Preface

This volume contains the papers presented at the 8th International Conference on Discovery Science (DS 2005) held in Singapore, Republic of Singapore, during the days from 8–11 of October 2005.

The main objective of the Discovery Science (DS) conference series is to provide an open forum for intensive discussions and the exchange of new ideas and information among researchers working in the area of automating scientific discovery or working on tools for supporting the human process of discovery in science. It has been a successful arrangement in the past to co-locate the DS conference with the International Conference on Algorithmic Learning Theory (ALT). This combination of ALT and DS allows for a comprehensive treatment of the whole range, from theoretical investigations to practical applications. Continuing in this tradition, DS 2005 was co-located with the 16th ALT conference (ALT 2005). The proceedings of ALT 2005 were published as a twin volume 3734 of the LNCS series.

The International Steering Committee of the Discovery Science conference series provided important advice on a number of issues during the planning of Discovery Science 2005. The members of the Steering Committee are Hiroshi Motoda, (Osaka University), Alberto Apostolico (Purdue University), Setsuo Arikawa (Kyushu University), Achim Hoffmann (University of New South Wales), Klaus P. Jantke (DFKI and FIT Leipzig, Germany), Massimo Melucci (University of Padua), Masahiko Sato (Kyoto University), Ayumi Shinohara (Tohoku University), Einoshin Suzuki (Yokohama National University), and Thomas Zeugmann (Hokkaido University).

We received 112 full paper submissions out of which 21 long papers (up to 15 pages), 7 regular papers (up to 9 pages), and 9 project reports (3 pages) were accepted for presentation and are published in this volume. Each submission was reviewed by at least two members of the Program Committee of international experts in the field. The selection was made after careful evaluation of each paper based on originality, technical quality, relevance to the field of discovery science, and clarity.

The Discovery Science 2005 conference had three types of presentations: long papers were presented in a plenary session; regular papers were presented in a short spotlight presentation to generate interest and a presentation during a poster session for intensive discussions and presentation of details; project reports were presented in a poster session to allow intensive discussion on ongoing work and interesting ideas that had not been developed to the same degree of maturity as long and regular papers.

The Carl Smith Award was presented this year for the first time in honor of Professor Carl Smith to the student author of the best paper in the Discovery Science conference authored or co-authored by a student. The prize of 555 Euro
was awarded to Qianjun Xu for the paper entitled *Active Constrained Clustering by Examining Spectral Eigenvectors*.

This volume consists of four parts. The first part contains invited talks of ALT 2005 and DS 2005. Since the talks were shared between the two conferences, for the speakers invited specifically for ALT 2005 only abstracts are contained in this volume, while the full paper is found in the twin volume LNCS 3734 (the proceedings of ALT 2005). We were delighted that Gary Bradshaw (Invention and Artificial Intelligence), Vasant Honovar (Algorithms and Software for Collaborative Discovery from Autonomous, Semantically Heterogeneous, Distributed, Information Sources), Chih-Jen Lin (Optimization Issues in Training Support Vector Machines), Ross D. King (The Robot Scientist Project), and Neil Smalheiser (The Arrowsmith Project: 2005 Status Report) followed our invitation to present their work.

The second part of this volume contains the papers accepted as long papers (acceptance rate of less than 21%). The third part of this volume contains the regular papers, which were found to belong to the best 27% of all submissions. Finally, the fourth part of this volume contains the project reports; the total acceptance rate for all three paper categories sums to 37% of all submissions.

We are deeply indebted to the Program Committee members as well as their subreferees who had the critically important role of reviewing the submitted papers and contributing to the intense discussions which resulted in the selection of the papers published in this volume. Without this enormous effort, ensuring the high quality of the work presented at Discovery Science 2005 would not have been possible.

We also thank all the authors who submitted their work to Discovery Science 2005 for their efforts.

We wish to express our gratitude to the invited speakers for their acceptance of the invitation and their stimulating contributions to the conference.

Finally, we wish to thank everyone who contributed to make Discovery Science 2005 a success: the DS Steering committee, the ALT conference chairs, invited speakers, and last but not least Lee Wee Sun, the Local Arrangements Chair and his team of supporters.

