

Proceedings

ICDE 2005

Table of Contents

Message from the General Chairs	xvii
Welcome from the Program Chairs	xviii
Conference Officers	xix
Program Committee	xxiv
Demo Program Committee	xxviii
External Reviewers	xxix

KEYNOTES

“One Size Fits All”: An Idea Whose Time Has Come and Gone	2
<i>M. Stonebraker and U. Çetintemel</i>	
Top Five Data Challenges for the Next Decade	12
<i>P. Selinger</i>	
IC Tag Based Traceability: System and Solutions	13
<i>Y. Taniguchi and N. Sagawa</i>	

RESEARCH SESSIONS

Research Session 1: Data Stream Mining

Effective Computation of Biased Quantiles over Data Streams	20
<i>G. Cormode, F. Korn, S. Muthukrishnan, and D. Srivastava</i>	
Range-Efficient Computation of F_0 over Massive Data Streams	32
<i>A. Pavan and S. Tirthapura</i>	
A Unified Framework for Monitoring Data Streams in Real Time	44
<i>A. Bulut and A. K. Singh</i>	

Research Session 2: Schema Matching

Corpus-Based Schema Matching.....	57
<i>J. Madhavan, P. A. Bernstein, A. Doan, and A. Halevy</i>	
Schema Matching Using Duplicates.....	69
<i>A. Bilke and F. Naumann</i>	
Representing and Querying Data Transformations.....	81
<i>Y. Velegrakis, R. J. Miller, and J. Mylopoulos</i>	

Research Session 3: Caching and View Maintenance

Bypass Caching: Making Scientific Databases Good Network Citizens.....	94
<i>T. Malik, R. Burns, and A. Chaudhary</i>	
Asymmetric Batch Incremental View Maintenance.....	106
<i>H. He, J. Xie, J. Yang, and H. Yu</i>	
Adaptive Caching for Continuous Queries.....	118
<i>S. Babu, K. Munagala, J. Widom, and R. Motwani</i>	

Research Session 4: Sensor Networks

Snapshot Queries: Towards Data-Centric Sensor Networks.....	131
<i>Y. Kotidis</i>	
Exploiting Correlated Attributes in Acquisitional Query Processing.....	143
<i>A. Deshpande, C. Guestrin, W. Hong, and S. Madden</i>	
Data Triage: An Adaptive Architecture for Load Shedding in TelegraphCQ.....	155
<i>F. Reiss and J. M. Hellerstein</i>	
Energy-Efficient Data Organization and Query Processing in Sensor Networks.....	157
<i>R. Gummadi, X. Li, R. Govindan, C. Shahabi, and W. Hong</i>	
AutoLag: Automatic Discovery of Lag Correlations in Stream Data.....	159
<i>Y. Sakurai, S. Papadimitriou, and C. Faloutsos</i>	

Research Session 5: Querying Ordered Data

Adaptive Processing of Top- <i>k</i> Queries in XML.....	162
<i>A. Marian, S. Amer-Yahia, N. Koudas, and D. Srivastava</i>	
Progressive Distributed Top- <i>k</i> Retrieval in Peer-to-Peer Networks.....	174
<i>W.-T. Balke, W. Nejdl, W. Siberski, and U. Thaden</i>	
Reverse Nearest Neighbors in Large Graphs.....	186
<i>M. L. Yiu, D. Papadias, N. Mamoulis, and Y. Tao</i>	

Optimizing Access Cost for Top-k Queries over Web Sources: A Unified Cost-Based Approach	188
<i>S. Hwang and K. C.-C. Chang</i>	

Efficient Processing of Skyline Queries with Partially-Ordered Domains	190
<i>C.-Y. Chan, P.-K. Eng, and K.-L. Tan</i>	

Research Session 6: Privacy-Preserving Data Management

A Framework for High-Accuracy Privacy-Preserving Mining	193
<i>S. Agrawal and J. R. Haritsa</i>	

Top-Down Specialization for Information and Privacy Preservation	205
<i>B. C. M. Fung, K. Wang, and P. S. Yu</i>	

Data Privacy through Optimal k-Anonymization	217
<i>R. J. Bayardo and R. Agrawal</i>	

