Yahiko Kambayashi  Mukesh Mohania  Wolfram Wöß (Eds.)

Data Warehousing and Knowledge Discovery

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

Springer
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)

June 2004
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
Mahmoooh 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, Jožef 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, Jožef 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
Program Committee

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
  Vicky Nassis, R. Rajugan, Tharam S. Dillon, and Wenny Rahayu

Bringing Together Partitioning, Materialized Views and Indexes
to Optimize Performance of Relational Data Warehouses ...................... 15
  Ladjel Bellatreche, Michel Schneider, Hervé Lorinquer,  
  and Mukesh Mohania

GeoDWFrame: A Framework for Guiding the Design
of Geographical Dimensional Schemas ........................................ 26
  Robson N. Fidalgo, Valéria C. Times, Joel Silva,  
  and Fernando F. Souza

Workload-Based Placement and Join Processing
in Node-Partitioned Data Warehouses ........................................... 38
  Pedro Furtado

Knowledge Discovery Framework and XML Data Minig

Novelty Framework for Knowledge Discovery in Databases .................... 48
  Ahmed Sultan Al-Hegami, Vasudha Bhatnagar, and Naveen Kumar

Revisiting Generic Bases of Association Rules ................................. 58
  S. Ben Yahia and E. Mephu Nguifo

Mining Maximal Frequently Changing Subtree Patterns
from XML Documents ....................................................................... 68
  Ling Chen, Sourav S. Bhowmick, and Liang-Tien Chia

Discovering Pattern-Based Dynamic Structures from Versions
of Unordered XML Documents ...................................................... 77
  Qiankun Zhao, Sourav S. Bhowmick, and Sanjay Madria

Data Cubes and Queries

Space-Efficient Range-Sum Queries in OLAP ..................................... 87
  Fredrik Bengtsson and Jingsen Chen

Answering Approximate Range Aggregate Queries on OLAP Data Cubes
with Probabilistic Guarantees ...................................................... 97
  Alfredo Cuzzocrea, Wei Wang, and Ugo Matrangolo
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
## Novels Clustering Approach that Employs Genetic Algorithm with New Representation Scheme and Multiple Objectives

*Jun Du, Erkan Korkmaz, Reda Alhajj, and Ken Barker*

### Data Visualization and Exploration

**Categorical Data Visualization and Clustering Using Subjective Factors**

*Chia-Hui Chang and Zhi-Kai Ding*

**Multidimensional Data Visualization by Interactive Information Segments**

*Francisco J. Ferrer-Troyano, Jesús S. Aguilar-Ruiz, and José C. Riquelme*

**Metadata to Support Transformations and Data & Metadata Lineage in a Warehousing Environment**

*Aurisan Souza de Santana and Ana Maria de Carvalho Moura*

### Data Classification, Extraction and Interpretation

**Classification Based on Attribute Dependency**

*Yue-Shi Lee and Show-Jane Yen*

**OWDEAH: Online Web Data Extraction Based on Access History**

*Zhao Li, Wee-Keong Ng, and Kok-Leong Ong*

**Data Mining Approaches to Diffuse Large B–Cell Lymphoma Gene Expression Data Interpretation**

*Jesús S. Aguilar-Ruiz, Francisco Azuaje, and José C. Riquelme*

### Data Semantics

**Deriving Multiple Topics to Label Small Document Regions**

*Henner Graubitz and Myra Spiliopoulou*

**Deriving Efficient SQL Sequences via Read-Aheads**

*A. Soydan Bilgin, Rada Y. Chirkova, Timo J. Salo, and Munindar P. Singh*

**Diversity in Random Subspacing Ensembles**

*Alexey Tsymbal, Mykola Pechenizkiy, and Pádraig Cunningham*

### Association Rule Mining

**Partitioned Approach to Association Rule Mining over Multiple Databases**

*Himavalli Kona and Sharma Chakravarthy*
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