Studies in Computational Intelligence, Volume 225

Editor-in-Chief
Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Vol. 204. Ajith Abraham, Aboul-Ella Hassanien, and André Ponce de Leon F. de Carvalho (Eds.)
Foundations of Computational Intelligence Volume 4, 2009
ISBN 978-3-642-01087-3

Vol. 205. Ajith Abraham, Aboul-Ella Hassanien, and Václav Snášel (Eds.)
Foundations of Computational Intelligence Volume 5, 2009
ISBN 978-3-642-01535-9

Vol. 206. Ajith Abraham, Aboul-Ella Hassanien, André Ponce de Leon F. de Carvalho, and Václav Snášel (Eds.)
Foundations of Computational Intelligence Volume 6, 2009
ISBN 978-3-642-01090-3

Vol. 207. Santo Fortunato, Giuseppe Mangioni, Ronaldo Menezes, and Vincenzo Nicosia (Eds.)
Complex Networks, 2009
ISBN 978-3-642-01205-1

Vol. 208. Roger Lee, Gongzu Hu, and Huaikou Miao (Eds.)
Computer and Information Science 2009, 2009
ISBN 978-3-642-01208-2

Vol. 209. Roger Lee and Naohiro Ishii (Eds.)
Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2009
ISBN 978-3-642-01202-0

Vol. 210. Andrew Lewis, Sanaz Mostaghim, and Marcus Randall (Eds.)
Biologically-Inspired Optimisation Methods, 2009
ISBN 978-3-642-01261-7

Vol. 211. Godfrey C. Onwubolu (Ed.)
Hybrid Self-Organizing Modeling Systems, 2009
ISBN 978-3-642-01529-8

Vol. 212. Viktor M. Kureychik, Sergey P. Malyukov, Vladimir V. Kureychik, and Alexander S. Malyoukov
Genetic Algorithms for Applied CAD Problems, 2009
ISBN 978-3-540-85280-3

Vol. 213. Stefano Cagnoni (Ed.)
Evolutionary Image Analysis and Signal Processing, 2009
ISBN 978-3-642-01635-6

Vol. 214. Been-Chian Chien and Tzung-Pei Hong (Eds.)
Opportunities and Challenges for Next-Generation Applied Intelligence, 2009
ISBN 978-3-540-92813-3

Vol. 215. Habib M. Ammari
Opportunities and Challenges of Connected k-Covered Wireless Sensor Networks, 2009
ISBN 978-3-642-01876-3

Vol. 216. Matthew Taylor
Transfer in Reinforcement Learning Domains, 2009
ISBN 978-3-642-01881-7

Vol. 217. Horia-Nicolai Teodorescu, Junzo Watada, and Lakhmi C. Jain (Eds.)
Intelligent Systems and Technologies, 2009
ISBN 978-3-642-01884-8

Vol. 218. Maria do Carmo Nicoletti and Lakhmi C. Jain (Eds.)
Computational Intelligence Techniques for Bioprocess Modelling, Supervision and Control, 2009
ISBN 978-3-642-01887-9

Vol. 219. Maja Hadzic, Elizabeth Chang, Pornpit Wongthongtham, and Tharam Dillon
Ontology-Based Multi-Agent Systems, 2009
ISBN 978-3-642-01903-6

Vol. 220. Bettina Berendt, Dunja Mladenic, Marco de de Gemmis, Giovanni Semeraro, Myra Spiliopoulou, Gerd Stumme, Vojtech Svatek, and Filip Zelezny (Eds.)
Knowledge Discovery Enhanced with Semantic and Social Information, 2009
ISBN 978-3-642-01890-9

Vol. 221. Tassilo Pellegrini, Sören Auer, Klaus Tochtermann, and Sebastian Schaffert (Eds.)
Networked Knowledge - Networked Media, 2009
ISBN 978-3-642-02183-1

Vol. 222. Elisabeth Rakus-Andersson, Ronald R. Yager, Nikhil Ichalkaranje, and Lakhmi C. Jain (Eds.)
Recent Advances in Decision Making, 2009
ISBN 978-3-642-02186-2

Vol. 223. Zbigniew W. Ras and Agnieszka Dardzinska (Eds.)
Advances in Data Management, 2009
ISBN 978-3-642-02189-3

Vol. 224. Amandeep S. Sidhu and Tharam S. Dillon (Eds.)
Biomedical Data and Applications, 2009
ISBN 978-3-642-02192-3

Vol. 225. Danuta Zakrzewska, Ernestina Menasalvas, and Liliana Byczkowska-Lipinska (Eds.)
Methods and Supporting Technologies for Data Analysis, 2009
ISBN 978-3-642-02195-4

Danuta Zakrzewska, Ernestina Menasalvas,
and Liliana Byczkowska-Lipinska (Eds.)

