
DATA MINING AND DECISION SUPPORT

Integration and Collaboration

**The Kluwer International Series
in Engineering and Computer Science**

DATA MINING AND DECISION SUPPORT

Integration and Collaboration

Edited by

Dunja Mladenić, Nada Lavrač,
Marko Bohanec, and Steve Moyle



Springer Science+Business Media, LLC

Library of Congress Cataloging-in-Publication Data

Data mining and decision support : integration and collaboration / edited by Dunja Mladenic ... [et al.].

p. cm. -- (The Kluwer international series in engineering and computer science ; SECS 745)

Includes bibliographical references and index.

ISBN 978-1-4613-5004-0 ISBN 978-1-4615-0286-9 (eBook)

DOI 10.1007/978-1-4615-0286-9

1. Data mining. 2. Decision support systems. I. Mladenic, Dunja, 1967- II. Series.

QA76.9.D343D376 2003

006.3--dc21

Copyright © 2003 Springer Science+Business Media New York

Originally published by Kluwer Academic Publishers in 2003

Softcover reprint of the hardcover 1st edition 2003

All rights reserved. No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without the written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Permission for books published in Europe: permissions@wkap.nl

Permissions for books published in the United States of America: permissions@wkap.com

Printed on acid-free paper.

SolEuNet is a registered trademark of Alarix d.o.o. Ljubljana. Used with permission.

Contents

Preface	ix
Acknowledgments	xi
Foreword	xiii
Contributing Authors	xvii

PART I BASIC TECHNOLOGIES	1
Edited by Dunja Mladenić	

1.	DATA MINING	3
	Nada Lavrač and Marko Grobelnik	
2.	TEXT AND WEB MINING	15
	Dunja Mladenić and Marko Grobelnik	
3.	DECISION SUPPORT	23
	Marko Bohanec	
4.	INTEGRATION OF DATA MINING AND DECISION SUPPORT	37
	Nada Lavrač and Marko Bohanec	
5.	COLLABORATION IN A DATA MINING VIRTUAL ORGANIZATION	49
	Steve Moyle, Jane McKenzie, and Alípio Jorge	
6.	DATA MINING PROCESSES AND COLLABORATION PRINCIPLES	63
	Alípio Jorge, Steve Moyle, Hendrik Blockeel, and Angi Voß	

PART II INTEGRATION ASPECTS OF DATA MINING AND DECISION SUPPORT **79**

Edited by Marko Bohanec

7. **DECISION SUPPORT FOR DATA MINING: AN INTRODUCTION
TO ROC ANALYSIS AND ITS APPLICATIONS** **81**
Peter Flach, Hendrik Blockeel, Cèsar Ferri, José Hernández-Orallo,
and Jan Struyf
8. **DATA MINING FOR DECISION SUPPORT: SUPPORTING
MARKETING DECISIONS THROUGH SUBGROUP DISCOVERY** **91**
Bojan Cestnik, Nada Lavrač, Peter Flach, Dragan Gamberger, and
Mihael Kline
9. **PREPROCESSING FOR DATA MINING AND DECISION
SUPPORT** **107**
Olga Štěpánková, Petr Aubrecht, Zdeněk Kouba, and Petr
Mikšovský
10. **DATA MINING AND DECISION SUPPORT INTEGRATION
THROUGH THE PREDICTIVE MODEL MARKUP
LANGUAGE STANDARD AND VISUALIZATION** **119**
Dietrich Wettschereck, Alípio Jorge, and Steve Moyle

