

Springer

Berlin

Heidelberg

New York

Barcelona

Hong Kong

London

Milan

Paris

Singapore

Tokyo

Yahiko Kambayashi Mukesh Mohania
A Min Tjoa (Eds.)

Data Warehousing and Knowledge Discovery

Second International Conference, DaWaK 2000
London, UK, September 4-6, 2000
Proceedings



Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany
Juris Hartmanis, Cornell University, NY, USA
Jan van Leeuwen, Utrecht University, The Netherlands

Volume Editors

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

Mukesh Mohania
Western Michigan University, Computer Science Department
Kalamazoo, MI 49008, USA
E-mail: mohania@cs.wmich.edu

A Min Tjoa
Vienna University of Technology, IFS
Favoritenstr. 9-11/188, 1040 Vienna, Austria

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Data warehousing and knowledge discovery : second international
conference ; proceedings / DaWaK 2000, London, UK, September 4 - 6,
2000. Yahiko Kambayashi . . . (ed.). - Berlin ; Heidelberg ; New York ;
Barcelona ; Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo :
Springer, 2000
(Lecture notes in computer science ; Vol. 1874)
ISBN 3-540-67980-4

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

ISSN 0302-9743

ISBN 3-540-67980-4 Springer-Verlag 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-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH
© Springer-Verlag Berlin Heidelberg 2000
Printed in Germany

Typesetting: Camera-ready by author

Printed on acid-free paper SPIN 10722353 06/3142 5 4 3 2 1 0

Preface

The Second International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2000) was held in Greenwich, UK 4–6 September. DaWaK 2000 was a forum where researchers from data warehousing and knowledge discovery disciplines could exchange ideas on improving next generation decision support and data mining systems.

The conference focused on the logical and physical design of data warehousing and knowledge discovery systems. The scope of the papers covered the most recent and relevant topics in the areas of data warehousing, multidimensional databases, OLAP, knowledge discovery and mining complex databases. These proceedings contain the technical papers selected for presentation at the conference.

We received more than 90 papers from over 20 countries and the program committee finally selected 31 long papers and 11 short papers. The conference program included three invited talks, namely, “A Foolish Consistency: Technical Challenges in Consistency Management” by Professor Anthony Finkelstein, University College London, UK; “European Plan for Research in Data Warehousing and Knowledge Discovery” by Dr. Harald Sonnberger (Head of Unit A4, Eurostat, European Commission); and “Security in Data Warehousing” by Professor Bharat Bhargava, Purdue University, USA.

We would like to thank the DEXA 2000 workshop general chair (Professor Roland Wagner) and the organizing committee of the 11th International Conference on Database and Expert Systems Applications (DEXA 2000) for their support and cooperation. Many many thanks are due Ms Gabriela Wagner for providing a great deal of help and assistance. We are very indebted to all program committee members and outside reviewers who have very carefully and timely reviewed the papers. We would also like to thank all the authors who submitted their papers to DaWaK 2000; they provided us with an excellent technical program.

Yahiko Kambayashi, General Chair
Mukesh Mohania and A Min Tjoa, Program Committee Chairs
Tok Wang Ling, Panel Chair

September 2000

Program Committee

General Chair:

Y. Kambayashi (Kyoto University Sakyo, Japan)

Program Chairs:

M. Mohania (University of South Australia)

A M. Tjoa (Vienna University of Technology, Austria)

Panel Chair:

T.W. Ling (National University of Singapore)

Program Committee:

D. Agrawal (University California Santa Barbara)

P. Atzeni (Università di Roma Tre, Italy)

E. Baralis (Politecnico di Torino, Italy)

S.A. Becker (Florida Institute of Technology, USA)

S. Berchtold (stb Software Technologie Beratung GmbH, Germany)

B. Bhargava (Purdue University, USA)

T. Catarci (Università degli Studi di Roma "La Sapienza", Italy)

S. Chakravarthy (University of Florida, USA)

P. Chamoni (University Duisburg, Germany)

Q. Chen (HP Labs, USA)

Arbee L.P. Chen (National Tsing-Hua University, Taiwan)

C.-M. Chen (Telcordia Technologies, USA)

S. Choenni (National Aerospace Laboratory, The Netherlands)

P.K. Chrysanthis (University of Pittsburgh, USA)

