

University of Wollongong
Research Online

University of Wollongong Thesis Collection
1954-2016

University of Wollongong Thesis Collections

2007

Design for conceptual knowledge processing: case studies in applied formal concept analysis

Jon R. Ducrou
jond@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/theses>

University of Wollongong

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following: This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author. Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material.

Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Unless otherwise indicated, the views expressed in this thesis are those of the author and do not necessarily represent the views of the University of Wollongong.

Recommended Citation

Ducrou, Jon R., Design for conceptual knowledge processing: case studies in applied formal concept analysis, Doctor of Philosophy thesis, School of Information Technology and Computer Science, University of Wollongong, 2007. <https://ro.uow.edu.au/theses/760>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

NOTE

This online version of the thesis may have different page formatting and pagination from the paper copy held in the University of Wollongong Library.

UNIVERSITY OF WOLLONGONG

COPYRIGHT WARNING

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

**Design for Conceptual
Knowledge Processing: Case
Studies in Applied Formal
Concept Analysis**

A thesis submitted in fulfilment of the
requirements for the award of the degree

Doctor of Philosophy

from

University of Wollongong

by

Jon Ducrou

Faculty of Informatics

2007

CERTIFICATION

I, Jon Robert Ducrou, declare that this thesis, submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Informatics, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Jon Robert Ducrou
December 4, 2007

CONTENTS

1. <i>Introduction</i>	21
1.1 Approach	22
1.2 Method	23
1.3 Summary of Results	25
1.4 Structure of the Thesis	26
2. <i>Formal Concept Analysis</i>	27
2.1 Definition	27
2.2 Algorithms for FCA	31
2.2.1 Complete Lattice Computation	31
2.2.2 Conceptual Neighbourhood Computation	32
2.2.3 Lattice Diagram Layout	32
2.2.4 Conclusion	39
3. <i>Ten Tasks of CKP</i>	45
3.1 <i>Exploration</i>	46
3.2 <i>Search</i>	47
3.3 <i>Recognition</i>	48
3.4 <i>Identification</i>	48
3.5 <i>Analysis</i>	49
3.6 <i>Investigation</i>	50
3.7 <i>Decision</i>	50
3.8 <i>Improvement</i>	51
3.9 <i>Restructuring</i>	51
3.10 <i>Memorisation</i>	52
3.11 Summary	53
4. MailSleuth	61
4.1 Domain	61
4.2 Knowledge Storage and Access	63
4.2.1 Knowledge Base	64
4.2.2 Attribute Hierarchy	64
4.2.3 Visualisation	66
4.2.4 Usage of the MailSleuth Model	68
4.3 Usability Evaluation	68
4.3.1 Comparative Functionality Review	69
4.3.2 User-based Evaluation	70
4.3.3 Findings and Actions	71
4.3.4 Design Aids for Interpreting Line Diagrams	75
4.4 Evolution	81

4.4.1	MailSleuth 's Evolution	81
4.4.2	MailSleuth and the CKP Tasks	82
5.	SurfMachine	87
5.1	Domain	88
5.1.1	The Mechanics of Beach Selection in Surfing	88
5.1.2	Sources of Data	90
5.2	Approach	93
5.2.1	SurfMachine	93
5.3	Evolution	97
5.3.1	SurfMachine and the CKP Tasks	97
5.3.2	Changes from MailSleuth	101
6.	DSift	103
6.1	Domain	104
6.2	Approach	105
6.2.1	Case Scenario One	108
6.2.2	Case Scenario Two	111
6.3	Evolution	113
6.3.1	DSift and the CKP Tasks	113
6.3.2	Changes from MailSleuth and SurfMachine	114
7.	ImageSleuth	117
7.1	Domain	117
7.1.1	Test Collection	118
7.2	Approach	118
7.2.1	Restriction of the attribute set	121
7.2.2	Moving to upper and lower neighbours	121
7.2.3	Search and Query-By-Example	122
7.2.4	Similarity	122
7.3	Evolution	128
7.3.1	ImageSleuth and the CKP Tasks	129
7.3.2	Changes from MailSleuth , SurfMachine and DSift	130
8.	SearchSleuth	133
8.1	Domain	133
8.2	Approach	136
8.2.1	An Example Interaction	140
8.3	Evolution	142
8.3.1	Changes from MailSleuth , SurfMachine , DSift and ImageSleuth	144
9.	<i>Conclusion</i>	145
9.1	Exposure of Conceptual Structures	145
9.1.1	<i>Lattice View/ Neighbourhood View</i>	146
9.1.2	<i>Conceptual Scales</i>	146
9.1.3	<i>Meaningful Concept Realisation</i>	147
9.1.4	<i>Limited Attribute Vocabulary</i>	148
9.1.5	<i>Intentional Focus</i>	148
9.1.6	<i>Extensional Focus</i>	149