PART III APPLICATIONS OF DATA MINING AND DECISION SUPPORT **131**

Edited by Steve Moyle

11. **ANALYSIS OF SLOVENIAN MEDIA SPACE** **133**
Marko Grobelnik, Maja Škrjanc, and Darko Zupanič
12. **ON THE ROAD TO KNOWLEDGE: MINING 21 YEARS OF UK
TRAFFIC ACCIDENT REPORTS** **143**
Peter Flach, Hendrik Blockeel, Thomas Gärtner, Marko Grobelnik,
Branko Kavšek, Martin Kejkula, Darek Krzywania, Nada Lavrač,
Peter Ljubič, Dunja Mladenić, Steve Moyle, Stefan Raeymaekers,
Jan Rauch, Simon Rawles, Rita Ribeiro, Gert Sclap, Jan Struyf,
Ljupčo Todorovski, Luis Torgo, Dietrich Wettschereck, and
Shaomin Wu
13. **ANALYSIS OF A DATABASE OF RESEARCH PROJECTS
USING TEXT MINING AND LINK ANALYSIS** **157**
Marko Grobelnik and Dunja Mladenić
14. **WEB SITE ACCESS ANALYSIS FOR A NATIONAL
STATISTICAL AGENCY** **167**
Alípio Jorge, Mário A. Alves, Marko Grobelnik, Dunja Mladenić,
and Johann Petrak
15. **FIVE DECISION SUPPORT APPLICATIONS** **177**
Marko Bohanec, Vladislav Rajkovič, and Bojan Cestnik

16.	LARGE AND TALL BUILDINGS: A CASE STUDY IN THE APPLICATION OF DECISION SUPPORT AND DATA MINING	191
	Steve Moyle, Marko Bohanec, and Eric Osrowski	
17.	A COMBINED DATA MINING AND DECISION SUPPORT APPROACH TO EDUCATIONAL PLANNING	203
	Silvana Gasar, Marko Bohanec, and Vladislav Rajkovič	
 PART IV COLLABORATION ASPECTS		213
	Edited by Nada Lavrač	
18.	COLLABORATIVE DATA MINING WITH RAMSYS AND SUMATRA TT: PREDICTION OF RESOURCES FOR A HEALTH FARM	215
	Olga Štěpánková, Jiří Kléma, and Petr Mikšovský	
19.	COLLABORATIVE DECISION MAKING: AN ENVIRONMENTAL CASE STUDY	227
	Tanja Urbančič, Marko Bohanec, and Branko Kontić	
20.	LESSONS LEARNED FROM DATA MINING, DECISION SUPPORT AND COLLABORATION	237
	Dunja Mladenić, Nada Lavrač, and Marko Bohanec	
21.	INTERNET SUPPORT TO COLLABORATION: A KNOWLEDGE MANAGEMENT AND ORGANIZATIONAL MEMORY VIEW	247
	Alípio Jorge, Damjan Bojadžiev, Mário Amado Alves, Olga Štěpánková, Dunja Mladenić, Jiří Palouš, Peter Flach, and Johann Petrak	
22.	MIND THE GAP: ACADEMIA-BUSINESS PARTNERSHIP MODELS AND E-COLLABORATION LESSONS LEARNED	261
	Nada Lavrač and Tanja Urbančič	
 Subject index		271

Preface

The main aim of this book is to present a framework, methods, and tools for the integration of data mining and decision support, as well as their application to business problems in a collaborative setting.

Data mining is concerned with solving problems by analyzing existing data. It is the core of the knowledge discovery process, which aims at the extraction of interesting, non-trivial, implicit, previously unknown and potentially useful information from data. It is an interdisciplinary area involving databases, machine learning, pattern recognition, statistics, visualization, and others.

Decision support is concerned with developing systems aimed at helping decision makers solve problems and make decisions. Decision support provides a selection of data analysis, simulation, visualization and modeling techniques, and software tools such as decision support systems, group decision support and mediation systems, expert systems, databases and data warehouses.

Data mining and decision support are, each on their own, well-developed research areas, but until now there has been no systematic attempt to integrate them. The bridging of these two technologies has a significant impact on the developments of both fields, largely by improving approaches for problem solving in real settings, enabling the fusion of knowledge from experts and knowledge extracted from data, and consequently enabling the successful solution of new types of problems. The chapters presented in this book provide a foundation for this new research area at the intersection of data mining and decision support. This is achieved by proposing a conceptual framework and methods for their integration and by discussing the collaboration aspects and the lessons learned from practical applications.

The book consists of four parts: Part I: Basic Technologies; Part II: Integration Aspects of Data Mining and Decision Support; Part III: Applications of Data Mining and Decision Support; Part IV: Collaboration Aspects.