G. Dong (Wright State University, USA)

V. Estivill-Castro (The University of Newcastle, Australia)

U. Fayyad (Microsoft, USA)

S. Goel (Informix, USA)

A. Gupta (Western Michigan University, USA)

S. K. Gupta (Indian Institute of Technology Delhi, India)

M. Gyssens (University of Limburg, Belgium)

J. Han (Simon Fraser University, Canada)

J. Harrison (University of Queensland, Australia)

K. Karlapalem (Hong Kong University of Science and Technology, China)

S. Kaski (Helsinki University of Technology, Finland)

H. Kawano (Kyoto University, Japan)

L. Kerschberg (George Mason University, USA)

M. Kitsuregawa (University of Tokyo, Japan)

F. Korn (AT&T, USA)

V. Kumar (University of Missouri-Kansas City, USA)

L. V.S. Lakshmanan (Concordia University, Canada)

D. Lee (Hong Kong University of Science and Technology, China)

W. Lehner (IBM Almaden Research Center, USA)

L. Libkin (Bell Labs, USA)

Q. Li (City University of Hong Kong, China)
T.W. Ling (National University of Singapore)
H. Lu (Hong Kong University of Science and Technology, China)
S. Madria (Purdue University, USA)
H. Mannila (Microsoft, USA)
S. Navathe (Georgia Institute of Technology, USA)
R. Ng (University of British Columbia, Canada)
W.-K. Ng (Nanyang Technological University, Singapore)
S. Nishio (Osaka University, Japan)
S. Paraboschi (Politecnico di Milano, Italy)
D.S. Parker (University of California, USA)
S. Pramanik (Michigan State University, USA)
C. Rainsford (Defence Science Technology Organisation, Australia)
J. Roddick (University of South Australia)
E. A. Rundensteiner (Worcester Polytechnic Institute, USA)
D. Sacca (Università della Calabria, Italy)
S. Samtani (Telcordia Technologies, USA)
N.L. Sarda (Indian Institute of Technology Mumbai, India)
M. Schrefl (University of South Australia, Australia)
T. Sellis (National Technical University of Athens, Greece)
K. Shim (Bell Labs, USA)
I.-Y. Song (Drexel University, USA)
J. Srivastava (University of Minnesota, USA)
E. Teniente (Universitat Politècnica de Catalunya, Spain)
H. Toivonen (Nokia Research Center, Finland)
M. Vincent (University of South Australia)
B. Wuthrich (Hong Kong University of Science and Technology, China)
J. Zytow (University of North Carolina, USA)

Table of Contents

Data Warehouse Design

The Design and Development of a Logical System for OLAP	1
<i>Cabibbo L., Torlone R.; Italy</i>	

Applying Vertical Fragmentation Techniques in Logical Design of 11	
Multidimensional Databases	
<i>Golfarelli M., Maio D., Rizzi S.; Italy</i>	

Space-Efficient Data Cubes for Dynamic Environments	24
<i>Riedewald M., Agrawal D., El Abbadi A., Pajarola R.; USA</i>	

On Making Data Warehouses Active	34
<i>Schrefl M., Thalhammer T.; Australia, Austria</i>	

Materialized Views

Supporting Hot Spots with Materialized Views	47
<i>Albrecht J., Bauer A., Redert M.; Germany</i>	

Evaluation of Materialized View Indexing in Data Warehousing Environments	57
<i>Bellatreche L., Karlapalem K., Li Q.; China</i>	

View Derivation Graph with Edge Fitting for Adaptive Data Warehousing	67
<i>Stanoi I., Agrawal D., El Abbadi A.; USA</i>	

On the Importance of Tuning in Incremental View Maintenance: An Experience Case Study	77
<i>O’Gorman K., Agrawal D., El Abbadi A.; USA</i>	

Warehouse Data Creation and Maintenance

BEDAWA - A Tool for Generating Sample Data for Data Warehouses	83
<i>Huynh T.N., Nguyen B.T., Schiefer J., Tjoa A M.; Austria</i>	

DyDa: Dynamic Data Warehouse Maintenance in a Fully Concurrent Environment	94
<i>Zhang X., Rundensteiner E.A.; USA</i>	

Scalable Maintenance of Multiple Interrelated Data Warehousing Systems	104
<i>Ding L., Zhang X., Rundensteiner E.A.; USA</i>	

View Maintenance for Hierarchical Semistructured Data 114
Liefke H., Davidson S.B.; USA

Maintaining Horizontally Partitioned Warehouse Views 126
Xu M., Ezeife C.I.; Canada

Invited Talk:

