

Proceedings

16th International Conference on Data Engineering



DATA ENGINEERING

Proceedings

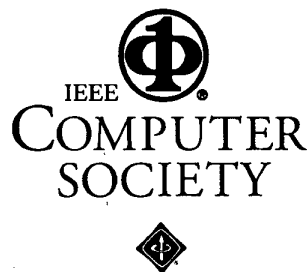
16th International Conference on Data Engineering

29 February – 3 March 2000

San Diego, California

Sponsored by

IEEE Computer Society Technical Committee on Data Engineering



Los Alamitos, California

Washington • Brussels • Tokyo

Copyright © 2000 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries may photocopy beyond the limits of US copyright law, for private use of patrons, those articles in this volume that carry a code at the bottom of the first page, provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Other copying, reprint, or republication requests should be addressed to: IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 133, Piscataway, NJ 08855-1331.

The papers in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and, in the interests of timely dissemination, are published as presented and without change. Their inclusion in this publication does not necessarily constitute endorsement by the editors, the IEEE Computer Society, or the Institute of Electrical and Electronics Engineers, Inc.

IEEE Catalog Number 00CB37073
ISBN: 0-7695-0506-6 (Softbound)
ISBN: 0-7803-6388-4 (Casebound)
ISBN: 0-7695-0508-2 (Microfiche)
ISSN: 1063-6382
PR0506

Additional copies may be ordered from:

IEEE Computer Society
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: + 1-714-821-8380
Fax: + 1-714-821-4641
E-mail: cs.books@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: + 1-732-981-0060
Fax: + 1-732-981-9667
mis.custserv@computer.org

IEEE Computer Society
Asia/Pacific Office
Watanabe Building,
1-4-2 Minami-Aoyama
Minato-ku, Tokyo 107-0062 JAPAN
Tel: + 81-3-3408-3118
Fax: + 81-3-3408-3553
tokyo.ofc@computer.org

Editorial production by Danielle C. Young

Cover art production by Joseph Daigle/Studio Productions

Printed in the United States of America by The Printing House

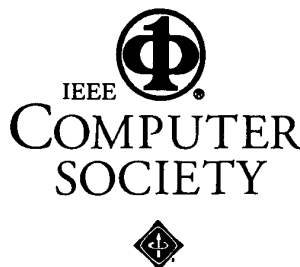


Table of Contents

—16th International Conference on Data Engineering (ICDE'00) —

Message from the General Chair	xv
Message from the Program Committee Co-Chairs.....	xvi
Organizing Committee	xviii
Program Vice-Chairs and Award Committee Members.....	xix
Program Committee Members	xx
External Reviewers.....	xxii
Author Index.....	701

Session 1: Keynote Address

Rules of Thumb in Data Engineering	3
<i>Jim Gray and Prashant Shenoy</i>	

Session 2: Time Series

Online Data Mining for Co-Evolving Time Sequences	13
<i>B.-K. Yi, N. Sidiropoulos, T. Johnson, H. Jagadish, C. Faloutsos, and A. Biliris</i>	
Efficient Searches for Similar Subsequences of Different Lengths in Sequence Databases	23
<i>S. Park, W. Chu, J. Yoon, and C. Hsu</i>	
Landmarks: A New Model for Similarity-Based Pattern Querying in Time Series Databases	33
<i>C.-S. Perng, H. Wang, S. Zhang, and D. Parker</i>	

Session 3: Transactions and Workflow

Managing Escalation of Collaboration Processes in Crisis Mitigation Situations	45
<i>D. Georgakopoulos, H. Schuster, D. Baker, and A. Cichocki</i>	
Semantic Conditions for Correctness at Different Isolation Levels	57
<i>A. Bernstein, P. Lewis, and S. Lu</i>	
Generalized Isolation Level Definitions	67
<i>A. Adya, B. Liskov, and P. O'Neil</i>	