Methods and Supporting Technologies for Data Analysis

Springer

Danuta Zakrzewska
Institute of Computer Science
Technical University of Lodz
Wolczanska 215
90-924 Lodz
Poland
E-mail: dzakrz@ics.p.lodz.pl

Liliana Byczkowska-Lipinska
Institute of Computer Science
Technical University of Lodz
Wolczanska 215
90-924 Lodz
Poland
E-mail: lilip@ics.p.lodz.pl

Ernestina Menasalvas
Facultad de Informatica
Universidad Politecnica de Madrid
Campus de Montegancedo s/n
28660 Boadilla del Monte Madrid
Spain
E-mail: emenasalvas@fi.upm.es

ISBN 978-3-642-02195-4 e-ISBN 978-3-642-02196-1

DOI 10.1007/978-3-642-02196-1

Studies in Computational Intelligence ISSN 1860-949X

Library of Congress Control Number: Applied for

Typeset & *Cover Design:* Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed in acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

The overwhelming pace of evolution in technology has made it possible to develop intelligent systems which help users in their dayly life activities. Accordingly, methods of recording, managing and analysing data have evolved from the very simple file systems into complex ambient supportive intelligent systems.

This book arises as a compilation of methods, techniques and tools connected with data related issues: from modelling to analysis. A broad range of approaches such as database self-* techniques for ubiquitous environments, multimedia data, or data driven models will be reviewed. Different areas of applications, in which data models conceptualize nowadays reality, starting from e-learning to electric transformers will be considered.

The book is a collection of representative contributions to cover the spectrum related to data bases, which support decision making and data mining methods as well as conceptualization. Datawarehouse technology and modeling are presented in the first chapter together with the deep review of datawarehouse techniques for supporting e-learning processes with special emphasis on data cubes, all the tools are considered in the context of implementation of software application. The second chapter continues with the similar technology and deals with the community data warehouse architecture. Authors propose integrating a parallel query optimizer and grid architecture for query scheduling and optimization. Once database support has been reviewed, in the third chapter authors integrate datawarehouse techniques and data mining methods, for web query categorization, they concentrate on analysing the process and the database support required for its automatization. Subsequently, the deep survey of Fuzzy and Rough Sets methods for data analysis is presented in the chapter entitled Applications of Fuzzy and Rough Set Theory in Data Mining. Continuing with data analysis methods, in the fifth chapter, the author deals with the important topic of user driven modeling techniques. Data mining techniques, in particular clustering and sequential pattern mining are used for analysing navigational behaviour and learning styles of users. The following chapter refers to multimedia databases

with focus on the techniques for image retrieval. After presenting supporting standards and solutions such as SQL/MM, Oracle and MPEG-7, the authors discuss using frequency domain techniques for content based image retrieval in multimedia databases. The last part of the book is more related to technology underlying database support. In the chapter entitled Query Relaxation in Cooperative Query Processing the authors deal with query mechanisms that enable to formulate meaningful queries by relaxing query constraints, they use semantic, structural and topological query relaxation. Special focus is done on similarities aspects and methodologies of obtained results evaluations. On the other hand, techniques for self-adaptation and interoperabolity are presented in the chapter entitled Ensuring Mobile Databases Interoperability in Ad Hoc Configurable Environments: A Plug-and-play Approach. Authors propose fully distributed agent-based architecture for building mobile database communities, taking into account physical mobility of hosts and logical mobility of database queries. The books ends by the presentation of the industry application, where data base is built according to requirements of expert system for monitoring and diagnosing in large power transformers.