Research Session 7: P2P and Data Dissemination

A Comparative Evaluation of Transparent Scaling Techniques for Dynamic Content Servers	230
<i>C. Amza, A. L. Cox, and W. Zwaenepoel</i>	

SemCast: Semantic Multicast for Content-Based Data Dissemination	242
<i>O. Papaemmanouil and U. Çetintemel</i>	

A Distributed Quadtree Index for Peer-to-Peer Settings	254
<i>E. Tanin, A. Harwood, and H. Samet</i>	

Adlib: A Self-Tuning Index for Dynamic Peer-to-Peer Systems	256
<i>P. Ganesan, Q. Sun, and H. Garcia-Molina</i>	

DUP: Dynamic-Tree Based Update Propagation in Peer-to-Peer Networks	258
<i>L. Yin and G. Cao</i>	

Research Session 8: XML Query Processing (I)

Vectorizing and Querying Large XML Repositories	261
<i>P. Buneman, B. Choi, W. Fan, R. Hutchison, R. Mann, and S. D. Viglas</i>	

IMAX: Incremental Maintenance of Schema-Based XML Statistics	273
<i>M. Ramanath, L. Zhang, J. Freire, and J. R. Haritsa</i>	

BOXes: Efficient Maintenance of Order-Based Labeling for Dynamic XML Data	285
<i>A. Silberstein, H. He, K. Yi, and J. Yang</i>	

Research Session 9: DBMS Architecture and Performance

Efficient Inverted Lists and Query Algorithms for Structured Value Ranking in Update-Intensive Relational Databases	298
<i>L. Guo, J. Shanmugasundaram, K. Beyer, and E. Shekita</i>	

Compressing Bitmap Indices by Data Reorganization	310
<i>A. Pinar, T. Tao, and H. Ferhatosmanoglu</i>	
Dynamic Load Management for Distributed Continuous Query Systems.....	322
<i>Y. Zhou, B. C. Ooi, and K.-L. Tan</i>	
Practical Data Management Techniques for Vehicle Tracking Data.....	324
<i>S. Brakatsoulas, D. Pfoser, and N. Tryfona</i>	
Filter Based Directory Replication and Caching	326
<i>A. Kumar</i>	

Research Session 10: Clustering

On Discovery of Extremely Low-Dimensional Clusters Using Semi-supervised Projected Clustering	329
<i>K. Y. Yip, D. W. Cheung, and M. K. Ng</i>	
Clustering Aggregation.....	341
<i>A. Gionis, H. Mannila, and P. Tsaparas</i>	
Mining Cross-Graph Quasi-cliques in Gene Expression and Protein Interaction Data	353
<i>J. Pei, D. Jiang, and A. Zhang</i>	
CLICKS: Mining Subspace Clusters in Categorical Data via K-Partite Maximal Cliques	355
<i>M. J. Zaki and M. Peters</i>	
Mining Closed Relational Graphs with Connectivity Constraints	357
<i>X. Yan, X. J. Zhou, and J. Han</i>	

Research Session 11: XML Indexing and Dissemination

Efficient Creation and Incremental Maintenance of the HOPI Index for Complex XML Document Collections	360
<i>R. Schenkel, A. Theobald, and G. Weikum</i>	
On the Sequencing of Tree Structures for XML Indexing.....	372
<i>H. Wang and X. Meng</i>	
Efficient Algorithms for Pattern Matching on Directed Acyclic Graphs.....	384
<i>L. Chen, A. Gupta, and M. E. Kurul</i>	
VLEI Code: An Efficient Labeling Method for Handling XML Documents in an RDB.....	386
<i>K. Kobayashi, W. Liang, D. Kobayashi, A. Watanabe, and H. Yokota</i>	
BlossomTree: Evaluating XPathS in FLWOR Expressions	388
<i>N. Zhang, S. K. Agrawal, and M. T. Özsu</i>	

Research Session 12: Mobility

Change Tolerant Indexing for Constantly Evolving Data	391
<i>R. Cheng, Y. Xia, S. Prabhakar, and R. Shah</i>	