Part I gives an overview of the technologies used in practical data analysis and decision-making problem solving, briefly describing each of the technologies, selected methods and applications. The first two chapters cover *data mining* and the specific area of *text and web mining*. The next two chapters describe *decision support* itself and the proposed technology for *data mining and decision support integration*. Collaboration in a data mining virtual organization and technology for *collaborative data mining* are addressed in the last two chapters of Part I.

Part II addresses the aspects of data mining and decision support integration. There are four chapters, presenting four possible integration approaches. The first

chapter presents the *decision support for data mining* approach, which is aimed at improving the data mining process by decision support methods, in particular decision support based on the ROC analysis. The second chapter addresses the complementary *data mining for decision support* approach, illustrated by two methods applied in marketing. The final two chapters of Part II describe the integration of data mining and decision support in *data pre-processing*, and through *model standardization and visualization*.

Part III describes in detail real-life applications of data mining and decision support. These applications are drawn from the following domains: mass media, road traffic accidents, collaborating research organizations, Web site logs, loan allocation, building construction, and education. The first two chapters in the part report on pure data mining applications. The second two chapters describe text and Web mining applications. The final three chapters report on decision support applications, two of which include the integration of decision support and data mining techniques.

Part IV presents practical advantages and limitations of collaborative problem solving framework in a virtual enterprise formed from remote teams collaborating mostly via the Internet. The first two chapters of this part report on collaborative data mining and collaborative decision-making. The third chapter addresses the knowledge management aspects of collaboration. The last two chapters report on the lessons learned from the collaboration of academic and business partners in the virtual enterprise framework.

How to read this book

The book is structured so that it gradually introduces the main concepts and methods covering the whole area of data mining, decision support, their integration and collaboration aspects, all illustrated with a number of applications. However, the book allows also for selective reading of specific areas in the manner which follows.

A reader interested in *data mining* only is suggested to read Chapters 1 and 2 of Part I and followed by reading about the real-life applications of data mining in Chapters 11 and 12 of Part III.

A reader interested exclusively in *decision support* is suggested to read Chapter 3 of Part I and then read about the applications of decision support in Chapter 15 of Part III.

Those interested in the *integration* aspects of data mining and decision support are suggested to read Chapters 1-4 of Part I, followed by reading about the approaches to data mining and decision support integration in all Part II chapters, the applications of integrated data mining and decision support technology in the last two chapters of Part III, and the lessons learned in the last two chapters of Part IV.

A reader interested in the aspects of *collaboration* is suggested to read Chapters 5 and 6 of Part I and then read about different collaboration aspects in Part IV.

Editors:

Dunja Mladenić, Nada Lavrač, Marko Bohanec, and Steve Moyle

Ljubljana and Oxford

Acknowledgments

The work reported in this book is a result of the research project '*Data Mining and Decision Support for Business Competitiveness: A European Virtual Enterprise*' (SolEuNet IST-11495, 2000-2003), enabling twelve academic and business teams to form a virtual enterprise aimed at the development of practical data mining and decision-making solutions for clients. The successful remote collaboration of partners in this project was funded by the Commission of European Communities.

We are grateful to the project officer Ralf Hansen and project reviewers Luis Camarinha-Matos and Ann Macintosh for setting high scientific standards, which were achieved with the assistance of their gentle guidance and invaluable support. Ralf Hansen is 'responsible' for suggesting that the consortium address the novel research area of data mining and decision support integration which may have remained untouched without his explicit request. Luis Camarinha-Matos has guided us on a tour of – for us – the unknown world of virtual enterprises and organizations, and Ann Macintosh made us aware of the difficulties of practical business solutions, providing guidance in focusing the project activities towards the main project goals of integration, collaboration, education, and business issues. In our work, we were supported also by the Advisory Board, whose main members were Maarten van Someren, David Pearce, Jane McKenzie, Jörg-Uwe Kietz and Sarabjot Singh Anand, each contributing expertise, critical project assessment and guidance.

We are grateful also to Damjan Demšar for the technical support in producing this manuscript.

