..................... 76
Figure 4.19: Triple extraction from an example sentence using (a) baseline algorithm (Rusu, et al., 2007) (b) CMMF ............................................................... 77
Figure 4.20: Parser tree of an example complex sentence ......................................................... 78
Figure 4.21: Linkage diagram of an example complex sentence ................................................. 79
Figure 4.22: Concepts occur in multiple levels of the slides ................................................... 82
Figure 4.23: An example of hierarchy extraction (with ‘relation label’) ..................................... 84
Figure 4.24: An example of hierarchy extraction (without ‘relationship label’) ......................... 84
Figure 4.25: Sample CXL file .................................................................................................. 90
Figure 4.26: An example CourseCMap ..................................................................................... 92
Figure 5.1: Example question 1 (a) text-based answer (b) task-adapted concept map .......... 95
Figure 5.2: Example question 2 (a) text-based answer (b) task-adapted concept map .......... 96
Figure 5.3: High-level architecture of the task-adapted scaffolding framework ....................... 99
Figure 5.4: (a) CourseCMap (b) TSKCMap ............................................................................. 105
Figure 5.5: Parser tree of an example question .......................................................................... 107
Figure 5.6: A skeleton tree used to describe the algorithm 5.1 (Dali et al., 2009) .................... 109
Figure 5.7: parser tree and corresponding triples for (a) question 1 (b) question 2 ............... 109
Figure 5.8: An example of concept map merging ...................................................................... 111
Figure 5.9: Interaction flow diagram ......................................................................................... 115
Figure 5.10: Database design for the web-based prototype .................................................... 117
Figure 5.11: Screen shots of (a) user login (b) quiz (c) scaffolding resource ......................... 118
Figure 5.12: Final overview of the research ............................................................................... 120
Figure 6.1: Design of the evaluation studies ............................................................................. 121
Figure 6.2: Distance between human and computer ranks against percentage of concepts (r_s = 0.813) .......................................................................................... 141
Figure 6.3: Experimental design ............................................................................................... 145
Figure 6.4: Histogram of pre-test scores ................................................................................... 153
Figure 6.5: Normal Q-Q plot of learning gain in each group ..................................................... 155
Figure 6.6: Students’ opinion on having concept maps or lecture slides as scaffolding to answer questions ............................................................................. 160
Figure 6.7: Students’ issues related to concept maps ................................................................. 161
Figure 6.8: Students’ opinion on CMMF ................................................................................. 162
Figure 6.9: Students’ preferred form of scaffolding ................................................................. 162
Figure 6.10: Students’ opinion on task-adapted concept maps as scaffolding to answer questions ................................................................. 163

Figure 6.11: Most suitable Computer Science courses for concept mapping ......................... 164

Figure C.1: Task-adapted concept map as scaffolding for question 1 .................................... 185

Figure C.2: Concept map as scaffolding for question 1 ........................................................ 186

Figure C.3: Concept map with highlighted problem solving context as scaffolding for question 1 ........................................................................... 187

Figure C.4: Task-adapted concept map as scaffolding for question 2 ................................. 188

Figure C.5: Concept map as scaffolding for question 2 ........................................................ 189

Figure C.6: Concept map with highlighted problem solving context as scaffolding for question 2 ........................................................................... 190


