

LNCS 3181

Yahiko Kambayashi  
Mukesh Mohania  
Wolfram Wöß (Eds.)

# Data Warehousing and Knowledge Discovery

6th International Conference, DaWaK 2004  
Zaragoza, Spain, September 2004  
Proceedings



Yahiko Kambayashi Mukesh Mohania  
Wolfram WöB (Eds.)

# Data Warehousing and Knowledge Discovery

6th International Conference, DaWaK 2004  
Zaragoza, Spain, September 1-3, 2004  
Proceedings

## Volume Editors

Yahiko Kambayashi  
Kyoto University  
Graduate School of Informatics  
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan  
E-mail: kambayashi@db.soc.i.kyoto-u.ac.jp

Mukesh Mohania  
IBM India Research Lab  
Block No. 1, I.I.T., Hauz Khas, New Delhi 110016, India  
E-mail: mkmukesh@in.ibm.com

Wolfram Wöß  
University of Linz, FAW Institute for Applied Knowledge Processing  
Altenberger Str. 69, 4040 Linz, Austria  
E-mail: wwoess@faw.uni-linz.ac.at

Library of Congress Control Number: 2004110973

CR Subject Classification (1998): H.2, H.3, H.4, C.2, H.5, I.2

ISSN 0302-9743

ISBN 3-540-22937-X 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 2004  
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik  
Printed on acid-free paper      SPIN: 11312215      06/3142      5 4 3 2 1 0