The work described in this book was supported also by the following national research grants:

- The Slovenian Ministry of Education, Science and Sport,
- The FEDER and Programa de Financiamento Plurianual de Unidades de I&D (FCT), Portugal,
- Fund for Scientific Research of Flanders, Belgium, and
- The Ministry of Education, Youth and Sports of the Czech republic.

The authors of individual chapters would like to thank:

- Chapter 10: A. and G. Andrienko, AIS, FhG, Sankt Augustin, Germany for developing some of the visualizations presented in Chapter 10,

- Chapter 12: John Bullas of Hampshire County Council for his enthusiasm and guidance throughout the project. John Bullas acknowledges support from the AA Foundation for Road Safety Research and the CSS (formerly the County Surveyors Society),
- Chapter 16: Knowledge Management Forum at Henley Management College UK for bringing them together,
- Chapter 18: The company, Lauryn, s.r.o., for providing the data for the SPA project, and Štěpán Lauryn who was very helpful in all phases of the project,
- Chapter 22: Matjaž Maček for his consultancy concerning the principal-agent theory and Andrej Orel for the collaboration in studying IPR issues and virtual enterprise models.

Foreword

The area of data mining is concerned with the discovery of interesting and useful patterns in data. Most scientific research in data mining is aimed at developing better methods that can find more types of patterns in noisy and incomplete data and at measures of interestingness and usefulness. Over the past years the research efforts have been quite successful, resulting in a wide range of methods and commercial tools. Industrial development of data mining involves the development of tools for visualisation of data and the resulting patterns, standards for representing data and patterns and business aspects such as project planning and feasibility assessment. In practice data mining is becoming an established technology with applications in a wide range of areas that include marketing, health care, finance, environment, economic planning, career planning, and military.

The ultimate purpose of industrial data mining is the *use* of the resulting patterns to solve some problem: predicting sales or customer loyalty, recognising disease or predicting the effect of medical intervention, recognising fraudulent financial transactions, recognising threats to ecological systems, etc. In such applications, data mining is only part of the total process. Experts and clients are needed in most stages of the project: to define and redefine the problem, to determine relevant aspects of the problem, to supply the data, to remove errors from the data, to provide constraints on possible patterns, to interpret patterns and possibly reject those that are implausible, to evaluate predicted effects, and so on.

A complementary approach to such problem solving that does not rely on collecting observational data is decision making. In this approach the human decision maker makes the alternatives and the criteria for preferring one to the other explicit and then uses this to make a rational decision. This process can be supported by computational decision support systems that systematically guide a decision maker through the process.

The vision behind this book is the leverage that can be obtained by integrating data mining and decision support systems. For example, in medical decisions, data mining can produce a pattern that can be used to predict the effect of possible treatments. Adding this information to the information pool will allow better decisions than when only the observations on the patient are available. Decision support systems can then be used to make a rational choice between alternative treatments.

This integration of these two paradigms introduces a number of issues that were not of central importance in data mining or decision support systems. Decision

support systems are based on information from the decision maker and possibly from additional experts and decision makers. To achieve integrated data mining / decision making requires including the state-of-the art data mining methods and tools but also a methodology and support for cleaning and preparing the data and standards for representing intermediate results.

Several practice-oriented research efforts in data mining have recognised the need for methods and tools that include a larger part of the problem solving process than data analysis. A methodology that covers the process from problem definition to presentation and delivery of the resulting patterns was developed in the CRISP-DM project (Chapman, et al., 2000) and is becoming a de facto industrial standard. The MiningMart project (Morik and Scholz, 2003) developed methods and tools that include the preprocessing stage and support the construction and use of a database of solutions. The Japanese Active Mining project (Motoda, 2002) extends the scope to the active acquisition of data and emphasises the role of a domain expert in all stages of the data mining process. The US project on Evidence Extraction and Link Discovery addresses similar issues but in the context of specific applications such as military decision making and discovering terrorist networks.

The vision of this book, which is a result of the European SolEuNet project 'Data Mining and Decision Support for Business Competitiveness: A European Virtual Enterprise' (Mladeníć and Lavrač, 2003), is to take the scope of the methods and tools one step further, covering the business and collaboration aspects, as well as the decision making stage of the problem solving process. The latter raises a number of new problems. One type of problem is that domain experts and decision makers must be able to participate in the process. This requires a methodology and tools that make the process transparent.