Posters

Session 4: Internet, Performance, and Systems Management

Squeezing the Most out of Relational Database Systems	81
<i>J. Goldstein and R. Ramakrishnan</i>	
Creating a Customized Access Method for Blobworld	82
<i>M. Thomas, C. Carson, and J. Hellerstein</i>	
Efficient Query Subscription Processing in a Multicast Environment	83
<i>A. Crespo, O. Buyukkokten, and H. Garcia-Molina</i>	
Distributed Query Processing on the Web	84
<i>N. Gupta, J. Haritsa, and M. Ramanath</i>	
Probabilistic Data Consistency for Wide-Area Applications.....	85
<i>H. Zou, N. Soparkar, and F. Jahanian</i>	
Dynamic Histograms: Capturing Evolving Data Sets	86
<i>D. Donjerkovic, Y. Ioannidis, and R. Ramakrishnan</i>	
Metadata Propagation in Large, Multi-Layer Database Systems.....	87
<i>A. Rosenthal and E. Sciore</i>	

Industrial

Session 5: Fast and Reliable Database Engines

Extensible Indexing: A Framework for Integrating Domain-Specific Indexing Schemes into Oracle8i	91
<i>J. Srinivasan, R. Murthy, S. Sundara, N. Agarwal, and S. DeFazio</i>	
DB2 Advisor: An Optimizer Smart Enough to Recommend Its Own Indexes.....	101
<i>G. Valentin, M. Zuliani, D. Zilio, G. Lohman, and A. Skelley</i>	
Taming the Downtime: High Availability in Sybase ASE 12	111
<i>S. Raghuram, S. Ranganath, S. Olson, and S. Nandi</i>	

Session 6: Query Processing

Accurate Estimation of the Cost of Spatial Selections	123
<i>A. Aboulnaga and J. Naughton</i>	
User Defined Aggregates in Object-Relational Systems	135
<i>H. Wang and C. Zaniolo</i>	

Scalable Algorithms for Large Temporal Aggregation	145
<i>B. Moon, I. López, and V. Immanuel</i>	

Session 7: Mobile and Embedded Systems

Power Conservative Multi-Attribute Queries on Data Broadcast	157
<i>Q. Hu, W.-C. Lee, and D. Lee</i>	

Multi-Level Multi-Channel Air Cache Designs for Broadcasting in a Mobile Environment	167
<i>K. Prabhakara, K. Hua, and J.-H. Oh</i>	

An Algebraic Compression Framework for Query Results	177
<i>Z. Chen and P. Seshadri</i>	

Posters

Session 8: New Applications

ACQ: An Automatic Clustering and Querying Approach for Large Image Databases	191
<i>D. Yu and A. Zhang</i>	

A Semi-Structured Data Cartridge for Relational Databases	192
<i>F. Sha, G. Gardarin, and L. Némirovski</i>	

The MARIFlow Workflow Management System	193
<i>A. Dogac, M. Ezbiderli, Y. Tambag, C. Icdem, A. Tumer, N. Tatbul, N. Hamali, and C. Beeri</i>	

Device Database Systems	194
<i>P. Bonnet and P. Seshadri</i>	

Distance Exponent: A New Concept for Selectivity Estimation in Metric Trees	195
<i>C. Traina, Jr., A. Traina, and C. Faloutsos</i>	

Efficient Query Refinement in Multimedia Databases	196
<i>K. Chakrabarti, K. Porkaew, and S. Mehrotra</i>	

Interactive-Time Similarity Search for Large Image Collections Using Parallel VA-Files	197
<i>R. Weber, K. Böhm, and H.-J. Schek</i>	

Efficient Storage of XML Data	198
<i>C.-C. Kanne and G. Moerkotte</i>	

Industrial

Session 9: OLAP and Data Warehousing

A Data-Warehouse/OLAP Framework for Scalable Telecommunication Tandem Traffic Analysis	201
<i>Q. Chen, M. Hsu, and U. Dayal</i>	
MetaComm: A Meta-Directory for Telecommunications	211
<i>J. Freire, D. Lieuwen, J. Ordille, L. Garg, M. Holder, H. Urroz, G. Michael, J. Orbach, L. Tucker, Q. Ye, and R. Arlein</i>	
Extracting Delta for Incremental Data Warehouse Maintenance.....	220
<i>P. Ram and L. Do</i>	

