

# **2009 IEEE 25th International Conference on Data Engineering (ICDE 2009)**

**Shanghai, China  
29 March – 2 April 2009**

**Pages 1-591**



**IEEE Catalog Number: CFP09026-PRT  
ISBN: 978-1-4244-3422-0**

# IEEE International Conference on Data Engineering

## ICDE 2009

### Table of Contents

Message from the General Co-chairs

Letter from the Program Chairs

Organization Committee

Program Committee

External Reviewers

Sponsors

---

#### Invited Speakers

Search Computing .....	1
<i>Stefano Ceri</i>	
Why Can't I Find My Data the Way I Find My Dinner? .....	4
<i>David Carlson</i>	
Data Management in the Cloud .....	5
<i>Raghu Ramakrishnan</i>	

#### Research Session 1: Data Privacy

Modeling and Integrating Background Knowledge in Data Anonymization .....	6
<i>Tiancheng Li, Ninghui Li, and Jian Zhang</i>	
Deriving Private Information from Association Rule Mining Results .....	18
<i>Zutao Zhu, Guan Wang, and Wenliang Du</i>	
Light-Weight, Runtime Verification of Query Sources .....	30
<i>Tingjian Ge and Stan Zdonik</i>	
Auditing a Database under Retention Restrictions .....	42
<i>Wentian Lu and Gerome Miklau</i>	

#### Research Session 2: Web I

ApproxRank: Estimating Rank for a Subgraph .....	54
<i>Yao Wu and Louiqa Raschid</i>	
BinRank: Scaling Dynamic Authority-Based Search Using Materialized SubGraphs .....	66
<i>Heasoo Hwang, Andrey Balmin, Berthold Reinwald, and Erik Nijkamp</i>	

AJAX Crawl: Making AJAX Applications Searchable .....	78
<i>Cristian Duda, Gianni Frey, Donald Kossmann, Reto Matter, and Chong Zhou</i>	
A Data Structure for Sponsored Search .....	90
<i>Arnd Christian König, Kenneth Church, and Martin Markov</i>	

### **Research Session 3: Streams I**

A Framework for Clustering Massive-Domain Data Streams .....	102
<i>Charu Aggarwal</i>	
Self-Tuning, Bandwidth-Aware Monitoring for Dynamic Data Streams .....	114
<i>Navendu Jain, Praveen Yalagandula, Mike Dahlin, and Yin Zhang</i>	
Towards Efficient Processing of General-Purpose Joins in Sensor Networks .....	126
<i>Mirco Stern, Erik Buchmann, and Klemens Böhm</i>	
Forward Decay: A Practical Time Decay Model for Streaming Systems .....	138
<i>Graham Cormode, Vladislav Shkapenyuk, Divesh Srivastava, and Bojian Xu</i>	

### **Research Session 4: Data Security and Provenance**

Unified Declarative Platform for Secure Networked Information Systems .....	150
<i>Wenchao Zhou, Yun Mao, Boon Thau Loo, and Martín Abadi</i>	
An Architecture for Regulatory Compliant Database Management .....	162
<i>Soumyadeb Mitra, Marianne Winslett, Richard T. Snodgrass, Shashank Yaduvanshi, and Sumedh Ambekar</i>	
Perm: Processing Provenance and Data on the Same Data Model through Query Rewriting .....	174
<i>Boris Glavic and Gustavo Alonso</i>	
Join Optimization of Information Extraction Output: Quality Matters! .....	186
<i>Alpa Jain, Panagiotis G. Ipeirotis, AnHai Doan, and Luis Gravano</i>	

### **Research Session 5: Web II**

STAIRS: Towards Efficient Full-Text Filtering and Dissemination in a DHT Environment .....	198
<i>Weixiong Rao, Ada Wai-Chee Fu, Lei Chen, and Hanhua Chen</i>	
iVA-File: Efficiently Indexing Sparse Wide Tables in Community Systems .....	210
<i>Boduo Li, Mei Hui, Jianzhong Li, and Hong Gao</i>	
SP <sup>2</sup> Bench: A SPARQL Performance Benchmark .....	222
<i>Michael Schmidt, Thomas Hornung, Georg Lausen, and Christoph Pinkel</i>	
Weighted Proximity Best-Joins for Information Retrieval .....	234
<i>Risi Thonangi, Hao He, Anhai Doan, Haixun Wang, and Jun Yang</i>	