The chapters in this book contribute steps towards integrated, collaborative data mining and decision making methods, tools and methodology. They also include overviews of the basic technologies and industrial cases.

The contributions in this book are thus of interest to researchers and practitioners in decision support and in data mining. Integrated use of these two technologies is likely to increase their usability and to raise additional research questions. Graduate students specializing in data mining and decision support will benefit from increased awareness of some new, non-conventional methods, which proved useful in the practical applications described in the book. The work on collaborative problem solving at this high technical level provides interesting tools and observations that are worth reading as they go beyond the areas of data mining and decision support. The industrial case studies illustrate the practical value of the approach. In this respect, the book will be of interest also to researchers, practitioners and students of knowledge management. The virtual enterprise organizational models and testimonies of practical collaboration experiences of business and academic institutions will certainly be illuminating for the reader. Finally, some explicit 'lessons learned' are worth considering when entering into collaborative projects involving business and academia.

References

- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide, CRISP-DM consortium, <http://www.crisp-dm.org>
- Mladenić, D. and Lavrač, N. (eds.), (2003). *Data Mining and Decision Support for Business Competitiveness: A European Virtual Enterprise, Final Report*, <http://soleunet.ijs.si>.
- Morik, K. and Scholz, M. (2003). The MiningMart Approach to Knowledge Discovery in Databases, In (eds. Zhong, N. and Liu, J.), *Handbook of Intelligent IT*, IOS Press.
- Motoda, H. (ed.) (2002), *Active Mining: New Directions of Data Mining*, IOS Press.

Contributing Authors

Mário Amado **Alves** (maa@liacc.up.pt) is a researcher at the Laboratory of artificial intelligence and computer science (LIACC) of the University of Porto, Portugal. He works on adaptive hypertext and data mining, and teaches artificial intelligence and software engineering at the Open University of Portugal (Universidade Aberta). He has worked on natural language processing and computational logic at the Center for artificial intelligence of the New University of Lisbon, and has published in the areas of data mining, software engineering and natural language processing.

Petr **Aubrecht** (aubrech@fel.cvut.cz) is a research fellow of the Gerstner Laboratory at the Czech Technical University in Prague. His research interests include data preprocessing, data warehousing, data mining, and knowledge engineering. He is the designer of SumatraTT, and coauthor of articles about data preprocessing for data mining and data warehousing.

Hendrik **Blockeel** (hendrik.blockeel@cs.kuleuven.ac.be) is a fellow of the Fund for Scientific Research of Flanders, and assistant professor at the Department of Computer Science of the Katholieke Universiteit Leuven, Belgium. His research interests include (relational) data mining, machine learning and inductive logic programming, with applications in various scientific domains. He has co-authored over 50 publications in these areas.

Marko **Bohanec** (marko.bohanec@ijs.si) is a senior researcher at Jožef Stefan Institute, Department of Intelligent Systems, Ljubljana, and assistant professor in information systems at the University of Ljubljana, School of Public Administration. His research interests relate to decision support systems and data mining, in particular to qualitative hierarchical modeling and machine learning. His publications include papers in the journals Machine Learning, Information & Management, Artificial Intelligence, IEEE Intelligent Systems, and International Journal of Medical Informatics.

Damjan **Bojadžiev** (damjan.bojadziev@ijs.si) is a researcher at the Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia, working on self-referential aspects of logic and computation. His publications include papers on (computational) logic in the Journal of Logic Programming, Artificial Intelligence,

Journal of Consciousness Studies and a chapter in a handbook of Phenomenology and Cognitive Science.

Bojan **Cestnik** (bojan.cestnik@temida.si) is the general manager of the software company Temida and a researcher at the Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia. His professional and research interests include knowledge based information systems, decision support systems and machine learning. He has been responsible for several large-scale software development and maintenance projects.

Cèsar **Ferri** (cferri@dsic.upv.es) is assistant professor at the Department of Information Systems and Computation, Technical University of Valencia, Spain. His research interests cover machine learning, cost-sensitive learning, relational data mining, and declarative programming.