Session 10: Multimedia Retrieval

Image Database Retrieval with Multiple-Instance Learning Techniques.....	233
<i>C. Yang and T. Lozano-Pérez</i>	
PAC Nearest Neighbor Queries: Approximate and Controlled Search in High-Dimensional and Metric Spaces	244
<i>P. Ciaccia and M. Patella</i>	
Efficiently Supporting Multiple Similarity Queries for Mining in Metric Databases	256
<i>B. Braunmüller, M. Ester, H.-P. Kriegel, and J. Sander</i>	

Session 11: Storage and Process Optimization

Declustering Using Golden Ratio Sequences	271
<i>R. Bhatia, R. Sinha, and C.-M. Chen</i>	
Optimization Techniques for Data-Intensive Decision Flows.....	281
<i>R. Hull, F. Llirbat, B. Kumar, G. Zhou, G. Dong, and J. Su</i>	
Optimal Index and Data Allocation in Multiple Broadcast Channels	293
<i>S.-C. Lo and A. Chen</i>	

Posters

Session 12: OLAP, DW, and Data Mining

Clustering Categorical Data.....	305
<i>Y. Zhang, A. Fu, C. Cai, and P. Heng</i>	
Discovering Temporal Association Rules: Algorithms, Language and System	306
<i>X. Chen and I. Petrounias</i>	

Mining Bases for Association Rules Using Closed Sets	307
<i>R. Taouil, N. Pasquier, Y. Bastide, and L. Lakhal</i>	
Speeding up View Maintenance Using Cheap Filters at the Warehouse	308
<i>N. Huyn</i>	
Approximate Query Answering with Frequent Sets and Maximum Entropy	309
<i>H. Mannila and P. Smyth</i>	
Association-Based Multiple Imputation in Multivariate Datasets: A Summary	310
<i>W. Zhang</i>	
Optimization of Hypothetical Queries in an OLAP Environment	311
<i>A. Balmin, Y. Papakonstantinou, and T. Papadimitriou</i>	
An Extensible Framework for Data Cleaning	312
<i>H. Galhardas, D. Florescu, D. Shasha, and E. Simon</i>	
<u>Panel</u>	
<u>Session 13: Object/Database Standards Soup</u>	
<i>Moderator: Nelson Mattos, IBM Santa Teresa Laboratory, USA</i>	
<i>Panelists: Linda DeMichiel, JavaSoft/Sun Microsystems, USA</i>	
<i>Stefan Dessloch, IBM Software Division, USA</i>	
<i>Donald Ferguson, IBM Research Division, USA</i>	
<i>Jim Melton, Oracle Corporation, USA</i>	
<i>Mike Pizzo, Microsoft Corporation, USA</i>	
<u>Session 14: Keynote Address</u>	
The Changing Art of Computer Research	319
<i>Dennis Tsichritzis</i>	
<u>Session 15: System Administration</u>	
On-Line Schema Update for a Telecom Database	329
<i>M. Ronström</i>	
Automating Statistics Management for Query Optimizers	339
<i>S. Chaudhuri and V. Narasayya</i>	
A Novel Deadline Driven Disk Scheduling Algorithm for Multi-Priority Multimedia Objects	349
<i>I. Kamel, T. Niranjan, and S. Ghandeharizedah</i>	

Panel

Session 16: Data Mining: Niche Market or Killer App?