Proactive Caching for Spatial Queries in Mobile Environments	403
<i>H. Hu, J. Xu, W. S. Wong, B. Zheng, D. L. Lee, and W.-C. Lee</i>	
Improving Data Accessibility for Mobile Clients through Cooperative Hoarding.....	415
<i>K. Y. Lai, Z. Tari, and P. Bertok</i>	
A Fully Distributed Spatial Index for Wireless Data Broadcast	417
<i>W.-C. Lee and B. Zheng</i>	
Efficient Data Management on Lightweight Computing Devices	419
<i>R. Sen and K. Ramamritham</i>	

Research Session 13: Transactions

Postgres-R(SI): Combining Replica Control with Concurrency Control Based on Snapshot Isolation.....	422
<i>S. Wu and B. Kemme</i>	
SNAP: Efficient Snapshots for Back-in-Time Execution.....	434
<i>L. Shrira and H. Xu</i>	
Improving Preemptive Prioritization via Statistical Characterization of OLTP Locking	446
<i>D. T. McWherter, B. Schroeder, A. Ailamaki, and M. Harchol-Balter</i>	

Research Session 14: Data Integration

A Probabilistic XML Approach to Data Integration.....	459
<i>M. van Keulen, A. de Keijzer, and W. Alink</i>	
A Relationally Complete Visual Query Language for Heterogeneous Data Sources and Pervasive Querying	471
<i>S. Polyviou, G. Samaras, and P. Evmiridou</i>	
Triggers over XML Views of Relational Data	483
<i>F. Shao, A. Novak, and J. Shanmugasundaram</i>	
THALIA: Test Harness for the Assessment of Legacy Information Integration Approaches.....	485
<i>J. Hammer, M. Stonebraker, and O. Topsakal</i>	
Integrating Data from Disparate Sources: A Mass Collaboration Approach.....	487
<i>R. McCann, A. Kramnik, W. Shen, V. Varadarajan, O. Sobulo, and A. Doan</i>	

Research Session 15: Query Processing

On the Optimal Ordering of Maps and Selections under Factorization.....	490
<i>T. Neumann, S. Helmer, and G. Moerkotte</i>	
Stabbing the Sky: Efficient Skyline Computation over Sliding Windows	502
<i>X. Lin, Y. Yuan, W. Wang, and H. Lu</i>	
RankFP: A Framework for Supporting Rank Formulation and Processing	514
<i>H. Yu, S. Hwang, and K. C.-C. Chang</i>	

Evaluation of Spatio-temporal Predicates on Moving Objects	516
<i>M. Schneider</i>	

Towards Exploring Interactive Relationship between Clusters and Outliers in Multi-dimensional Data Analysis	518
<i>Y. Shi and A. Zhang</i>	

Research Session 16: Privacy and Security

Privacy and Ownership Preserving of Outsourced Medical Data.....	521
<i>E. Bertino, B. C. Ooi, Y. Yang, and R. H. Deng</i>	

Configurable Security Protocols for Multi-party Data Analysis with Malicious Participants.....	533
<i>B. Malin, E. Airoidi, S. Edoho-Eket, and Y. Li</i>	

Privacy-Preserving Top- <i>k</i> Queries.....	545
<i>J. Vaidya and C. Clifton</i>	

Secure Third Party Distribution of XML Data	547
<i>B. Carminati, E. Ferrari, and E. Bertino</i>	

Improving Performance of Cluster-Based Secure Application Servers with User-Level Communication	549
<i>J.-H. Kim, G. S. Choi, and C. R. Das</i>	

Research Session 17: Data Warehousing

GPIVOT: Efficient Incremental Maintenance of Complex ROLAP Views	552
<i>S. Chen and E. A. Rundensteiner</i>	

Optimizing ETL Processes in Data Warehouses	564
<i>A. Simitsis, P. Vassiliadis, and T. Sellis</i>	

PnP: Parallel and External Memory Iceberg Cube Computation.....	576
<i>Y. Chen, F. Dehne, T. Eavis, and A. Rau-Chaplin</i>	

Cost-Driven General Join View Maintenance over Distributed Data Sources	578
<i>B. Liu and E. A. Rundensteiner</i>	