Funding Research in Data Warehousing and Knowledge Discovery
EPROS: The European Plan for Research in Official Statistics 134
Mercy J.-L., Sonnberger H.; Luxembourg

Warehouse Views Selection and Evolution

Elimination of Redundant Views in Multidimensional Aggregates 146
Kotsis N., McGregor D.R.; United Kingdom

Data Cube Compression with QuantiCubes 162
Furtado P., Madeira H.; Portugal

History-Driven View Synchronization 168
Koeller A., Rundensteiner E.A.; USA

A Logical Model for Data Warehouse Design and Evolution 178
Bouzeghoub M., Kedad Z.; France

OLAP System Design and Query Analysis

An Alternative Relational OLAP Modeling Approach 189
Bauer A., Hümmel W., Lehner W.; Germany

Functional Dependencies in Controlling Sparsity of OLAP Cubes..... 199
Niemi T., Nummenmaa J., Thanisch P.; Finland

An OLAP-based Scalable Web Access Analysis Engine 210
Chen Q., Dayal U., Hsu M.; USA

PROMISE: Predicting Query Behavior to Enable Predictive Caching
Strategies for OLAP Systems 224
Sapia C.; Germany

OLAP Query Evaluation

Supporting Online Queries in ROLAP 234
Barbará D., Wu X.; USA

Optimal Multidimensional Query Processing Using Tree Striping	244
<i>Berchtold S., Böhm C., Keim D.A., Kriegel H.-P., Xu X.; Germany</i>	

Enhancing Preprocessing in Data-Intensive Domains using Online-Analytical Processing.....	258
<i>Maedche A., Hotho A., Wiese M.; Germany</i>	

Meta-queries - Computation and Evaluation	265
<i>Ben-Eliyahu-Zohary R., Gudes E.; Israel</i>	

Partitioning Algorithms for the Computation of Average Iceberg Queries	276
<i>Bae J., Lee S.; Korea</i>	

Invited Talk:

Security in Data Warehousing.....	287
<i>Bhargava B.; USA</i>	

Association Rules

Mining of Association Rules in Text Databases Using Inverted Hashing and Pruning	290
<i>Holt J.D., Chung S.M.; USA</i>	

SQL Based Association Rule Mining Using Commercial RDBMS (IBM DB2 UBD EEE)	301
<i>Yoshizawa T., Pramudiono I., Kitsuregawa M.; Japan</i>	

On Supporting Interactive Association Rule Mining	307
<i>Goethals B., Van den Bussche J.; Belgium</i>	

Temporal Association Rules

Discovering Temporal Patterns for Interval-Based Events	317
<i>Kam P.-s., Fu A.W.-c.; China</i>	

An Integrated Query and Mining System for Temporal Association Rules	327
<i>Chen X., Petrounias I. ; United Kingdom</i>	

Mining Changes for Real-Life Applications	337
<i>Liu B., Hsu W., Han H.-S., Xia Y.; Singapore</i>	

AIM: Approximate Intelligent Matching for Time Series Data	347
<i>Kim E.D., Lam J.M.W., Han J.; Canada</i>	

Mining Complex Databases

COFE: A Scalable Method for Feature Extraction from Complex Objects	358
<i>Hristescu G., Farach-Colton M.; USA</i>	
The Pruning Power: Theory and Heuristics for Mining Databases with Multiple k -Nearest-Neighbor Queries	372
<i>Böhm C., Braunmüller B., Kriegel H.-P.; Germany</i>	
Data Mining Support in Database Management Systems	382
<i>Morzy T., Wojciechowski M., Zakrzewicz M.; Poland</i>	
Decision Trees for Probabilistic Data	393
<i>Aboa J.-P., Emilion R.; France</i>	
Mining Frequent Binary Expressions	399
<i>Calders T., Paredaens J.; Belgium</i>	
A Fast Algorithm for Hierarchical Text Classification	409
<i>Chuang W.T., Tiyyagura A., Yang J., Giuffrida G.; USA</i>	
A Hybrid Technique for Data Mining on Balance-Sheet Data	419
<i>Dattilo G., Greco S., Masciari E., Pontieri L.; Italy</i>	
Mondou: Information Navigator with Visual Interface	425
<i>Kawano H., Kawahara M.; Japan</i>	
<i>Vmhist</i> : Efficient Multidimensional Histograms with Improved Accuracy	431
<i>Furtado P., Madeira H.; Portugal</i>	
Author Index	437