

Yahiko Kambayashi Mukesh Mohania
Wolfram Wöß (Eds.)

Data Warehousing and Knowledge Discovery

5th International Conference, DaWaK 2003
Prague, Czech Republic, September 3-5, 2003
Proceedings



Springer

Table of Contents

Invited Talk

XML for Data Warehousing Chances and Challenges <i>P. Fankhauser and T. Klement</i>	1
--	---

Data Cubes and Queries

CPM: A Cube Presentation Model for OLAP <i>A. Maniatis, P. Vassiliadis, S. Skiadopoulos, and Y. Vassiliou</i>	4
Computation of Sparse Data Cubes with Constraints <i>C. Chen, J. Feng, and L. Xiang</i>	14
Answering Joint Queries from Multiple Aggregate OLAP Databases <i>E. Pourabbas and A. Shoshani</i>	24
An Approach to Enabling Spatial OLAP by Aggregating on Spatial Hierarchy <i>L. Zhang, Y. Li, F. Rao, X. Yu, Y. Chen, and D. Liu</i>	35

Multidimensional Data Model

A Multidimensional Aggregation Object (MAO) Framework for Computing Distributive Aggregations <i>M.-F. Tsai and W. Chu</i>	45
The <i>GMD</i> Data Model for Multidimensional Information: A Brief Introduction <i>E. Franconi and A. Kamble</i>	55
An Application of Case-Based Reasoning in Multidimensional Database Architecture <i>D. Simić, V. Kurbalija, and Z. Budimac</i>	66

Web Warehousing

MetaCube XTM: A Multidimensional Metadata Approach for Semantic Web Warehousing Systems <i>T.B. Nguyen, A Min Tjoa, and O. Mangisengi</i>	76
---	----

Designing Web Warehouses from XML Schemas <i>B. Vrdoljak, M. Banek, and S. Rizzi</i>	89
Building XML Data Warehouse Based on Frequent Patterns in User Queries <i>J. Zhang, T.W. Ling, R.M. Bruckner, and A Min Tjoa</i>	99
 Change Detection	
A Temporal Study of Data Sources to Load a Corporate Data Warehouse <i>C. Martín and A. Abelló</i>	109
Automatic Detection of Structural Changes in Data Warehouses <i>J. Eder, C. Koncilia, and D. Mitsche</i>	119
Performance Tests in Data Warehousing ETLM Process for Detection of Changes in Data Origin <i>R.L.A. Rocha, L.F. Cardoso, and J.M. de Souza</i>	129
 Web Mining and Association Rule	
Recent Developments in Web Usage Mining Research <i>F.M. Facca and P.L. Lanzi</i>	140
Parallel Vector Computing Technique for Discovering Communities on the Very Large Scale Web Graph <i>K. Kawase, M. Kawahara, T. Iwashita, H. Kawano, and M. Kawazawa</i> ...	151
 Association Rules and Decision Trees	
Ordinal Association Rules towards Association Rules <i>S. Guillaume</i>	161
Rough Set Based Decision Tree Model for Classification <i>S. Minz and R. Jain</i>	172
Inference Based Classifier: Efficient Construction of Decision Trees for Sparse Categorical Attributes <i>S.-H. Lo, J.-C. Ou, and M.-S. Chen</i>	182
Generating Effective Classifiers with Supervised Learning of Genetic Programming <i>B.-C. Chien, J.-H. Yang, and n. W.-Y. Lin</i>	192

Clustering I

Clustering by Regression Analysis <i>M. Motoyoshi, T. Miura, and I. Shioya</i>	202
Handling Large Workloads by Profiling and Clustering <i>M. Golfarelli</i>	212
Incremental OPTICS: Efficient Computation of Updates in a Hierarchical Cluster Ordering <i>H.-P. Kriegel, P. Kröger, and I. Gotlibovich</i>	224

Clustering II

On Complementarity of Cluster and Outlier Detection Schemes <i>Z. Chen, A. W.-C. Fu, and J. Tang</i>	234
Cluster Validity Using Support Vector Machines <i>V. Estivill-Castro and J. Yang</i>	244
FSSM: Fast Construction of the Optimized Segment Support Map <i>K.-L. Ong, W.-K. Ng, and E.-P. Lim</i>	257

Association Rule Mining

Using a Connectionist Approach for Enhancing Domain Ontologies: Self-Organizing Word Category Maps Revisited <i>M. Dittenbach, D. Merkl, and H. Berger</i>	267
Parameterless Data Compression and Noise Filtering Using Association Rule Mining <i>Y.-K. Woon, X. Li, W.-K. Ng, and W.-F. Lu</i>	278
Performance Evaluation of SQL-OR Variants for Association Rule Mining <i>P. Mishra and S. Chakravarthy</i>	288

Data Analysis and Discovery

A Distance-Based Approach to Find Interesting Patterns <i>C. Zheng and Y. Zhao</i>	299
Similarity Search in Structured Data <i>H.-P. Kriegel and S. Schönauer</i>	309

Ontologies and Improving Data Quality

Using an Interest Ontology for Improved Support in Rule Mining
X. Chen, X. Zhou, R. Scherl, and J. Geller 320

Fraud Formalization and Detection
B. Bhargava, Y. Zhong, and Y. Lu 330

Combining Noise Correction with Feature Selection
C.M. Teng 340

Queries and Data Patterns

Pre-computing Approximate Hierarchical Range Queries
in a Tree-Like Histogram
F. Buccafurri and G. Laz 350

Comprehensive Log Compression with Frequent Patterns
*K. Hätönen, J.F. Boulicaut, M. Klemettinen, M. Miettinen,
and C. Masson* 360

Non Recursive Generation of Frequent K-itemsets
from Frequent Pattern Tree Representations
M. El-Hajj and O.R. Zaïane 371

Improving Database Query Engine

A New Computation Model for Rough Set Theory Based
on Database Systems
J. Han, X. Hu, and T.Y. Lin 381

Computing SQL Queries with Boolean Aggregates
A. Badia 391

Fighting Redundancy in SQL
A. Badia and D. Anand 401

Sampling and Vector Classification

"On-the-fly" VS Materialized Sampling and Heuristics
P. Furtado 412

Incremental and Decremental Proximal Support Vector Classification
Using Decay Coefficients
A. Tveit, M.L. Hetland, and H. Engum 422

Author Index 431