*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*

# Lecture Notes in Computer Science

For information about Vols. 1–3083

please contact your bookseller or Springer

- Vol. 3220: J.C. Lester, R.M. Vicari, F. Paragau\c cu (Eds.), *Intelligent Tutoring Systems*. XXI, 920 pages. 2004.
- Vol. 3208: H.J. Ohlbach, S. Schaffert (Eds.), *Principles and Practice of Semantic Web Reasoning*. VII, 165 pages. 2004.
- Vol. 3207: L.T. Jang, M. Guo, G.R. Gao, N.K. Jha, *Embedded and Ubiquitous Computing*. XX, 1116 pages. 2004.
- Vol. 3205: N. Davies, E. Mynatt, I. Siio (Eds.), *UbiComp 2004: Ubiquitous Computing*. XVI, 452 pages. 2004.
- Vol. 3198: G.-J. de Vreede, L.A. Guerrero, G. Marín Raventós (Eds.), *Groupware: Design, Implementation and Use*. XI, 378 pages. 2004.
- Vol. 3194: R. Camacho, R. King, A. Srinivasan (Eds.), *Inductive Logic Programming*. XI, 361 pages. 2004. (Subseries LNAI).
- Vol. 3186: Z. Bellahsene, T. Milo, M. Rys, D. Suciu, R. Unland (Eds.), *Database and XML Technologies*. X, 235 pages. 2004.
- Vol. 3184: S. Katsikas, J. Lopez, G. Pernul (Eds.), *Trust and Privacy in Digital Business*. XI, 299 pages. 2004.
- Vol. 3183: R. Traunmüller (Ed.), *Electronic Government*. XIX, 583 pages. 2004.
- Vol. 3182: K. Bauknecht, M. Bichler, B. Pröll (Eds.), *E-Commerce and Web Technologies*. XI, 370 pages. 2004.
- Vol. 3181: Y. Kambayashi, M. Mohania, W. Wöb (Eds.), *Data Warehousing and Knowledge Discovery*. XIV, 412 pages. 2004.
- Vol. 3180: F. Galindo, M. Takizawa, R. Traunmüller (Eds.), *Database and Expert Systems Applications*. XXI, 945 pages. 2004.
- Vol. 3178: W. Jonker, M. Petkovic (Eds.), *Secure Data Management*. VIII, 219 pages. 2004.
- Vol. 3177: Z.R. Yang, H. Yin, R. Everson (Eds.), *Intelligent Data Engineering and Automated Learning – IDEAL 2004*. XVIII, 852 pages. 2004.
- Vol. 3175: C.E. Rasmussen, H.H. Bülthoff, B. Schölkopf, M.A. Giese (Eds.), *Pattern Recognition*. XVIII, 581 pages. 2004.
- Vol. 3174: F. Yin, J. Wang, C. Guo (Eds.), *Advances in Neural Networks - ISNN 2004*. XXXV, 1021 pages. 2004.
- Vol. 3172: M. Dorigo, M. Birattari, C. Blum, L. M. Gambardella, F. Mondada, T. Stützle (Eds.), *Ant Colony, Optimization and Swarm Intelligence*. XII, 434 pages. 2004.
- Vol. 3170: P. Gardner, N. Yoshida (Eds.), *CONCUR 2004 - Concurrency Theory*. XIII, 529 pages. 2004.
- Vol. 3166: M. Rauterberg (Ed.), *Entertainment Computing – ICEC 2004*. XXIII, 617 pages. 2004.
- Vol. 3158: I. Nikolaidis, M. Barbeau, E. Kranakis (Eds.), *Ad-Hoc, Mobile, and Wireless Networks*. IX, 344 pages. 2004.
- Vol. 3157: C. Zhang, H. W. Guesgen, W.K. Yeap (Eds.), *PRICAI 2004: Trends in Artificial Intelligence*. XX, 1023 pages. 2004. (Subseries LNAI).
- Vol. 3156: M. Joye, J.-J. Quisquater (Eds.), *Cryptographic Hardware and Embedded Systems - CHES 2004*. XIII, 455 pages. 2004.
- Vol. 3155: P. Funk, P.A. González Calero (Eds.), *Advances in Case-Based Reasoning*. XIII, 822 pages. 2004. (Subseries LNAI).
- Vol. 3154: R.L. Nord (Ed.), *Software Product Lines*. XIV, 334 pages. 2004.
- Vol. 3153: J. Fiala, V. Koubek, J. Kratochvíl (Eds.), *Mathematical Foundations of Computer Science 2004*. XIV, 902 pages. 2004.
- Vol. 3152: M. Franklin (Ed.), *Advances in Cryptology – CRYPTO 2004*. XI, 579 pages. 2004.
- Vol. 3150: G.-Z. Yang, T. Jiang (Eds.), *Medical Imaging and Augmented Reality*. XII, 378 pages. 2004.
- Vol. 3149: M. Danelutto, M. Vanneschi, D. Laforenza (Eds.), *Euro-Par 2004 Parallel Processing*. XXXIV, 1081 pages. 2004.
- Vol. 3148: R. Giacobazzi (Ed.), *Static Analysis*. XI, 393 pages. 2004.
- Vol. 3146: P. Érdi, A. Esposito, M. Marinaro, S. Scarpetta (Eds.), *Computational Neuroscience: Cortical Dynamics*. XI, 161 pages. 2004.
- Vol. 3144: M. Papatriantafilou, P. Hunel (Eds.), *Principles of Distributed Systems*. XI, 246 pages. 2004.
- Vol. 3143: W. Liu, Y. Shi, Q. Li (Eds.), *Advances in Web-Based Learning – ICWL 2004*. XIV, 459 pages. 2004.
- Vol. 3142: J. Diaz, J. Karhumäki, A. Lepistö, D. Sannella (Eds.), *Automata, Languages and Programming*. XIX, 1253 pages. 2004.
- Vol. 3140: N. Koch, P. Fraternali, M. Wirsing (Eds.), *Web Engineering*. XXI, 623 pages. 2004.
- Vol. 3139: F. Iida, R. Pfeifer, L. Steels, Y. Kuniyoshi (Eds.), *Embodied Artificial Intelligence*. IX, 331 pages. 2004. (Subseries LNAI).
- Vol. 3138: A. Fred, T. Caelli, R.P.W. Duin, A. Campilho, D. d. Ridder (Eds.), *Structural, Syntactic, and Statistical Pattern Recognition*. XXII, 1168 pages. 2004.
- Vol. 3137: P. De Bra, W. Nejdl (Eds.), *Adaptive Hypermedia and Adaptive Web-Based Systems*. XIV, 442 pages. 2004.
- Vol. 3136: F. Meziane, E. Métais (Eds.), *Natural Language Processing and Information Systems*. XII, 436 pages. 2004.