We expect, that the book will be of interest for students, researchers and practicionners in any field related to development of intelligent systems, in which data bases and data analysis play crucial role. The editors would like to thank the contributors of the book for their effort in preparing the chapters. We would also like to give special thanks to Editor of the Serie prof. Janusz Kacprzyk for his help and his support and to prof. Mykhaylo Yatsymirskyy, for the book inspiration. Thanks to all the people who help in completing the book. We hope that our work will contribute to the evolution of databases research in the development of ubiquituos intelligent systems

March, 2009

Lodz, Danuta Zakrzewska

Madrid, Ernestina Menasalvas

Lodz, Liliana Byczkowska-Lipińska

Contents

Data Warehouse Technology for E-Learning 1
Marta E. Zorrilla
1 Business Intelligence Overview 1
2 Why a Data Warehouse for E-Learning? 2
3 E-Learning Data Warehouse Architecture 4
4 E-Learning Data Warehouse Development Lifecycle 6
4.1 Business Requirements 7
4.2 Dimensional Modelling 9
4.3 ETL Processes 12
4.4 OLAP Cubes 14
4.5 User Interface 15
5 Summary 17
References 19

Optimizer and Scheduling for the Community Data Warehouse Architecture 21
Rogério Luís de Carvalho Costa, Ricardo Antunes, and Pedro Furtado
1 Introduction 21
2 Background 23
2.1 Query Optimization 25
2.2 Query Scheduling in Grids 27
3 Grid-DWPA Architecture 30
4 Generic Query Optimizer for DWPA 31
4.1 GAP Concepts 32
4.2 GAP Algorithm 34
4.3 Contributions to Parallel Optimization 37
4.4 GAP Results 38
5 Scheduling in Grid Data Warehouses 44
5.1 Data Placement and Availability 45
5.2 Dynamic Query Scheduling in Grid-DWPA 47

5.3 Experimental Evaluation 48
6 Conclusions 51
References 52

Database Support for Automatic Web Queries Categorization 57
Ernestina Menasalvas Ruiz and Santiago Eibe Garcia
1 Introduction 57
2 Related Work 58
2.1 Preliminaries 58
2.2 Query Understanding 60
2.3 Temporal Profiles of Queries 61
3 Database Support for Web Query Categorization 62
3.1 The Process of Categorization 62
3.2 Taxonomies of Concepts 65
3.3 Visibility of Words in Queries 65
3.4 The Database Support 66
4 Conclusion and Outline 68
References 69

Applications of Fuzzy and Rough Set Theory in Data Mining 71
Dan Li and Jitender S. Deogun
1 Data Mining 72
1.1 Data Mining and Related Fields 73
1.2 Data Mining Functionalities 74
2 Fuzzy Set Theory and Rough Set Thoery 76
2.1 An Introduction to Fuzzy Set Theory 76
2.2 An Introduction to Rough Set Theory 77
3 Integrating Fuzzy and Rough Set Theory into Data Mining 78
3.1 APPLICATION 1: Fuzzy Bayesian Classification for Anomaly Detection 80
3.2 APPLICATION 2: Fuzzy K-Nearest Neighbor Classification for Gene Function Prediction 84
3.3 APPLICATION 3: Rough-Fuzzy K-Means Clustering for Missing Data Imputation 93
4 Summary 109
References 109

Data Driven Users' Modeling 115
Danuta Zakrzewska
1 Introduction 115
2 Modeling according to Navigational Patterns 117
2.1 Log Files 117
2.2 Finding Navigational Patterns 119
2.3 Sequential Pattern Mining 120
2.4 Application of OLAP Operations 121

2.5 Clustering 122
2.6 Evaluation 124
3 Modeling according to Learning Styles 128
3.1 Learning Styles' Data 128
3.2 Cluster Analysis 129
3.3 Dealing with Outliers 129
3.4 Experiments 131
4 Summary 133
References 134

Frequency Domain Methods for Content-Based Image Retrieval in Multimedia Databases 137
Bartłomiej Stasiak and Mykhaylo Yatsymirskyy
1 Introduction 137
2 The Basics of Automatic Image Retrieval and Classification 138
3 Solutions and Standards Supporting Content-Based Image Retrieval 140
3.1 SQL Multimedia and Application Packages (SQL/MM) 141
3.2 Oracle 9i/10g interMedia 142
3.3 MPEG-7 144
4 Time Domain Methods – Overview 146
5 Frequency Domain Methods 148
5.1 Inspiration 149
5.2 Boundary-Based Shape Descriptors 150
5.3 Texture Descriptors 153
5.4 Fourier-Mellin Transform and Its Application to Content-Based Image Retrieval 155
6 Summary 161
References 162