Mining Evolving Customer-Product Relationships in Multi-dimensional Space.....	580
<i>X. Li, J. Han, X. Yin, and D. Xin</i>	

Research Session 18: Text Processing

Bootstrapping Semantic Annotation for Content-Rich HTML Documents.....	583
<i>S. Mukherjee, I. V. Ramakrishnan, and A. Singh</i>	

Text Classification without Labeled Negative Documents.....	594
<i>G. P. C. Fung, J. X. Yu, H. Lu, and P. S. Yu</i>	

Modeling and Managing Content Changes in Text Databases	606
<i>P. G. Ipeirotis, A. Ntoulas, J. Cho, and L. Gravano</i>	

Research Session 19: Nearest Neighbors

Fast Approximate Similarity Search in Extremely High-Dimensional Data Sets.....	619
<i>M. E. Houle and J. Sakuma</i>	
Monitoring k -Nearest Neighbor Queries over Moving Objects.....	631
<i>X. Yu, K. Q. Pu, and N. Koudas</i>	
SEA-CNN: Scalable Processing of Continuous K-Nearest Neighbor Queries in Spatio-temporal Databases	643
<i>X. Xiong, M. F. Mokbel, and W. G. Aref</i>	

Research Session 20: Data and Dynamics

Architecture and Performance of Application Networking in Pervasive Content Delivery	656
<i>M. Su and C.-H. Chi</i>	
A Multiresolution Symbolic Representation of Time Series.....	668
<i>V. Megalooikonomou, Q. Wang, G. Li, and C. Faloutsos</i>	
Venn Sampling: A Novel Prediction Technique for Moving Objects	680
<i>Y. Tao, D. Papadias, J. Zhai, and Q. Li</i>	

Research Session 21: XML Query Processing (II)

XML Views as Integrity Constraints and Their Use in Query Translation	693
<i>R. Krishnamurthy, R. Kaushik, and J. F. Naughton</i>	
Full-Fledged Algebraic XPath Processing in Natix.....	705
<i>M. Brantner, S. Helmer, C.-C. Kanne, and G. Moerkotte</i>	
RDF Aggregate Queries and Views.....	717
<i>E. Hung, Y. Deng, and V. S. Subrahmanian</i>	

Research Session 22: Indexing and Data Aggregation

Maintaining Implicated Statistics in Constrained Environments.....	730
<i>Y. Sismanis and N. Roussopoulos</i>	
On the Signature Trees and Balanced Signature Trees.....	742
<i>Y. Chen</i>	
Index Support for Frequent Itemset Mining in a Relational DBMS	754
<i>E. Baralis, T. Cerquitelli, and S. Chiusano</i>	

Research Session 23: Distributed Streams

Finding (Recently) Frequent Items in Distributed Data Streams.....	767
<i>A. Manjhi, V. Shkapenyuk, K. Dhamdhere, and C. Olston</i>	
High-Availability Algorithms for Distributed Stream Processing.....	779
<i>J.-H. Hwang, M. Balazinska, A. Rasin, U. Çetintemel, M. Stonebraker, and S. Zdonik</i>	

Dynamic Load Distribution in the Borealis Stream Processor	791
<i>Y. Xing, S. Zdonik, and J.-H. Hwang</i>	

Research Session 24: Storage Systems

Deep Store: An Archival Storage System Architecture	804
<i>L. L. You, K. T. Pollack, and D. D. E. Long</i>	
QoS Mig: Adaptive Rate-Controlled Migration of Bulk Data in Storage Systems	816
<i>K. Dasgupta, S. Ghosal, R. Jain, U. Sharma, and A. Verma</i>	
Adaptive Overlapped Declustering: A Highly Available Data-Placement Method Balancing Access Load and Space Utilization	828
<i>A. Watanabe and H. Yokota</i>	

Research Session 25: Enhancing Usability

Personalized Queries under a Generalized Preference Model	841
<i>G. Koutrika and Y. Ioannidis</i>	
Increasing the Accuracy and Coverage of SQL Progress Indicators	853
<i>G. Luo, J. F. Naughton, C. J. Ellmann, and M. W. Watzke</i>	
Robust Identification of Fuzzy Duplicates	865
<i>S. Chaudhuri, V. Ganti, and R. Motwani</i>	