9.1.7	Interface Exposure	149
9.2	Ten Tasks	150
9.3	Conclusion	156
 <i>Appendix</i>		157
A.	<i>Sources</i>	159
A.1	Chapter 2: Formal Concept Analysis	159
A.2	Chapter 4: MailSleuth	159
A.3	Chapter 5: SurfMachine	159
A.4	Chapter 6: DSift	160
A.5	Chapter 7: ImageSleuth	160
A.6	Chapter 8: SearchSleuth	161
B.	SurfMachine Context	163
C.	ImageSleuth v1 Questions	165
C.1	Test Script	165
C.1.1	Part A: <i>Windows Explorer</i>	166
C.1.2	Part B: ImageSleuth	168
C.2	Survey	171
C.2.1	Participant Survey	171

LIST OF FIGURES

1.1	General methodology of design research. Taken from [VK04]	24
2.1	Lattice diagram of the planets context	30
2.2	Pseudocode for the Objects Intersection algorithm	40
2.3	Pseudocode for computing Lower Neighbours	40
2.4	Vector Layout Example	41
2.5	Padded Vector Layout Example	41
2.6	Vector with Layers Layout Example	42
2.7	Force Directed Layout Example	42
2.8	Heuristic Layout Example	43
2.9	Heuristic Vector Layout Example	43
3.1	Carpineto and Romano's <i>CREDO</i> Web Search Application	54
3.2	Cigarran's <i>JBraindead</i> Web Search Application	55
3.3	Becker's <i>Docco</i> File Search Application	56
3.4	Becker's <i>Elba</i> Analysis Preparation Application	57
3.5	Becker's <i>ToscanaJ</i> Analysis Application	58
3.6	Becker's <i>ToscanaJ</i> Analysis Application showing Nesting	59
3.7	Koester's <i>FooCA</i> Web Search Application	60
3.8	Cole, Tilley and Becker's <i>CASS</i> Application	60
4.1	Example of traditional email categorisation	63
4.2	Complete list of index types in MailSleuth	65
4.3	MailSleuth drill down folder and its produced lattice	67
4.4	View of arrow widget used in MailSleuth	70
4.5	Final version of MailSleuth	71
4.6	August 2003 version of MailSleuth	77
4.7	Example of layer shading algorithm producing confusing results	78
4.8	May 2003 version of MailSleuth	79
4.9	Screenshot of HierMail	81
4.10	Comparison of interface evolution in MailSleuth	85
4.11	Screenshot of MailSleuth showing <i>Analysis</i>	86
5.1	Waves optimally break along the beach	89
5.2	Photos of good surfing conditions	92
5.3	Input devices used in SurfMachine (blank)	93
5.4	Input devices used in SurfMachine (with query)	93
5.5	Lattice output of SurfMachine	94
5.6	Example of map system used in SurfMachine	95
5.7	The complete interface view of SurfMachine .	96
5.8	Block Diagram of SurfMachine 's architecture.	97

5.9	Example of <i>analysis</i> in SurfMachine	100
6.1	Concept lattices showing the different types of numeric scaling	106
6.2	Diagram generated by DSift	108
6.3	Case Scenario One: Initial Search	109
6.4	Case Scenario One: Second Search	110
6.5	Case Scenario Two: Initial Search	111
6.6	Case Scenario Two: Second Search	112
6.7	Case Scenario Two: Third Search	113
7.1	Screenshot of ImageSleuth and its conceptual neighbourhood	125
7.2	An example of lattice traversal	126
7.3	Pseudocode representation of search traversal	127
7.4	Results of a concept traversal	128
8.1	Diagram demonstrating the <i>sibling</i> concepts	138
8.2	SearchSleuth: ‘formal concept analysis’	140
8.3	SearchSleuth : ‘tiger’	141
8.4	SearchSleuth : ‘tiger information’	141
8.5	SearchSleuth : ‘tiger os’	141
8.6	SearchSleuth : ‘tiger cat’	142
8.7	SearchSleuth : ‘tiger feature’	142
8.8	SearchSleuth : ‘tiger woods’	142
9.1	Concept lattice corresponding to the context shown in Table 9.1.	151
9.2	Concept lattice corresponding to the context shown in Table 9.2.	152
9.3	Combination of lattices from Fig. 9.1 and Fig. 9.2	154

