

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

University of Dortmund, Germany

Madhu Sudan

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

New York University, NY, USA

Doug Tygar

University of California, Berkeley, CA, USA

Moshe Y. Vardi

Rice University, Houston, TX, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Bart Goethals Arno Siebes (Eds.)

Knowledge Discovery in Inductive Databases

Third International Workshop, KDID 2004

Pisa, Italy, September 20, 2004

Revised Selected and Invited Papers



Springer

Volume Editors

Bart Goethals

University of Antwerp, Department of Mathematics and Computer Science

Middelheimlaan 1, 2020 Antwerp, Belgium

E-mail: bart.goethals@ua.ac.be

Arno Siebes

Utrecht University, Institute of Information and Computing Sciences

PO Box 80.089, 3508TB Utrecht, The Netherlands

E-mail: arno.siebes@cs.uu.nl

Library of Congress Control Number: 2005921108

CR Subject Classification (1998): H.2, I.2

ISSN 0302-9743

ISBN 3-540-25082-4 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2005

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper SPIN: 11400059 06/3142 5 4 3 2 1 0

Preface

The 3rd International Workshop on Knowledge Discovery in Inductive Databases (KDID 2004) was held in Pisa, Italy, on September 20, 2004 as part of the 15th European Conference on Machine Learning and the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2004).

Ever since the start of the field of data mining, it has been realized that the knowledge discovery and data mining process should be integrated into database technology. This idea has been formalized in the concept of inductive databases, introduced by Imielinski and Mannila (CACM 1996, 39(11)).

In general, an inductive database is a database that supports data mining and the knowledge discovery process in a natural and elegant way. In addition to the usual data, it also contains inductive generalizations (e.g., patterns, models) extracted from the data. Within this framework, knowledge discovery is an interactive process in which users can query the inductive database to gain insight to the data and the patterns and models within that data.

Despite many recent developments, there still exists a pressing need to understand the central issues in inductive databases. This workshop aimed to bring together database and data mining researchers and practitioners who are interested in the numerous challenges that inductive databases offers.

This workshop followed the previous two workshops: KDID 2002 held in Helsinki, Finland, and KDID 2003 held in Cavtat-Dubrovnik, Croatia.

The scientific program of the workshop included 9 papers, selected out of 23 submissions, and an invited talk by Sunita Sarawagi. During the workshop, only informal proceedings were distributed. The papers in this volume were revised by the authors based on the comments from the refereeing stage and ensuing discussions during the workshop, and were subjected to a final acceptance by the Program Committee.

We wish to thank the invited speaker, all the authors who submitted their papers to the workshop, the Program Committee members for their help in the reviewing process, and the ECML/PKDD Organization Committee for their help and local organization.

December 2004

Bart Goethals
Arno Siebes

Program Chairs

Bart Goethals

Helsinki Institute for Information Technology – Basic Research Unit

Department of Computer Science

University of Helsinki

Finland

<http://www.cs.helsinki.fi/bart.goethals/>

Arno Siebes

Institute of Information and Computing Sciences

Utrecht University

The Netherlands

<http://www.cs.uu.nl/staff/siebes.html>

Program Committee

Roberto Bayardo, *IBM Almaden, USA*

Francesco Bonchi, *ISTI-CNR, Italy*

Jean-François Boulicaut, *INSA Lyon, France*

Toon Calders, *University of Antwerp, Belgium*

Luc De Raedt, *Albert-Ludwigs-Universitaet Freiburg, Germany*

Saso Dzeroski, *Jozef Stefan Institute, Slovenia*

Minos N. Garofalakis, *Bell Labs, USA*

Johannes Gehrke, *Cornell University, USA*

Mika Klemettinen, *Nokia, Finland*

Heikki Mannila, *HIIT-BRU, University of Helsinki,*

Helsinki University of Technology, Finland

Rosa Meo, *University of Turin, Italy*

Ryszard S. Michalski, *George Mason University, USA*

Taneli Mielikäinen, *HIIT-BRU, University of Helsinki, Finland*

Mohammed Zaki, *Rensselaer Polytechnic Institute, USA*

Table of Contents

Invited Paper

Models and Indices for Integrating Unstructured Data with a Relational Database

Sunita Sarawagi 1

Contributed Papers

Constraint Relaxations for Discovering Unknown Sequential Patterns

Cláudia Antunes, Arlindo L. Oliveira 11

Mining Formal Concepts with a Bounded Number of Exceptions from Transactional Data

Jérémy Besson, Céline Robardet, Jean-François Boulicaut 33

Theoretical Bounds on the Size of Condensed Representations

Nele Dexters, Toon Calders 46

Mining Interesting XML-Enabled Association Rules with Templates

Ling Feng, Tharam Dillon 66

Database Transposition for Constrained (Closed) Pattern Mining

Baptiste Jeudy, François Rioult 89

An Efficient Algorithm for Mining String Databases Under Constraints

Sau Dan Lee, Luc De Raedt 108

An Automata Approach to Pattern Collections

Taneli Mielikäinen 130

Implicit Enumeration of Patterns

Taneli Mielikäinen 150

Condensed Representation of EPs and Patterns Quantified by Frequency-Based Measures

Arnaud Soulet, Bruno Crémilleux, François Rioult 173

Author Index 191