Research Session 26: XML Filtering

Cache-Conscious Automata for XML Filtering	878
<i>B. He, Q. Luo, and B. Choi</i>	
Bloom Filter-Based XML Packets Filtering for Millions of Path Queries	890
<i>X. Gong, W. Qian, Y. Yan, and A. Zhou</i>	
Batched Processing for Information Filters	902
<i>P. M. Fischer and D. Kossmann</i>	

INDUSTRIAL PAPER SESSIONS

Industrial Paper Session 1: Query Processing and Optimization

Uncovering Database Access Optimizations in the Middle Tier with TORPEDO	916
<i>B. E. Martin</i>	
Load and Network Aware Query Routing for Information Integration	927
<i>W.-S. Li, V. S. Batra, V. Raman, W. Han, K. S. Candan, and I. Narang</i>	
Predicate Derivation and Monotonicity Detection in DB2 UDB	939
<i>T. Malkemus, S. Padmanabhan, B. Bhattacharjee, L. Cranston, T. Lai, and F. Koo</i>	

Industrial Paper Session 2: DB Research in Japanese Industry

Acceleration Technique of Snake-Shaped Regions Retrieval Method for Telematics Navigation Service System.....	949
<i>M. Tanizaki, K. Maruyama, and S. Shimada</i>	
TRMeister: A DBMS with High-Performance Full-Text Search Functions	958
<i>T. Ikeda, H. Mano, H. Itoh, H. Takegawa, T. Hiraoka, S. Horibe, and Y. Ogawa</i>	
Distributed XML Stream Filtering System with High Scalability	968
<i>H. Uchiyama, M. Onizuka, and T. Honishi</i>	
Network-Based Problem Detection for Distributed Systems	978
<i>H. Kashima, T. Tsumura, T. Idé, T. Nogayama, R. Hirade, H. Etoh, and T. Fukuda</i>	

Industrial Paper Session 3: New System Functions for DBMS

Towards an Industrial Strength SQL/XML Infrastructure	991
<i>M. Krishnaprasad, Z. H. Liu, A. Manikuttu, J. W. Warner, and V. Arora</i>	
Distributed/Heterogeneous Query Processing in Microsoft SQL Server.....	1001
<i>J. A. Blakeley, C. Cunningham, N. Ellis, B. Rathakrishnan, and M.-C. Wu</i>	
Extending Relational Database Systems to Automatically Enforce Privacy Policies.....	1013
<i>R. Agrawal, P. Bird, T. Grandison, J. Kiernan, S. Logan, and W. Rjaibi</i>	

Industrial Paper Session 4: DBMS Implementation

Design, Implementation, and Evaluation of a Repairable Database Management System.....	1024
<i>T. Chiueh and D. Pilania</i>	
NFM: An Inter-domain Network Fault Management System	1036
<i>Q. Jiang, R. Adaikkalavan, and S. Chakravarthy</i>	
SVL: Storage Virtualization Engine Leveraging DBMS Technology.....	1048
<i>L. Qiao, B. R. Iyer, D. Agrawal, and A. El Abbadi</i>	

Industrial Paper Session 5: New Applications for Databases

Knowledge Discovery from Transportation Network Data	1061
<i>W. Jiang, J. Vaidya, Z. Balaporia, C. Clifton, and B. Banich</i>	
Sentiment Mining in WebFountain	1073
<i>J. Yi and W. Niblack</i>	
iBOM: A Platform for Intelligent Business Operation Management.....	1084
<i>M. Castellanos, F. Casati, M.-C. Shan, and U. Dayal</i>	

DEMO SESSIONS

Demo Session 1: Web Search and Query Processing

Towards Building a MetaQuerier: Extracting and Matching Web Query Interfaces.....	1098
<i>B. He, Z. Zhang, and K. C.-C. Chang</i>	
DCbot: Exploring the Web as Value-Added Service for Location-Based Applications	1100
<i>M. Jakob, M. Grossmann, N. Hönle, and D. Nicklas</i>	
RelaxImage: A Cross-Media Meta-Search Engine for Searching Images from Web Based on Query Relaxation	1102
<i>A. Kuwabara and K. Tanaka</i>	
Odysseus: A High-Performance ORDBMS Tightly-Coupled with IR Features.....	1104
<i>K.-Y. Whang, M.-J. Lee, J.-G. Lee, M.-S. Kim, and W.-S. Han</i>	
Querying and Visualizing Gridded Datasets for e-Science	1106
<i>B. Howe and D. Maier</i>	