- Vol. 3134: C. Zannier, H. Erdogmus, L. Lindstrom (Eds.), *Extreme Programming and Agile Methods - XP/Agile Universe 2004*. XIV, 233 pages. 2004.
- Vol. 3133: A.D. Pimentel, S. Vassiliadis (Eds.), *Computer Systems: Architectures, Modeling, and Simulation*. XIII, 562 pages. 2004.
- Vol. 3132: B. Demoen, V. Lifschitz (Eds.), *Logic Programming*. XII, 480 pages. 2004.
- Vol. 3131: V. Torra, Y. Narukawa (Eds.), *Modeling Decisions for Artificial Intelligence*. XI, 327 pages. 2004. (Subseries LNAI).
- Vol. 3130: A. Syropoulos, K. Berry, Y. Haralambous, B. Hughes, S. Peter, J. Plaice (Eds.), *TeX, XML, and Digital Typography*. VIII, 265 pages. 2004.
- Vol. 3129: Q. Li, G. Wang, L. Feng (Eds.), *Advances in Web-Age Information Management*. XVII, 753 pages. 2004.
- Vol. 3128: D. Asonov (Ed.), *Querying Databases Privately*. IX, 115 pages. 2004.
- Vol. 3127: K.E. Wolff, H.D. Pfeiffer, H.S. Delugach (Eds.), *Conceptual Structures at Work*. XI, 403 pages. 2004. (Subseries LNAI).
- Vol. 3126: P. Dini, P. Lorenz, J.N.d. Souza (Eds.), *Service Assurance with Partial and Intermittent Resources*. XI, 312 pages. 2004.
- Vol. 3125: D. Kozen (Ed.), *Mathematics of Program Construction*. X, 401 pages. 2004.
- Vol. 3124: J.N. de Souza, P. Dini, P. Lorenz (Eds.), *Telecommunications and Networking - ICT 2004*. XXVI, 1390 pages. 2004.
- Vol. 3123: A. Belz, R. Evans, P. Piwek (Eds.), *Natural Language Generation*. X, 219 pages. 2004. (Subseries LNAI).
- Vol. 3122: K. Jansen, S. Khanna, J.D.P. Rolim, D. Ron (Eds.), *Approximation, Randomization, and Combinatorial Optimization*. IX, 428 pages. 2004.
- Vol. 3121: S. Nikolettseas, J.D.P. Rolim (Eds.), *Algorithmic Aspects of Wireless Sensor Networks*. X, 201 pages. 2004.
- Vol. 3120: J. Shawe-Taylor, Y. Singer (Eds.), *Learning Theory*. X, 648 pages. 2004. (Subseries LNAI).
- Vol. 3118: K. Miesenberger, J. Klaus, W. Zagler, D. Burger (Eds.), *Computer Helping People with Special Needs*. XXIII, 1191 pages. 2004.
- Vol. 3116: C. Rattray, S. Maharaj, C. Shankland (Eds.), *Algebraic Methodology and Software Technology*. XI, 569 pages. 2004.
- Vol. 3114: R. Alur, D.A. Peled (Eds.), *Computer Aided Verification*. XII, 536 pages. 2004.
- Vol. 3113: J. Karhumäki, H. Maurer, G. Paun, G. Rozenberg (Eds.), *Theory Is Forever*. X, 283 pages. 2004.
- Vol. 3112: H. Williams, L. MacKinnon (Eds.), *Key Technologies for Data Management*. XII, 265 pages. 2004.
- Vol. 3111: T. Hagerup, J. Katajainen (Eds.), *Algorithm Theory - SWAT 2004*. XI, 506 pages. 2004.
- Vol. 3110: A. Juels (Ed.), *Financial Cryptography*. XI, 281 pages. 2004.
- Vol. 3109: S.C. Sahinalp, S. Muthukrishnan, U. Dogrusoz (Eds.), *Combinatorial Pattern Matching*. XII, 486 pages. 2004.
- Vol. 3108: H. Wang, J. Pieprzyk, V. Varadharajan (Eds.), *Information Security and Privacy*. XII, 494 pages. 2004.
- Vol. 3107: J. Bosch, C. Krueger (Eds.), *Software Reuse: Methods, Techniques and Tools*. XI, 339 pages. 2004.
- Vol. 3106: K.-Y. Chwa, J.I. Munro (Eds.), *Computing and Combinatorics*. XIII, 474 pages. 2004.
- Vol. 3105: S. Göbel, U. Spierling, A. Hoffmann, I. Iurgel, O. Schneider, J. Dechau, A. Feix (Eds.), *Technologies for Interactive Digital Storytelling and Entertainment*. XVI, 304 pages. 2004.
- Vol. 3104: R. Kralovic, O. Sykora (Eds.), *Structural Information and Communication Complexity*. X, 303 pages. 2004.
- Vol. 3103: K. Deb, e. al. (Eds.), *Genetic and Evolutionary Computation - GECCO 2004*. XLIX, 1439 pages. 2004.
- Vol. 3102: K. Deb, e. al. (Eds.), *Genetic and Evolutionary Computation - GECCO 2004*. L, 1445 pages. 2004.
- Vol. 3101: M. Masoodian, S. Jones, B. Rogers (Eds.), *Computer Human Interaction*. XIV, 694 pages. 2004.
- Vol. 3100: J.F. Peters, A. Skowron, J.W. Grzymała-Busse, B. Kostek, R.W. Świniarski, M.S. Szczuka (Eds.), *Transactions on Rough Sets I*. X, 405 pages. 2004.
- Vol. 3099: J. Cortadella, W. Reisig (Eds.), *Applications and Theory of Petri Nets 2004*. XI, 505 pages. 2004.
- Vol. 3098: J. Desel, W. Reisig, G. Rozenberg (Eds.), *Lectures on Concurrency and Petri Nets*. VIII, 849 pages. 2004.
- Vol. 3097: D. Basin, M. Rusinowitch (Eds.), *Automated Reasoning*. XII, 493 pages. 2004. (Subseries LNAI).
- Vol. 3096: G. Melnik, H. Holz (Eds.), *Advances in Learning Software Organizations*. X, 173 pages. 2004.
- Vol. 3095: C. Bussler, D. Fensel, M.E. Orłowska, J. Yang (Eds.), *Web Services, E-Business, and the Semantic Web*. X, 147 pages. 2004.
- Vol. 3094: A. Nürnberger, M. Detryniecki (Eds.), *Adaptive Multimedia Retrieval*. VIII, 229 pages. 2004.
- Vol. 3093: S. Katsikas, S. Gritzalis, J. Lopez (Eds.), *Public Key Infrastructure*. XIII, 380 pages. 2004.
- Vol. 3092: J. Eckstein, H. Baumeister (Eds.), *Extreme Programming and Agile Processes in Software Engineering*. XVI, 358 pages. 2004.
- Vol. 3091: V. van Oostrom (Ed.), *Rewriting Techniques and Applications*. X, 313 pages. 2004.
- Vol. 3089: M. Jakobsson, M. Yung, J. Zhou (Eds.), *Applied Cryptography and Network Security*. XIV, 510 pages. 2004.
- Vol. 3087: D. Maltoni, A.K. Jain (Eds.), *Biometric Authentication*. XIII, 343 pages. 2004.
- Vol. 3086: M. Odersky (Ed.), *ECOOP 2004 - Object-Oriented Programming*. XIII, 611 pages. 2004.
- Vol. 3085: S. Berardi, M. Coppo, F. Damiani (Eds.), *Types for Proofs and Programs*. X, 409 pages. 2004.
- Vol. 3084: A. Persson, J. Stirna (Eds.), *Advanced Information Systems Engineering*. XIV, 596 pages. 2004.