Moderator: Ramakrishnan Srikant, IBM Almaden Research Center, USA

Panelists: Umesh Dayal, Hewlett-Packard Research Laboratories, USA

Christos Faloutsos, Carnegie Mellon University, USA

Jim Gray, Microsoft Research, USA

Brian Lent, Amazon.com, USA

Raghu Ramakrishnan, University of Wisconsin-Madison, USA

Session 17: Tutorial

Directories: Managing Data for Networked Applications

D. Srivastava

Session 18: Data Warehousing

Practical Lineage Tracing in Data Warehouses367

Y. Cui and J. Widom

The DC-Tree: A Fully Dynamic Index Structure for Data Warehouses379

M. Ester, J. Kohlhammer, and H.-P. Kriegel

Answering Regular Path Queries Using Views389

D. Calvanese, G. De Giacomo, M. Lenzerini, and M. Vardi

Session 19: Heterogeneous Queries

Query Planning with Limited Source Capabilities.....401

C. Li and E. Chang

Developing Cost Models with Qualitative Variables for Dynamic
Multidatabase Environments413

Q. Zhu, Y. Sun, and S. Matheramgari

Dynamic Query Scheduling in Data Integration Systems.....425

L. Bouganim, F. Fabret, C. Mohan, and P. Valduriez

Session 20: Tutorial

Indexing High-Dimensional Spaces: Database Support for Next
Generation's Applications

D. Keim and S. Berchtold

Session 21: New Trends in Data Mining

DEMON: Mining and Monitoring Evolving Data.....439

V. Ganti, J. Gehrke, and R. Ramakrishnan

CMP: A Fast Decision Tree Classifier Using Multivariate Predictions	449
<i>H. Wang and C. Zaniolo</i>	

Mining Recurrent Items in Multimedia with Progressive Resolution Refinement	461
<i>O. Zaïane, J. Han, and H. Zhu</i>	

Panel

Session 22: Is E-Commerce a New Wave for Database Research?

Moderator: Anant Jhingran, IBM T.J. Watson Research Center, USA

*Panelists: Sesh Murthy, IBM T.J. Watson Research Center, USA
Sham Navathe, Georgia Institute of Technology, USA
Hamid Pirahesh, IBM Almaden Research Center, USA
Krithi Ramamrithan, University of Massachusetts-Amherst, USA*

Industrial

Session 23: Java and Databases

Pure Java Databases for Deployed Applications	477
<i>N. Wyatt</i>	

Database Technology for Internet Applications	700
<i>A. Nori</i>	

Session 24: Association Rules and Correlations

Finding Interesting Associations without Support Pruning.....	489
<i>E. Cohen, M. Datar, S. Fujiwara, A. Gionis, P. Indyk, R. Motwani, J. Ullman, and C. Yang</i>	

Dynamic Miss-Counting Algorithms: Finding Implication and Similarity Rules with Confidence Pruning	501
<i>S. Fujiwara, J. Ullman, and R. Motwani</i>	

Efficient Mining of Constrained Correlated Sets	512
<i>G. Grahne, L. Lakshmanan, and X. Wang</i>	

Session 25: Spatial and Temporal Data

Analyzing Range Queries on Spatial Data.....	525
<i>J. Jin, N. An, and A. Sivasubramaniam</i>	

Data Redundancy and Duplicate Detection in Spatial Join Processing	535
<i>J.-P. Dittrich and B. Seeger</i>	

Query Plans for Conventional and Temporal Queries Involving Duplicates and Ordering	547
<i>G. Slivinskas, C. Jensen, and R. Snodgrass</i>	

Industrial

Session 26: XML and Databases

Oracle8i — The XML <i>Enabled</i> Data Management System	561
<i>S. Banerjee, V. Krishnamurthy, M. Krishnaprasad, and R. Murthy</i>	
XML and DB2.....	569
<i>J. Cheng and J. Xu</i>	

Session 27: High-Dimensional Data

Independent Quantization: An Index Compression Technique for High-Dimensional Data Spaces	577
<i>S. Berchtold, C. Böhm, H. Jagadish, H.-P. Kriegel, and J. Sander</i>	
Deflating the Dimensionality Curse Using Multiple Fractal Dimensions	589
<i>B.-U. Pagel, F. Korn, and C. Faloutsos</i>	
Similarity Search for Multidimensional Data Sequences	599
<i>S.-L. Lee, S.-J. Chun, D.-H. Kim, J.-H. Lee, and C.-W. Chung</i>	