Peter **Flach** (Peter.Flach@bristol.ac.uk) is Reader in Machine Learning at the Department of Computer Science at the University of Bristol, UK. His research interests include machine learning, inductive logic programming and multirelational data mining, intelligent reasoning, and first-order probabilistic models. He is the author of *Simply Logical: Intelligent Reasoning by Example* (John Wiley, 1994) and co-editor of *Abductive and Inductive Reasoning* (Kluwer Academic Publishers, 2000).

Dragan **Gamberger** (dragan.gamberger@irb.hr) is Head of the Laboratory for Information Systems at the Rudjer Bosković Institute, Zagreb, Croatia. His research interests include machine learning, knowledge discovery, and intelligent data analysis. He has worked on various induction-based data analysis applications in medical and marketing domains, and has set up the Internet Data Mining Server.

Thomas **Gärtner** (thomas.gaertner@ais.fraunhofer.de) is a PhD candidate at the Fraunhofer Institute for Autonomous Intelligent Systems at Sankt Augustin and a research assistant at the Universities of Bristol and Bonn. He received a Master of Science from the University of Bristol and a Diploma from the University of Cooperative Education, Mannheim. His main research interests are kernel methods and learning from structured data such as graphs, higher-order terms, and relational data. He has published several papers at international conferences, including the International Conference on Machine Learning and the International Conference on Inductive Logic Programming.

Silvana **Gasar** (silvana.gasar@telesat.si) holds masters degrees in psychology and computer science. Her interests and work cover human information processing, human sexuality, education, e-communication, machine learning, data mining, and decision making from both the human and the computer standpoint.

Marko **Grobelnik** (marko.grobelnik@ijs.si) works at the Department of Intelligent Systems of Jožef Stefan Institute, Ljubljana, Slovenia. His research interests are the study and development of data, text and Web mining techniques and their applications, especially learning from large text data sets. He has worked on several

national and international projects in text and data mining, Internet archiving, intelligent agents and strategic game theory.

José **Hernández-Orallo** (jorallo@dsic.upv.es) is Senior Lecturer at the Department of Information Systems and Computation, Technical University of Valencia, Spain. His research interests center on the areas of artificial intelligence, machine learning, data mining and software engineering. His publications include several books and book chapters in these areas.

Alípio **Jorge** (amjorge@liacc.up.pt) is a member of the Artificial Intelligence and Computer Science Laboratory and a Lecturer at the Faculty of Economics of the University of Porto, Portugal. His research covers (collaborative) data mining, including association and classification rules, and Web mining. He has organized several international conferences and other events in artificial intelligence and data mining. He is the coordinator of the data analysis and decision support systems masters program at the University of Porto, and vice-president of the Portuguese AI Society.

Martin **Kejkula** (kejkula@vse.cz) is a member of the Laboratory for Intelligent systems (LISp) in Prague, Czech Republic. He works on knowledge discovery from databases, and studies at the Department of Information and Knowledge Engineering at the University of Economics in Prague.

Jiří **Kléma** (klema@labe.felk.cvut.cz) is assistant professor at the Department of Cybernetics, Czech Technical University, Prague. His research interests include machine learning, data mining, intelligent diagnostics, decision support, and case based reasoning.

Mihael **Kline** (kline-kline@siol.net) is assistant professor of consumer psychology, integrated marketing communications and basics of visual communication at the Faculty of social sciences of the University of Ljubljana, Slovenia. He is also the general manager and creative director of the marketing agency Kline & Partner. He has been a management consultant to Slovenian companies in marketing, management and communications skills. He is the author of a book, four monographs and over 50 professional articles.

Branko **Kavšek** (branko.kavsek@ijs.si) is a member of the Intelligent Data Analysis and Computational Linguistics Research Group at Jožef Stefan Institute, Ljubljana, Slovenia. His research covers machine learning, inductive databases, and subgroup discovery.

Branko **Kontič** (branko.kontic@ijs.si) is an adviser on environmental impact assessment for industry, regulators and Slovenian government institutions. He has been engaged in preparing methodologies for strategic environmental assessment, developing the strategy of radioactive waste management, environmental modeling, and risk analysis. His publications include book chapters on environmental protection and predictive health risk assessment.

