

Yahiko Kambayashi Mukesh Mo
Wolfram Wöß (Eds.)



dandelion.com

2008 AGI Information Management Consultants
May be used for personal purposes only or by
libraries associated to dandelion.com network.

Data Warehousing and Knowledge Discovery

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



Springer

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 Mining

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

Data Visualization and Exploration

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

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

- Metadata to Support Transformations and Data & Metadata Lineage
in a Warehousing Environment 249
Aurisan Souza de Santana and Ana Maria de Carvalho Moura

Data Classification, Extraction and Interpretation

- Classification Based on Attribute Dependency 259
Yue-Shi Lee and Show-Jane Yen

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

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

Data Semantics

- Deriving Multiple Topics to Label Small Document Regions 289
Henner Graubitz and Myra Spiliopoulou

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

- Diversity in Random Subspacing Ensembles 309
Alexey Tsybmal, Mykola Pechenizkiy, and Pádraig Cunningham

Association Rule Mining

- Partitioned Approach to Association Rule Mining
over Multiple Databases 320
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