# In Memoriam

**Yahiko Kambayashi**

(1943–2004)



In great sadness we received notice of the sudden and untimely passing away of Prof. Yahiko Kambayashi on Friday, February 6, 2004.

Prof. Kambayashi, one of the leading international pioneers of database research, completed his Ph.D. at Kyoto University in 1970. In his early academic career, Prof. Kambayashi's research topics were logic circuits design, switching theory, and automata theory. From 1971 to 1973 he was a Visiting Research Associate at the University of Illinois, Urbana, where he developed a logic design method called the transduction method, which is now widely used by major US and Japanese logic design software vendors. After his return to Kyoto University as a faculty member in 1973, Prof. Kambayashi started to focus his research on databases. His major research results include an algorithm for the calculation of key dependencies, new inference rules for embedded multivalued dependencies, processing methods for cyclic queries, and new concurrency control mechanisms. In 1984, he became a professor at Kyushu University, where he extended his research area to geographic information systems, schema design in network data models, and concurrency control. In 1990, he was appointed as a professor at Kyoto University, where he conducted several very important practical research projects including studies of applications of database technologies to groupware and distance education systems. From April 2003, he served as Dean of the Graduate School of Informatics at Kyoto University.

Prof. Kambayashi published numerous articles in major journals and conferences. He also was the author and editor of many books and conference proceedings.

Prof. Kambayashi was also a great educator. A record number of Japanese and foreign students received M.S. and Ph.D. degrees under his supervision at Kyoto University and Kyushu University. Many of them are now serving as faculty members at universities in Japan and other countries. Prof. Kambayashi also taught courses at McGill University (1979), Kuwait University (1982) and Wuhan University (1984) as a visiting professor.

Prof. Kambayashi was an IEEE fellow, a trustee of the VLDB Endowment, a member of the SIGMOD Advisory Committee, a vice-chair of the ACM Tokyo/Japan Chapter, chair of the DASFAA Steering Committee, co-chair of the WISE Society and WISE Steering Committee, a member of the CODAS Steering Committee, a member of the ER Steering Committee, a member of the RIDE Steering Committee, a co-editor-in-chief of the World Wide Web Journal, an associate editor of ACM TODS, and a member of the editorial boards of several international journals. He was a winner of the ACM SIGMOD Contribution Award in 1995 for his many professional services in Japan and internationally.