Zdeněk Kouba (kouba@labe.felk.cvut.cz) is Head of the Information and Knowledge-based Systems group of the Gerstner laboratory at the Czech Technical University, Prague. He teaches information systems design and decision support systems. His research interests include database systems, uncertainty processing, object-oriented design and programming, and formal design of software systems. He is the (co-)author of more than 50 publications in these fields.

Darek Krzywania (darek.krzywania@cs.kuleuven.ac.be) is a student in the Department of Computer Science, Katholieke Universiteit Leuven, Belgium, in the research group Declarative Languages and Artificial Intelligence. His research interests mainly focus on privacy issues in distributed data mining.

Nada Lavrač (nada.lavrac@ijs.si) is Head of the Intelligent Data Analysis and Computational Linguistics Research Group at Jožef Stefan Institute, Ljubljana, Slovenia. Her research interests include machine learning, relational data mining, inductive logic programming, and intelligent data analysis in medicine. She is a co-author and co-editor of several books, including *Inductive Logic Programming: Techniques and Applications*, Ellis Horwood, 1994, and co-editor of the book *Relational Data Mining*, Springer, 2001.

Peter Ljubič (peter.ljubic@ijs.si) works at the Intelligent Data Analysis and Computational Linguistics Research Group at Jožef Stefan Institute, Ljubljana, Slovenia. His research interests include machine learning, data mining, and equation discovery.

Jane McKenzie (jane.mckenzie@henleymc.ac.uk) is a Fellow of the Chartered Institute of Certified Accountants and Director of Studies of MBA programmes in London and Henley, UK. She is the author of two books on business and knowledge management. Her research covers the effects of information technologies on learning, business and management, and the emergence of virtual organizations.

Petr Mikšovský (miksovsp@labe.felk.cvut.cz) is a researcher at the Gerstner Laboratory for Intelligent Decision Making and Control, Czech Technical University, Prague. His research activities are oriented towards knowledge-based and information systems.

Dunja Mladenić (Dunja.Mladenic@ijs.si) works in the Department of Intelligent Systems of the Jožef Stefan Institute, Ljubljana, Slovenia. She has been a visiting researcher and faculty member at the School of Computer Science, Carnegie Mellon University, Pittsburgh, USA. Her research includes the study and development of machine learning and data mining techniques and their application to real-world problems, especially learning from text and the Web.

Steve Moyle (steve.moyle@comlab.ox.ac.uk) is a researcher at the Oxford University Computing Laboratory, United Kingdom. His interests include the application of data mining to real-world problems, inductive logic programming, and relational data mining.

Eric **Ostrowski** (eric.ostrowski@echarris.com) is a Partner of EC Harris, and leads knowledge development within the organization. He has significant experience in knowledge management techniques and is actively involved in external specialist research in the area of knowledge management. His interests include the innovative use of information technologies to combat the problems of information overload.

Jiří **Palouš** (palous@labe.felk.cvut.cz) is a student at the Department of Cybernetics at the Czech Technical University, Prague. His research interests include machine learning and case-based reasoning.

Johann **Petrak** (johann@ai.univie.ac.at) is a researcher at the Austrian Research Institute for Artificial Intelligence, Vienna. His research interests are machine learning and data mining.

Vladislav **Rajkovič** (vladislav.rajkovic@ijs.si) is professor of information systems and Head of the Laboratory for Decision Processes and Knowledge-based Systems at the Faculty of Organisational Sciences, University of Maribor, Slovenia. He is also a research fellow at the Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia. His research focuses on information systems and artificial intelligence methods for supporting decision and educational processes.

Stefan **Raeymaekers** (stefan.raeymaekers@cs.kuleuven.ac.be) is a research assistant in the Machine Learning Research Group at the Katholieke Universiteit Leuven, Belgium. His research interests include machine learning, relational data mining and web mining.

Jan **Rauch** (rauch@vse.cz) is Head of the EuroMISE centrum Cardio department and member of the Department of Information and Knowledge Engineering at the University of Economics, Prague, Czech Republic. His research interests include theoretical foundations and applications of data mining. He is a coeditor of the Proceedings of the conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-99), Springer 1999.