LIST OF TABLES

2.1	Planets Context	28
4.1	CKP Task breakdown for MailSleuth	62
4.2	Statement ratings from usability testing of MailSleuth	74
5.1	CKP Task breakdown for SurfMachine	87
6.1	CKP Task breakdown for DSift	103
7.1	CKP Task breakdown for ImageSleuth	117
8.1	CKP Task breakdown for SearchSleuth	134
9.1	Breakdown of conceptual structure exposure types	151
9.2	Summary of the primary tasks of CKP	152

LIST OF ABBREVIATIONS

<i>ATC</i>	Access Testing Centre
<i>CASS</i>	Conceptual Analysis of Software Structures
<i>CEM</i>	Conceptual Email Manager
<i>CIS</i>	Conceptual Information System
<i>CKP</i>	Conceptual Knowledge Processing
<i>CLK</i>	Conceptual Landscapes of Knowledge
<i>CSV</i>	Comma Separated Values
<i>DSIFT</i>	Dynamic Simple Intuitive FCA Tool
<i>ECA</i>	Email Concept Analysis
<i>FCA</i>	Formal Concept Analysis
<i>GIS</i>	Geographical Information System
<i>LN</i>	Lower Neighbour
<i>TOSCANA</i>	TOolS for Conceptual ANalysis
<i>UN</i>	Upper Neighbour

ACKNOWLEDGEMENTS

Of course, the most important acknowledgement is to my supervisor and friend, Prof. Peter Eklund. His ability to convey his enthusiasm for the research area made the PhD process exciting rather than tedious.

This thesis would not have been possible without the support of collaborators – from writing papers to designing software together. Dr. Richard Cole and Peter Becker provided entertaining coffee breaks with arguments about the various facets of the latest programming languages. Bastian Wormuth and Björn Vormbrock, travelled from Darmstadt to argue about mathematical foundations, and provided a theoretical landscape for the construction of software. Honours students Shaun Domingo and Timothy Wilson put many more hours than was required into producing and analysing usability tests and surveys.

Other notable influences and supporters include: Dr. Frithjof Dau, always available for coffee and discussion of Peirce; Bjoern Koester whose enthusiasm for a visiting Aussie was greatly appreciated; and Dr. Mark Sifer who always made himself available for coffee. Finally, Dr. Tom Tilley, whose contribution was by far the most important – the project suffix *-sleuth*.

The support of my parents, Bob and Debra Ducrou, made this whole thesis possible. Dad with exasperatingly pedantic spelling and grammar checks and Mum with unconditional support.

Most importantly, I have to acknowledge for the infinite love and support of Amanda Ryan. Without you, I would never have made it this far in life.

'The purpose of computing is insight, not numbers.'

Mathematician Richard Hamming (1915-1998)

ABSTRACT

Conceptual Knowledge Processing (CKP) is a knowledge management and data analysis technique that makes use of conceptual structures. Formal Concept Analysis (FCA) is a CKP methodology that uses lattice theory to represent units of thought, or concepts. When FCA is used in software applications, it makes use of a process called Mixed Initiative. Mixed Initiative breaks down the roles of user and machine, allowing each to play to their strengths. This process allows the computer, which can process vast amounts of data, to produce interaction options from which the user can select. A human can interpret semantic knowledge contained within the data that a computer cannot. This synergy of user and computer allows complex tasks to be performed. Wille [Wil99] proposed ten atomic tasks of CKP which are combined to make these more complex tasks. The ten tasks are *exploration*, *search*, *recognition*, *identification*, *analysis*, *investigation*, *decision*, *improvement*, *restructuring* and *memorisation*. Individually, these tasks represent facets of interaction with conceptual systems.

This thesis uses the ten tasks of Conceptual Knowledge Processing as a framework for experimentation with applications that use Formal Concept Analysis. The applications used for this analysis are **MailSleuth**, **SurfMachine**, **DSift**, **ImageSleuth** and **SearchSleuth**. These applications approach various problems, using FCA as the primary knowledge structure and interaction framework. Each application uses various interface components and varying degrees and types of exposure to the FCA structures on which they are based. The connection between CKP tasks and interface exposure is then explored and reported.