Prof. Kambayashi helped to found the DEXA series of conferences and was one of the initiators of the DaWaK conference for which he served as General Chair from the very beginning in 1999.

Those who knew Prof. Kambayashi remember the energy and stamina with which he not only tackled his own research issues but also supported his colleagues, staff, students, collaborators and guests. Not only profound insights and expertise, but also his friendship and generous hospitality attracted many researchers and students.

Prof. Kambayashi is survived by his wife and two sons. His sudden leave is not only a tragic loss to his family but also a great loss to the whole international research community. Many of us will remember him as a friend, a mentor, a leader, an educator, and as our source of inspiration. We express our heartfelt condolence and our deepest sympathy to his family.

#### **PC Chairs of DaWaK 2004**

Mukesh Mohania (IBM India Research Lab, India)

Wolfram Wöß (FAW, Johannes Kepler University of Linz, Austria)

#### **DEXA President and Vice-President**

Roland Wagner (FAW, Johannes Kepler University of Linz, Austria)

A Min Tjoa (Vienna University of Technology, Austria)

#### **Former PC chairs of DaWaK**

Werner Winiwarter (University of Vienna, Austria)

Masatoshi Arikawa (University of Tokyo, Japan)

A Min Tjoa (Vienna University of Technology, Austria)



# Preface

Within the last few years, data warehousing and knowledge discovery technology has established itself as a key technology for enterprises that wish to improve the quality of the results obtained from data analysis, decision support, and the automatic extraction of knowledge from data.

The 6th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2004) continued a series of successful conferences dedicated to this topic. Its main objective was to bring together researchers and practitioners to discuss research issues and experience in developing and deploying data warehousing and knowledge discovery systems, applications, and solutions.

The conference focused on the logical and physical design of data warehousing and knowledge discovery systems. The scope of the papers covers the most recent and relevant topics in the areas of data cubes and queries, multidimensional data models, XML data mining, data semantics and clustering, association rules, data mining techniques, data analysis and discovery, query optimization, data cleansing, data warehouse design and maintenance, and applications. These proceedings contain the technical papers selected for presentation at the conference.

We received more than 100 papers, including 12 industrial papers, from over 33 countries, and the program committee finally selected 40 papers. The conference program included an invited talk by Kazuo Iwano, IBM Tokyo Research Lab, Japan.

We would like to thank the DEXA 2004 Workshop General Chairs (Prof. Roland Wagner, Prof. A Min Tjoa) and the Organizing Committee of the 15th International Conference on Database and Expert Systems Applications (DEXA 2004) for their support and their cooperation. Many thanks go to Ms. Gabriela Wagner for providing a great deal of assistance as well as to Mr. Raimund Angleitner-Flotzinger and Mr. Andreas Dreiling for administering the conference management software. We are very indebted to all the Program Committee members and external reviewers who reviewed the papers very carefully and in a timely manner. We would also like to thank all the authors who submitted their papers to DaWaK 2004; they provided us with an excellent technical program.