October 2005

Achim Hoffmann
Hiroshi Motoda
Tobias Scheffer
Organization

Conference Chair

Hiroshi Motoda Osaka University, Japan

Program Committee

Achim Hoffmann University of New South Wales, Sydney, Australia (Chair)
Tobias Scheffer Humboldt-Universität zu Berlin, Germany (Chair)
Jose Luis Balcázar University of Catalunya, Spain
Elisa Bertino Purdue University, USA
Wray Buntine Helsinki Institute of Information Technology, Finland
Vincent Corruble University of Pierre et Marie Curie, France
Manoranjan Dash Nanyang Technological University, Singapore
Andreas Dress Max Planck Institute for Mathematics in the Sciences, Germany
Sašo Džeroski Jožef Stefan Institute, Slovenia
Eibe Frank Tampere University of Technology, Finland
Johannes Fürnkranz Technical University of Darmstadt, Germany
João Gama University of Porto, Portugal
Gunter Grieser Technical University of Darmstadt, Germany
Fabrice Guillet Ecole Polytechnique of the University of Nantes, France
Mohand-Said Hacid University of Claude Bernard, Lyon, France
Udo Hahn Jena University, Germany
Tu Bao Ho, JAIST, Japan
Klaus P. Jantke FIT Leipzig, Germany
Szymon Jaroszewicz Technical University of Szczecin, Poland
Kristian Kersting Universität Freiburg, Germany
Ross King University of Wales, UK
Kevin Korb, Monash University, Melbourne, Australia
Ramamohanarao Kotagiri University of Melbourne, Australia
Stefan Kramer TU München, Germany
Nicolas Lachiche Univ. Strasbourg, France
Nada Lavrač Jožef-Stefan Institute, Ljubljana, Slovenia
Aleksandar Lazarević United Technologies Research Center, CT, USA
Jinyan Li Institute for Infocomm Research, Singapore
VIII Organization

Ashesh Mahidadia  University of New South Wales, Sydney, Australia
Michael May  Fraunhofer Institute for Autonomous Intelligent Systems, Germany
Katharina Morik  University of Dortmund, Germany
Ion Muslea  Language Weaver, USA
Lourdes Peña  Center for Intelligent Systems at the ITESM, Mexico
Bernhard Pfahringer  University of Waikato, New Zealand
Jan Rauch  University of Economics, Czech Republic
Domenico Saccà  University of Calabria and ICAR-CNR, Italy
Rudy Setiono  National University of Singapore, Singapore
Myra Spiliopoulou  Otto-von-Guericke University, Germany
Ashwin Srinivasan  IBM India, India
Einoshin Suzuki  Yokohama National University, Japan
Masayuki Takeda  Kyushu University, Japan
Kai Ming Ting  Monash University, Australia
Ljupčo Todorovski  Jožef Stefan Institute, Slovenia
Volker Tresp  Siemens AG, München, Germany
Alfonso Valencia  National Centre for Biotechnology, Spain
David Vogel  AI Insight, USA
Gerhard Widmer  Johannes-Kepler-Universität, Austria
Akihiro Yamamoto  Kyoto University, Japan
Mohammed Zaki  Rensselaer Polytechnic Institute, USA
Chengqi Zhang  University of Technology Sydney, Australia
Djamel A. Zighed  University of Lumière, France

Local Arrangements

Lee Wee Sun  National University of Singapore, Republic of Singapore
External Reviewers

Mohammed Al Hasan  Lucas Hope
Alexandre Aussem  Daisuke Ikeda
Hideo Bannai  Branko Kavsek
Maurice Bernadet  Gaelle Legrand
Steffen Bickel  Lee Wee Sun
Remco Bouckaert  Remi Lehn
Agnès Braud  Peter Ljubic
Ulf Brefeld  Chuan Lu
Michael Brückner  Carlo Mastroianni
Robert D. Burbidge  Igor Nai-Fovino
Mario Cannataro  Luigi Palopoli
Narendra S. Chaudhari  Esa Pitkänen
Vineet Chaoji  Dragoljub Pokrajac
Maria Luisa Damiani  Lothar Richter
Marko Debeljak  Ulrich Rückert
Damjan Demšar  Martin Scholz
Isabel Drost  Alexander K. Seewald
Timm Euler  Zujun Shentu
Tanja Euler  Giandomenico Spezzano
Feng Gao  Shyh Wei Teng
Gemma C. Garriga  Evimaria Terzi
Rémy Gaudin  Nguyen Truong Thang
Vivekanand Gopalkrishnan  Julien Thomas
Andrea Gualtieri  Salvatore Vitabile
Pietro H. Guzzo  Michael Wurst
Mounira Harzallah  Shipeng Yu
Kohei Hatano  Bernard Ženko
Phan Xuan Hieu

Sponsoring Institutions

We wish to thank the Air Force Office of Scientific Research and the Asian Office of Aerospace Research and Development for their contribution to the success of this conference.

AFOSR/AOARD support is not intended to express or imply endorsement by the U.S. Federal Government.
The Carl Smith Award

Starting with this year, the “Carl Smith Award” is presented to the most outstanding paper written or co-authored by a student. The selection is made by the actual program committee of the Discovery Science conference. The award carries a scholarship prize of 555 Euro.

The decision to introduce this award has been proposed at the ALT/DS-business meeting of last year’s conference in Padua after remembering Carl Smith, who passed away on July 21, 2004 after a long and valiant battle with cancer, with a minute of silence. Subsequently, this decision has been happily approved by Patricia Smith.