Demo Session 2: Advanced DB and Data Models

Scrutinizing Frequent Pattern Discovery Performance.....	1109
<i>O. R. Zaïane, M. El-Hajj, Y. Li, and S. Luk</i>	
ModelGen: Model Independent Schema Translation	1111
<i>P. Atzeni, P. Cappellari, and P. A. Bernstein</i>	
Adaptive Process Management with ADEPT2.....	1113
<i>M. Reichert, S. Rinderle, U. Kreher, and P. Dadam</i>	
SECONDO: An Extensible DBMS Platform for Research Prototyping and Teaching	1115
<i>R. H. Güting, V. Almeida, D. Ansorge, T. Behr, Z. Ding, T. Höse, F. Hoffmann, M. Spiekermann, and U. Telle</i>	

Demo Session 3: XML and Data Streams

ViteX: A Streaming XPath Processing System	1118
<i>Y. Chen, S. B. Davidson, and Y. Zheng</i>	
The XML Stream Query Processor SPEX.....	1120
<i>F. Bry, F. Coskun, S. Durmaz, T. Furche, D. Olteanu, and M. Spannagel</i>	
Constructing and Querying Peer-to-Peer Warehouses of XML Resources	1122
<i>S. Abiteboul, I. Manolescu, and N. Preda</i>	
XGuard: A System for Publishing XML Documents without Information Leakage in the Presence of Data Inference	1124
<i>X. Yang, C. Li, G. Yu, and L. Shi</i>	
Online Latent Variable Detection in Sensor Networks.....	1126
<i>J. Sun, S. Papadimitriou, and C. Faloutsos</i>	

Demo Session 4: Spatial and Temporal DB

Spatiotemporal Annotation Graph (STAG): A Data Model for Composite Digital Objects	1129
<i>S. Yamini and A. Gupta</i>	
MoDB: Database System for Synthesizing Human Motion	1131
<i>T. Edmunds, S. Muthukrishnan, S. Sadhukhan, and S. Sueda</i>	
An Enhanced Query Model for Soccer Video Retrieval Using Temporal Relationships	1133
<i>S.-C. Chen, M.-L. Shyu, and N. Zhao</i>	
Spatio-temporal Databases in Practice: Directly Supporting Previously Developed Land Data Using Tripod	1135
<i>T. Griffiths, A. A. A. Fernandes, N. W. Paton, S.-H. Jeon, N. Djafri, and K. T. Mason</i>	
Fuzzy Spatial Objects: An Algebra Implementation in SECONDO	1137
<i>T. Behr and R. H. Güting</i>	

PANELS

Paradigm Shift to New DBMS Architectures: Research Issues and Market Needs.....	1140
<i>S. K. Cha, A. Ailamaki, Y. Hara, and V. Sikka</i>	
Panel on Business Process Intelligence	1141
<i>M. Castellanos and F. Casati</i>	

ADVANCED SEMINARS

XQuery Midflight: Emerging Database-Oriented Paradigms and a Classification of Research Advances	1143
<i>I. Manolescu and Y. Papakonstantinou</i>	
Rank-Aware Query Processing and Optimization	1144
<i>I. F. Ilyas and W. G. Aref</i>	
Data Stream Query Processing	1145
<i>N. Koudas and D. Srivastava</i>	
Online Mining of Data Streams: Applications, Techniques and Progress	1146
<i>H. Wang, J. Pei, and P. S. Yu</i>	
Web Services and Service-Oriented Architectures	1147
<i>G. Alonso and F. Casati</i>	
Database Architectures for New Hardware	1148
<i>A. Ailamaki</i>	

Data Mining Techniques for Microarray Datasets.....	1149
<i>L. Liu, J. Yang, and A. K. H. Tung</i>	
Author Index	1151