September 2004

Mukesh Mohania  
Wolfram Wöß  
Pankaj Garg

# Program Committee

## General Chairperson

Yahiko Kambayashi, Kyoto University, Japan

## Conference Program Chairpersons

Mukesh Mohania, IBM India Research Lab, India

Wolfram Wöß, Johannes Kepler University of Linz, Austria

Industrial Chair: Pankaj Garg, Intel, Japan

## Program Committee Members

Tatsuya Akutsu, Kyoto University, Japan

Marcelo Arenas, University of Toronto, Canada

Hiroki Arimura, Kyushu University, Japan

Mahmooh Awan, Dubai University College, Dubai

Mike Bain, University of New South Wales, Australia

Elena Baralis, Politecnico di Torino, Italy

Denilson Barbosa, University of Toronto, Canada

Ladjel Bellatreche, ENSMA, Poitiers University, France

Jorge Bernardino, Coimbra Polytechnic, Portugal

Sourav S. Bhowmick, Nanyang Technological University, Singapore

Christian Böhm, UMIT Tirol, Austria

Janez Brank, Jozef Stefan Institute, Ljubljana, Slovenia

Damien Bright, University of South Australia, Australia

Alun Butler, University of Greenwich, London, UK

Luca Cabibbo, Università degli Studi Roma Tre, Italy

Tiziana Catarci, Università degli Studi di Roma La Sapienza, Italy

Chee-Yong Chan, National University of Singapore, Singapore

Elizabeth Chang, Curtin University of Technology, Bentley, Perth, Australia

Arbee L.P. Chen, National Dong Hwa University, Taiwan

Sunil Choenni, Nyenrode University/University of Twente, The Netherlands

Saso Dzeroski, Jozef Stefan Institute, Slovenia

Vladimir Estivill-Castro, Griffith University, Brisbane, Australia

Wenfei Fan, Temple University, USA

Ling Feng, University of Twente, The Netherlands

Jean-Gabriel Ganascia, Université Paris VI, France

Minos Garofalakis, Bell Laboratories, USA

Ananth Grama, Purdue University, USA

Jerzy Grzymala-Busse, University of Kansas, USA

Rajeev Gupta, IBM India Research Lab, India  
S.K. Gupta, IIT Delhi, India  
Marc Gyssens, University of Limburg, Belgium  
Mirsad Hadzikadic, University of North Carolina at Charlotte, USA  
Tu Bao Ho, Japan Advanced Institute of Science and Technology, Japan  
Se June Hong, IBM T.J. Watson Research Center, USA  
Andreas Hotho, University of Karlsruhe, Germany  
Samuel Kaski, University of Helsinki, Finland  
Hiroyuki Kawano, Kyoto University, Japan  
Larry Kerschberg, George Mason University, USA  
Masaru Kitsuregawa, University of Tokyo, Japan  
Guanling Lee, National Dong Hwa University, Taiwan  
Yue-Shi Lee, Ming Chuan University, Taiwan  
Leszek Lilien, Purdue University, USA  
Tok Wang Ling, National University of Singapore, Singapore  
Jixue Liu, University of South Australia, Australia  
Zhang Long, IBM China Research Laboratory, India  
Sanjay Kumar Madria, University of Missouri-Rolla, USA  
Mukesh Mohania, IBM India Research Lab, India  
Anirban Mondal, University of Tokyo, Japan  
Wee Keong Ng, Nanyang Technological University, Singapore  
Dimitris Papadias, Hong Kong University of Science and Technology, China  
Stefano Paraboschi, Politecnico di Milano, Italy  
Torben Bach Pedersen, Aalborg University, Denmark  
Jian Pei, State University of New York, USA  
Clara Pizzuti, ISI-CNR, Italy  
David Powers, Flinders University of South Australia, Australia  
Zbigniew W. Ras, University of North Carolina, USA  
Rajeev Rastogi, Bell Laboratories, USA  
Prasan Roy, IBM India Research Lab, India  
Elke A. Rundensteiner, Worcester Polytechnic Institute, USA  
Domenico Saccà, Università della Calabria, Italy  
Carmem Satie Hara, Universidade Federal do Parana, Brazil  
Michael Schrefl, Johannes Kepler University of Linz, Austria  
Il-Yeol Song, Drexel University, USA  
Bala Srinivasan, Monash University, Australia  
Biplav Srivastava, IBM India Research Lab, India  
Gerd Stumme, University of Karlsruhe, Germany  
Einoshin Suzuki, Yokohama National University, Japan  
Ernest Teniente, Universitat Politècnica de Catalunya, Spain  
A Min Tjoa, Vienna University of Technology, Austria  
Riccardo Torlone, Università Roma Tre, Italy  
Sérgio Viademonte da Rosa, Monash University, Australia  
Millist Vincent, University of South Australia, Australia  
Werner Winiwarter, University of Vienna, Austria

Wolfram Wöß, Johannes Kepler University of Linz, Austria  
Xindong Wu, University of Vermont, USA  
Yiyu Yao, University of Regina, Canada  
Show-Jane Yen, Fu Jen Catholic University, Taiwan  
Kazumasa Yokota, Okayama Prefectural University, Japan

## **External Reviewers**

Periklis Andritsos  
Vasudha Bhatnagar  
Mario Cannataro  
Ling Chen  
Silvia Chiusano  
Byron Choi  
Agnieszka Dardzinska  
Filippo Furfaro  
Paolo Garza  
Martin Goller  
John Horner  
Arto Klami  
Kuriakos Mouratidis  
Dung Nguyen Xuan  
Riccardo Ortale  
Günter Preuner  
Giovanni Quattrone  
Janne Sinkkonen  
Min Song  
Jarkko Venna  
Qiankun Zhao