Carl performed his undergraduate studies in Vermont and received his Bachelor of Science Degree from the University of Vermont in 1972. Then, he moved to State University of New York at Buffalo where he received his Ph.D. Subsequently, he was Assistant Professor of Computer Science at Purdue University. Then he was at the University of Maryland at College Park, where he got promoted to Associate and Full Professor. In 1993, Carl received the Habilitation degree from the University of Latvia in Riga. He is also one of the very few non-Latvian scientists who got elected to the Latvian Academy of Science.

Additionally, Carl spent several years as program manager at the National Science Foundation’s theoretical computer science program and continued to work for the National Science Foundation by working on programs and panel reviews for many years.

Carl also contributed to the computer science community as an editor of the International Journal of the Foundations of Computer Science, Theoretical Computer Science, and Fundamenta Informaticae.

The Discovery Science conference series is still a young one but many researchers remember Carl for a much longer time, because of his very active role in the algorithmic or computational learning communities.

Let us look back to 1986 when the 1st International Workshop on Analogical and Inductive Inference was held in Wendisch-Rietz near Berlin. This was the starting point of the first international conference series on learning theory which merged in 1994 with the Algorithmic Learning Theory series established in 1990. At this workshop Carl gave a talk “On the Inference of Sequence of Functions” (co-authored with Bill Gasarch) in which he developed a model of “learning how to learn.” Of course, by this time Carl was already well known through his work on comparison of identification criteria for machine inductive inference, his work on team learning, and the beautiful survey paper “Inductive Inference: Theory and Methods” (co-authored with Dana Angluin).

Besides the very fruitful scientific discussions we all enjoyed at this workshop, it was also the beginning or continuation of a lasting friendship many of us had with Carl which in turn led to many teams including Carl all over
the world. These long and fruitful collaborations included leading groups from Japan, Latvia, Germany, USA, Australia, and Singapore among many countries. As a result, papers on query learning, on memory limitation, on learning with anomalies, on the complexity of inductive inference, on Barzdins’s conjecture, on procrastination, on mind change complexity as well as on a logic of discovery emerged.

Besides his regular papers, Carl contributed in many ways to the ALT and DS conference series by serving for their Program Committees and the DS Steering Committee, and by serving as local chair, as conference chair and arrangements as invited speaker.

He also chaired IFIP WG 1.4 on Computational Learning Theory and organized many funding to support, in particular, young scientists.

Since Carl Smith did so much for the ALT and DS conferences, his spirit, his contributions, his passion, and his ideas will be remembered and passed to the young generations by the “Carl Smith Award.”

August 2005

Thomas Zeugmann
# Table of Contents

## Invited Papers

Invention and Artificial Intelligence
   *Gary Bradshaw* ............................................. 1

Algorithms and Software for Collaborative Discovery from Autonomous, Semantically Heterogeneous, Distributed Information Sources
   *Doina Caragea, Jun Zhang, Jie Bao, Jyotishman Pathak,*
   *Vasant Honavar* ............................................ 14

Training Support Vector Machines via SMO-Type Decomposition Methods
   *Pai-Hsuen Chen, Rong-En Fan, Chih-Jen Lin* ................. 15

The Robot Scientist Project
   *Ross D. King, Michael Young, Amanda J. Clare,*
   *Kenneth E. Whelan, Jem Rowland* .......................... 16

The Arrowsmith Project: 2005 Status Report
   *Neil R. Smalheiser* ........................................... 26

## Regular Contributions - Long Papers

Practical Algorithms for Pattern Based Linear Regression
   *Hideo Bannai, Kohei Hatano, Shunsuke Inenaga,*
   *Masayuki Takeda* .......................................... 44

Named Entity Recognition for the Indonesian Language: Combining Contextual, Morphological and Part-of-Speech Features into a Knowledge Engineering Approach
   *Indra Budi, Stéphane Bressan, Gatot Wahyudi, Zainal A. Hasibuan,*
   *Bobby A.A. Nazief* ......................................... 57

Bias Management of Bayesian Network Classifiers
   *Gladys Castillo, João Gama* .................................. 70

A Bare Bones Approach to Literature-Based Discovery: An Analysis of the Raynaud’s/Fish-Oil and Migraine-Magnesium Discoveries in Semantic Space
   *Richard J. Cole, Peter D. Bruza* ............................ 84
Assisting Scientific Discovery with an Adaptive Problem Solver

Christopher Dartnell, Jean Sallantin ........................................ 99

Cross-Language Mining for Acronyms and Their Completions from the Web

Udo Hahn, Philipp Daumke, Stefan Schulz, Kornél Markó .............. 113

Mining Frequent $\delta$-Free Patterns in Large Databases

Céline Hébert, Bruno Crémilleux ............................................. 124