Simon **Rawles** (rawles@cs.bris.ac.uk) is a student at the Machine Learning Group at the University of Bristol, United Kingdom. His research is concerned with the use of object orientation in knowledge representation for multi-relational data mining. He is also the author of the SINUS system for learning from structured data through propositionalisation.

Rita **Ribeiro** (rita@liacc.up.pt) is a research assistant in the Artificial Intelligence and Data Analysis Group at the University of Porto, Portugal. Her main interests focus on machine learning and data mining.

Gert **Sclep** (Gert.Sclep@rega.kuleuven.ac.be) is a student at the Laboratory of Medicinal Chemistry at the Rega Institute in Leuven, Belgium. His research interests

include structure based drug design, structure prediction of biopolymers and machine learning.

Maarten van Someren (maarten@swi.psy.uva.nl) is Lecturer in Artificial Intelligence at the University of Amsterdam. His research interests include machine learning, data mining, methods for adaptive interactive systems, information extraction and models of human learning.

Jan Struyf (Jan.Struyf@cs.kuleuven.ac.be) is a research assistant at the Fund for Scientific Research of Flanders and a student at the Declarative Languages and Artificial Intelligence Laboratory of the Katholieke Universiteit Leuven, Belgium. His research interests include machine learning, relational data mining, especially techniques for improving scalability and efficiency in relational data mining systems.

Maja Škrjanc (maja.skrjanc@ijs.si) is a member of the Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia. Her research focuses on machine learning and data mining, with applications to linguistic rule discovery and business problems.

Olga Štěpánková (step@labe.felk.cvut.cz) is Vice-head of the Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, and president of the Czech Society for Cybernetics and Informatics. Her research has been focused mainly on domains where mathematical logic meets artificial intelligence: machine learning, inductive logic programming and multi-agent systems. She has co-authored six textbooks in Czech, and co-edited three volumes published by Springer.

Ljupčo Todorovski (Ljupco.Todorovski@ijs.si) is a researcher in the Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia. His research covers machine learning, equation discovery, computational scientific discovery and automated modeling of dynamic systems. He has co-edited a book on Computational Discovery of Communicable Knowledge, Springer, 2003.

Luis Torgo (ltorgo@liacc.up.pt) is a member of the Artificial Intelligence and Computer Science Laboratory of the University of Porto, Portugal, and assistant professor at the Faculty of Economics. His research interests include machine learning and data mining, focusing on numerical prediction in financial and ecological time series. He has authored the regression tool RT, which won a prize in an international data analysis competition on ecological modeling. He has published several research papers on regression analysis and participated in several consortium research projects.

Tanja Urbančič (tanja.urbancic@ijs.si) is Head of the Center for Knowledge Transfer in Information Technologies, Jožef Stefan Institute, Ljubljana, Slovenia, and Head of the School of Economics and Engineering at Nova Gorica Polytechnic, Slovenia. Her research interests cover knowledge management and transfer, and artificial intelligence methods for modeling human skills.

Angi **Voß** (angelika.voss@ais.fraunhofer.de) is Head of the Department for Mediation Systems at the Fraunhofer Institute for Autonomous Intelligent Systems in Sankt Augustin, Germany. With her interdisciplinary team she develops software and methods for planning, conducting and moderating e-discourses. The solutions are applied to public participation processes, cooperative spatial planning projects and sustainable communication processes for decision-making, problem solving and knowledge sharing.

Dietrich **Wettschereck** (dw@comp.rgu.ac.uk) is a Senior Research Fellow at the Robert Gordon University, Aberdeen, Scotland. His research interests include machine learning (especially distance-based algorithms) and data mining (especially standards and visualization of data mining results).

Shaomin **Wu** (shaomin@cs.bris.ac.uk) is a Research Fellow in the Machine Learning Group in the Department of Computer Science at the University of Bristol, UK. His research interests include machine learning, reliability analysis, and statistical process control.

Darko **Zupanič** (darko.zupanic@ijs.si) is associated with the Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia. His research covers areas of inductive logic programming and constraint satisfaction systems.