# Table of Contents

## Data Warehousing Design

Conceptual Design of XML Document Warehouses .....	1
<i>Vicky Nassis, R. Rajugan, Tharam S. Dillon, and Wenny Rahayu</i>	
Bringing Together Partitioning, Materialized Views and Indexes to Optimize Performance of Relational Data Warehouses .....	15
<i>Ladjel Bellatreche, Michel Schneider, Hervé Loringuer, and Mukesh Mohania</i>	
GeoDWFrame: A Framework for Guiding the Design of Geographical Dimensional Schemas .....	26
<i>Robson N. Fidalgo, Valéria C. Times, Joel Silva, and Fernando F. Souza</i>	
Workload-Based Placement and Join Processing in Node-Partitioned Data Warehouses .....	38
<i>Pedro Furtado</i>	

## Knowledge Discovery Framework and XML Data Minig

Novelty Framework for Knowledge Discovery in Databases .....	48
<i>Ahmed Sultan Al-Hegami, Vasudha Bhatnagar, and Naveen Kumar</i>	
Revisiting Generic Bases of Association Rules .....	58
<i>S. Ben Yahia and E. Mephu Nguifo</i>	
Mining Maximal Frequently Changing Subtree Patterns from XML Documents .....	68
<i>Ling Chen, Sourav S. Bhowmick, and Liang-Tien Chia</i>	
Discovering Pattern-Based Dynamic Structures from Versions of Unordered XML Documents .....	77
<i>Qiankun Zhao, Sourav S. Bhowmick, and Sanjay Madria</i>	

## Data Cubes and Queries

Space-Efficient Range-Sum Queries in OLAP .....	87
<i>Fredrik Bengtsson and Jingsen Chen</i>	
Answering Approximate Range Aggregate Queries on OLAP Data Cubes with Probabilistic Guarantees .....	97
<i>Alfredo Cuzzocrea, Wei Wang, and Ugo Matrangolo</i>	

Computing Complex Iceberg Cubes by Multiway Aggregation  
and Bounding ..... 108  
*LienHua Pauline Chou and Xiuzhen Zhang*

**Multidimensional Schema and Data Aggregation**

An Aggregate-Aware Retargeting Algorithm  
for Multiple Fact Data Warehouses ..... 118  
*Karin Beckera and Duncan Dubugras Ruiz*

A Partial Pre-aggregation Scheme for HOLAP Engines ..... 129  
*Wo-Shun Luk and Chao Li*

Discovering Multidimensional Structure in Relational Data ..... 138  
*Mikael R. Jensen, Thomas Holmgren, and Torben Bach Pedersen*

**Inductive Databases and Temporal Rules**

Inductive Databases as Ranking ..... 149  
*Taneli Mielikäinen*

Inductive Databases of Polynomial Equations ..... 159  
*Sašo Džeroski, Ljupčo Todorovski, and Peter Ljubič*

From Temporal Rules to Temporal Meta-rules ..... 169  
*Paul Cotofrei and Kilian Stoffel*

**Industrial Track**

How Is BI Used in Industry?:  
Report from a Knowledge Exchange Network ..... 179  
*Torben Bach Pedersen*

Towards an Adaptive Approach for Mining Data Streams  
in Resource Constrained Environments ..... 189  
*Mohamed Medhat Gaber, Arkady Zaslavsky,  
and Shonali Krishnaswamy*

**Data Clustering**

Exploring Possible Adverse Drug Reactions  
by Clustering Event Sequences ..... 199  
*Hongxing He, Graham Williams, Jie Chen, Simon Hawkins,  
and Chris Kelman*

SCLOPE: An Algorithm for Clustering Data Streams  
of Categorical Attributes ..... 209  
*Kok-Leong Ong, Wenyuan Li, Wee-Keong Ng, and Ee-Peng Lim*



Novel Clustering Approach that Employs Genetic Algorithm with New Representation Scheme and Multiple Objectives .....	219
<i>Jun Du, Erkan Korkmaz, Reda Alhajj, and Ken Barker</i>	

## Data Visualization and Exploration

Categorical Data Visualization and Clustering Using Subjective Factors ...	229
<i>Chia-Hui Chang and Zhi-Kai Ding</i>	

Multidimensional Data Visual Exploration by Interactive Information Segments .....	239
<i>Francisco J. Ferrer-Troyano, Jesús S. Aguilar-Ruiz, and José C. Riquelme</i>	

Metadata to Support Transformations and Data & Metadata Lineage in a Warehousing Environment .....	249
<i>Aurisan Souza de Santana and Ana Maria de Carvalho Moura</i>	

## Data Classification, Extraction and Interpretation

Classification Based on Attribute Dependency .....	259
<i>Yue-Shi Lee and Show-Jane Yen</i>	

OWDEAH: Online Web Data Extraction Based on Access History .....	269
<i>Zhao Li, Wee-Keong Ng, and Kok-Leong Ong</i>	