Session 28: Web-Based Systems

XWRAP: An XML-Enabled Wrapper Construction System for Web Information Sources.....	611
<i>L. Liu, C. Pu, and W. Han</i>	
Self-Adaptive User Profiles for Large-Scale Data Delivery	622
<i>U. Çetintemel, M. Franklin, and C. Giles</i>	

Industrial

Session 29: Main Memory and Small Footprint Databases

In-Memory Data Management in the Application Tier.....	637
<i>The TimesTen Team</i>	
SQLServer for Windows CE — A Database Engine for Mobile and Embedded Platforms	642
<i>P. Seshadri and P. Garrett</i>	
Join Enumeration in a Memory-Constrained Environment	645
<i>I. Bowman and G. Paulley</i>	

Plenary Panel

Session 30: XML + Databases = ?

Moderator: *Michael Carey, IBM Almaden Research Center, USA*

Panelists: *Adam Bosworth, Microsoft Corporation, USA*

David DeWitt, University of Wisconsin-Madison, USA

Alon Levy, University of Washington, USA

Bruce Lindsay, IBM Almaden Research Center, USA

Jennifer Widom, Stanford University, USA

Demo Session 1

Web Query Optimizer	661
<i>V. Zadorozhny, L. Bright, L. Raschid, T. Urhan, and M. Vidal</i>	
ReQueSS: Relational Querying of Semi-Structured Data.....	664
<i>R. Sunderraman</i>	
The IDEAL Approach to Internet-Based Negotiation for E-Business.....	666
<i>J. Hammer, C. Huang, Y. Huang, C. Pluempitiwiriyawej, M. Lee, H. Li, L. Wang, Y. Liu, and S. Su</i>	
READY: A High Performance Event Notification Service	668
<i>R. Gruber, B. Krishnamurthy, and E. Panagos</i>	
A Multimedia Information Server with Mixed Workload Scheduling	670
<i>G. Nerjes</i>	
DISIMA: An Object-Oriented Approach to Developing an Image Database System.....	672
<i>V. Oria, T. Özsu, P. Iglinski, B. Xu, and L. Cheng</i>	

Demo Session 2

The Collaboration Management Infrastructure	677
<i>H. Schuster, D. Baker, A. Cichocki, D. Georgakopoulos, and M. Rusinkiewicz</i>	
Assisting the Integration of Taxonomic Data: The LITCHI Toolkit	679
<i>I. Sutherland, J. Robinson, S. Brandt, A. Jones, S. Embury, W. Gray, R. White, and F. Bisby</i>	
TheaterLoc: Using Information Integration Technology to Rapidly Build Virtual Applications	681
<i>G. Barish, Y.-S. Chen, D. DiPasquo, C. Knoblock, S. Minton, I. Muslea, and C. Shahabi</i>	
Lineage Tracing in a Data Warehousing System	683
<i>Y. Cui and J. Widom</i>	

The Mentor-Lite Prototype: A Light-Weight Workflow Management System	685
<i>J. Weissenfels, M. Gillmann, O. Roth, G. Shegalov, and W. Wonner</i>	
Location Prediction and Queries for Tracking Moving Objects	687
<i>O. Wolfson, B. Xu, and S. Chamberlain</i>	
Semiorder Database for Complex Activity Recognition in Multi-Sensory Environments	689
<i>S. Bhonsle, A. Gupta, S. Santini, and R. Jain</i>	
<u>Tutorials</u>	
Tutorial 1: Web Information Retrieval	693
<i>M. Henzinger</i>	
Tutorial 2: Mobile and Wireless Database Access for Pervasive Computing	694
<i>P. Chrysanthis and E. Pitoura</i>	
Tutorial 3: Data Mining with Decision Trees	696
<i>J. Gehrke</i>	
Tutorial 4: Directories: Managing Data for Networked Applications	697
<i>D. Srivastava</i>	
Tutorial 5: Indexing High-Dimensional Spaces: Database Support for Next Decade's Applications	698
<i>S. Berchtold and D. Keim</i>	