An Experiment with Association Rules and Classification: Post-Bagging and Conviction

Alípio M. Jorge, Paulo J. Azevedo ........................................... 137

Movement Analysis of Medaka (Oryzias Latipes) for an Insecticide Using Decision Tree

Sengtai Lee, Jeehoon Kim, Jae-Yeon Baek, Man-Wi Han,
Chang Woo Ji, Tae-Soo Chon ............................................... 150

Support Vector Inductive Logic Programming

Stephen Muggleton, Huma Lodhi, Ata Amini,
Michael J.E. Sternberg ...................................................... 163

Measuring Over-Generalization in the Minimal Multiple Generalizations of Biosequences

Yen Kaow Ng, Hirotaka Ono, Takeshi Shinohara ......................... 176

The $q$-Gram Distance for Ordered Unlabeled Trees

Nobuhito Ohkura, Kouichi Hirata, Tetsuji Kuboyama,
Masateru Harao .............................................................. 189

Monotone Classification by Function Decomposition

Viara Popova, Jan C. Bioch .................................................. 203

Learning On-Line Classification via Decorrelated LMS Algorithm:
Application to Brain-Computer Interfaces

Shiliang Sun, Changshui Zhang ............................................... 215

An Algorithm for Mining Implicit Itemset Pairs Based on Differences of Correlations

Tsuyoshi Taniguchi, Makoto Haraguchi ................................... 227

Pattern Classification via Single Spheres

Jigang Wang, Predrag Neskovic, Leon N. Cooper ....................... 241
SCALETRACK: A System to Discover Dynamic Law Equations Containing Hidden States and Chaos
   Takashi Washio, Fuminori Adachi, Hiroshi Motoda .................. 253

Exploring Predicate-Argument Relations for Named Entity Recognition in the Molecular Biology Domain
   Tuangthong Wattarujeekrit, Nigel Collier .......................... 267

Massive Biomedical Term Discovery
   Joachim Wermter, Udo Hahn ......................................... 281

Active Constrained Clustering by Examining Spectral Eigenvectors
   Qianjun Xu, Marie desJardins, Kiri L. Wagstaff .................... 294

Learning Ontology-Aware Classifiers
   Jun Zhang, Doina Caragea, Vasant Honavar .......................... 308

Regular Contributions - Regular Papers

Automatic Extraction of Proteins and Their Interactions from Biological Text
   Kiho Hong, Junhyung Park, Jihoon Yang, Eunok Paek ............ 322

A Data Analysis Approach for Evaluating the Behavior of Interestingness Measures
   Xuan-Hiep Huynh, Fabrice Guillet, Henri Briand ..................... 330

Unit Volume Based Distributed Clustering Using Probabilistic Mixture Model
   Keunjoon Lee, Jinu Joo, Jihoon Yang, Sungyong Park ............. 338

Finding Significant Web Pages with Lower Ranks by Pseudo-Clique Search
   Yoshiaki Okubo, Makoto Haraguchi, Bin Shi ....................... 346

CLASSIC’CL: An Integrated ILP System
   Christian Stolle, Andreas Karwath, Luc De Raedt .................. 354

Detecting and Revising Misclassifications Using ILP
   Masaki Yokoyama, Tohgoroh Matsui, Hayato Ohwada ............... 363

Project Reports

Self-generation of Control Rules Using Hierarchical and Nonhierarchical Clustering for Coagulant Control of Water Treatment Plants
   Hyeon Bae, Sungshin Kim, Yejin Kim, Chang-Won Kim ............. 371
A Semantic Enrichment of Data Tables Applied to Food Risk Assessment
Hélène Gagliardi, Ollivier Haemmerlé, Nathalie Pernelle, Fatiha Saïs ................................. 374

Knowledge Discovery Through Composited Visualization, Navigation and Retrieval
Wei-Ching Lim, Chien-Sing Lee .................................................. 377

A Tabu Clustering Method with DHB Operation and Mergence and Partition Operation
Yongguo Liu, Dong Zheng, Shiqun Li, Libin Wang, Kefei Chen .................. 380

Discovering User Preferences by Using Time Entries in Click-Through Data to Improve Search Engine Results
Parthasarathy Ramachandran .................................................. 383

Network Boosting for BCI Applications
Shijun Wang, Zhonglin Lin, Changshui Zhang ................................. 386

Rule-Based FCM: A Relational Mapping Model
Ying Yang, Tao-shen Li, Jia-jin Le .............................................. 389

Effective Classifier Pruning with Rule Information
Xiaolong Zhang, Mingjian Luo, Daoying Pi ................................. 392

Text Mining for Clinical Chinese Herbal Medical Knowledge Discovery
Xuezhong Zhou, Baoyan Liu, Zhaohui Wu ................................. 396

Author Index ................................................................. 399