Data Mining Approaches to Diffuse Large B-Cell Lymphoma Gene Expression Data Interpretation .....	279
<i>Jesús S. Aguilar-Ruiz, Francisco Azuaje, and José C. Riquelme</i>	

## Data Semantics

Deriving Multiple Topics to Label Small Document Regions .....	289
<i>Henner Graubitz and Myra Spiliopoulou</i>	

Deriving Efficient SQL Sequences via Read-Aheads .....	299
<i>A. Soydan Bilgin, Rada Y. Chirkova, Timo J. Salo, and Munindar P. Singh</i>	

Diversity in Random Subspacing Ensembles .....	309
<i>Alexey Tsymbal, Mykola Pechenizkiy, and Pádraig Cunningham</i>	

## Association Rule Mining

Partitioned Approach to Association Rule Mining over Multiple Databases .....	320
<i>Himavalli Kona and Sharma Chakravarthy</i>	

A Tree Partitioning Method for Memory Management  
in Association Rule Mining ..... 331  
*Shakil Ahmed, Frans Coenen, and Paul Leng*

Mining Interesting Association Rules for Prediction  
in the Software Project Management Area ..... 341  
*María N. Moreno García, Francisco J. García Peñalvo,  
and M. José Polo Martín*

**Mining Event Sequences**

PROWL: An Efficient Frequent Continuity Mining Algorithm  
on Event Sequences ..... 351  
*Kuo-Yu Huang, Chia-Hui Chang, and Kuo-Zui Lin*

Algorithms for Discovery of Frequent Superset,  
Rather than Frequent Subset ..... 361  
*Zhung-Xun Liao and Man-Kwan Shan*

Improving Direct Counting for Frequent Itemset Mining ..... 371  
*Adriana Prado, Cristiane Targa, and Alexandre Plastino*

**Pattern Mining**

Mining Sequential Patterns with Item Constraints ..... 381  
*Show-Jane Yen and Yue-Shi Lee*

Mining Borders of the Difference of Two Datacubes ..... 391  
*Alain Casali*

Mining Periodic Patterns in Sequence Data ..... 401  
*Kuo-Yu Huang and Chia-Hui Chang*

**Author Index** ..... 411

# Conceptual Design of XML Document Warehouses

Vicky Nassis<sup>1</sup>, R. Rajugan<sup>2</sup>, Tharam S. Dillon<sup>2</sup>, and Wenny Rahayu<sup>1</sup>

<sup>1</sup>Dept. of CS-CE, La Trobe University, Melbourne, Australia  
{vnassis, wenny}@cs.latrobe.edu.au

<sup>2</sup>Faculty of Information Technology, University of Technology, Sydney, Australia  
{tharam3, rajugan}@it.uts.edu.au

**Abstract.** EXtensible Markup Language (XML) has emerged as the dominant standard in describing and exchanging data among heterogeneous data sources. XML with its self-describing hierarchical structure and its associated XML Schema (XSD) provides the flexibility and the manipulative power needed to accommodate complex, disconnected, heterogeneous data. The issue of large volume of data appearing deserves investigating XML Document Warehouses. But due to XML's non-scalar, set-based semi-structured nature, traditional data design models lack the ability to represent XML design level constructs in an abstract and implementation-independent form which are crucial for designing complex domains such as data marts and data warehouses, but also during their operational and maintenance phase. We utilize Object Oriented (OO) concepts to develop a conceptual model for XML Document Warehouses. In this paper we propose a conceptual design formalism to build meaningful XML Document Warehouses (XDW). Our focus includes; (1) conceptually design and build meaningful XML (warehouse) repository (xFACT) using OO concepts in integration with XML Schema constructs, (2) conceptually model and design *virtual dimensions* using *XML conceptual views* [10a] [10b] to satisfy warehouse end-user requirements and (3) use UML *package* diagrams to help logically group and build hierarchical *conceptual views* to enhance semantics and expressiveness of the XDW.

## 1 Introduction

Data Warehousing (DW) has been an approach adopted for handling large volumes of historical data for detailed analysis and management support. Transactional data in different databases is cleaned, aligned and combined to produce good data warehouses. Since its introduction in 1996, eXtensible Markup Language (XML) has become the *defacto* standard for storing and manipulating self-describing information (meta-data), which creates vocabularies in assisting information exchange between heterogenous data sources over the web [22]. Due to this, there is considerable work to be achieved in order to allow electronic document handling, electronic storage, retrieval and exchange. It is envisaged that XML will also be used for logically encoding documents for many domains. Hence it is likely that a large number of XML documents will populate the would-be repository and several disparate transactional databases.