Query Relaxation in Cooperative Query Processing 167
Arianna D'Ulizia, Fernando Ferri, and Patrizia Grifoni
1 Introduction 167
2 Semantic Query Relaxation 169
3 Structural Query Relaxation 174
4 Topological Query Relaxation 177
5 Conclusion 182
References 183

Ensuring Mobile Databases Interoperability in Ad Hoc Configurable Environments: A Plug-and-Play Approach 187
Angelo Brayner, José de Aguiar Moraes Filho, Maristela Holanda, Eriko Werbet, and Sergio Fialho
1 Introduction 188
2 Mobile Computing Environment Model 189

3 Sharing Mobile Databases in MDBCs 190
4 Processing Queries over Mobile Databases 192
4.1 Query Engine 192
4.2 The Cost Model 196
4.3 Adaptive Features 197
4.4 Processing Queries with Unavailable Data Sources 206
5 Processing Mobile Transactions 207
5.1 ITS: A Self-adaptable Scheduler 207
5.2 Ensuring Self-adaptability 209
5.3 Experiments 211
6 Conclusions 215
References 216

Database Architecture of Diagnostic System for Large Power Transformers 219
Liliana Byczkowska-Lipińska and Agnieszka Wosiak
1 Introduction 219
2 Relevant Research 220
3 The Architecture of the Diagnosing System 221
4 Database Architecture 225
4.1 Traditional Data Storage Techniques 226
4.2 Relational Database Management Systems 227
4.3 Multiple Databases 228
4.4 Tables' Partitioning 228
4.5 The Diagnostic System's Database Architecture . . . 229
5 Monitoring System Interface 232
6 Data Compression Techniques 233
6.1 Run Length Encoding Algorithm Bases 234
6.2 Signal Compression Using Huffman Algorithm 234
6.3 Signal Compression Using LZW Algorithm 234
6.4 Data Compression Used in the Diagnostic System . 235
7 Conclusions 236
References 237

Subject Index 239

List of Contributors

Ricardo Antunes
University of Coimbra
rantunes@dei.uc.pt

Angelo Brayner
University of Fortaleza (UNIFOR)
brayner@unifor.br

Liliana Byczkowska - Lipińska
Technical University of Lodz
lilip@ics.p.lodz.pl

Rogério Luís de Carvalho Costa
University of Coimbra
rogcosta@dei.uc.pt

Jitender S. Deogun
University of Nebraska-Lincoln
deogun@cse.unl.edu

Santiago Eibe Garcia
Universidad Politecnica de Madrid
seibe@fi.upm.es

Fernando Ferri
National Research Council Italy
fernando.ferri@irpps.cnr.it

Sergio Fialho
Federal University of Rio Grande do Norte
fialho@pop-rn.rnp.br

José de Aguiar Moraes Filho
University of Fortaleza (UNIFOR)
jaguiar@unifor.br

Pedro Furtado
University of Coimbra
pnf@dei.uc.pt

Patrizia Grifoni
National Research Council Italy
patrizia.grifoni@irpps.cnr.it

Maristela Holanda
Federal University of Rio Grande do Norte
mholanda@dca.ufrn.br

Dan Li
Northern Arizona University
Dan.Li@nau.edu

Ernestina Menasalvas Ruiz
Universidad Politecnica de Madrid
emensalvas@fi.upm.es

Bartłomiej Stasiak
Technical University of Lodz
basta@ics.p.lodz.pl

Arianna D'Ulizia
National Research Council Italy
arianna.dulizia@irpps.cnr.it

Eriko Werbet
University of Fortaleza
(UNIFOR)
eriko@unifor.br

Agnieszka Wosiak
Technical University of Lodz
agnieszka@ics.p.lodz.pl

Mykhaylo Yatsymirskyy
Technical University of Lodz
jacym@ics.p.lodz.pl

Danuta Zakrzewska
Technical University of Lodz
dzakrz@ics.p.lodz.pl

Marta E. Zorrilla
University of Cantabria
marta.zorrila@unican.es