### **Research Session 6: Streams II**

Access Methods for Markovian Streams .....	246
<i>Julie Letchner, Christopher Re, Magdalena Balazinska, and Matthai Philipose</i>	
Keyword Search over Dynamic Categorized Information .....	258
<i>Manish Bhide, Venkatesan T. Chakaravarthy, Krithi Ramamritham, and Prasan Roy</i>	

Shared Winner Determination in Sponsored Search Auctions .....	270
<i>David J. Martin and Joseph Y. Halpern</i>	
Deductive Framework for Programming Sensor Networks .....	281
<i>Himanshu Gupta, Xianjin Zhu, and Xiang Xu</i>	

## Research Session 7: Data Uncertainty

Histograms and Wavelets on Probabilistic Data .....	293
<i>Graham Cormode and Minos Garofalakis</i>	
Semantics of Ranking Queries for Probabilistic Data and Expected Ranks .....	305
<i>Graham Cormode, Feifei Li, and Ke Yi</i>	
Ranking with Uncertain Scores .....	317
<i>Mohamed A. Soliman and Ihab F. Ilyas</i>	
Leveraging COUNT Information in Sampling Hidden Databases .....	329
<i>Arjun Dasgupta, Nan Zhang, and Gautam Das</i>	

## Research Session 8: Transactions

A Robust Technique to Ensure Serializable Executions with Snapshot Isolation DBMS .....	341
<i>Mohammad Alomari, Alan Fekete, and Uwe Röhm</i>	
Transaction Support for Log-Based Middleware Server Recovery .....	353
<i>Rui Wang, Betty Salzberg, and David Lomet</i>	
Adaptive Scheduling of Web Transactions .....	357
<i>Shenoda Guirguis, Mohamed A. Sharaf, Panos K. Chrysanthis, Alexandros Labrinidis, and Kirk Pruhs</i>	
BP-Wrapper: A System Framework Making Any Replacement Algorithms (Almost) Lock Contention Free .....	369
<i>Xiaoning Ding, Song Jiang, and Xiaodong Zhang</i>	

## Research Session 9: Streams and Interface

Sketching Sampled Data Streams .....	381
<i>Florin Rusu and Alin Dobra</i>	
Continuous Subgraph Pattern Search over Graph Streams .....	393
<i>Changliang Wang and Lei Chen</i>	
Top-k Exploration of Query Candidates for Efficient Keyword Search on Graph-Shaped (RDF) Data .....	405
<i>Thanh Tran, Haofen Wang, Sebastian Rudolph, and Philipp Cimiano</i>	
A Spreadsheet Algebra for a Direct Data Manipulation Query Interface .....	417
<i>Bin Liu and H. V. Jagadish</i>	

## Research Session 10: Data Mining I

Using Anonymized Data for Classification .....	429
<i>Ali Inan, Murat Kantarcioglu, and Elisa Bertino</i>	
Decision Trees for Uncertain Data .....	441
<i>Smith Tsang, Ben Kao, Kevin Y. Yip, Wai-Shing Ho, and Sau Dan Lee</i>	

Finding Time-Lagged 3D Clusters .....	445
<i>Xin Xu, Ying Lu, Kian-Lee Tan, and Anthony K. H. Tung</i>	
Contextual Ranking of Keywords Using Click Data .....	457
<i>Utku Irmak, Vadim von Brzeski, and Reiner Kraft</i>	

## Research Session 11: Query Optimization

Power Hints for Query Optimization .....	469
<i>Nicolas Bruno, Surajit Chaudhuri, and Ravi Ramamurthy</i>	
Double Index NEsted-Loop Reactive Join for Result Rate Optimization .....	481
<i>Mihaela A. Bornea, Vasilis Vassalos, Yannis Kotidis, and Antonios Deligiannakis</i>	
Join Reordering by Join Simulation .....	493
<i>Chaitanya Mishra and Nick Koudas</i>	
PSALM: Cardinality Estimation in the Presence of Fine-Grained Access Controls .....	505
<i>Huaxin Zhang, Ihab F. Ilyas, and Kenneth Salem</i>	

## Research Session 12: XML

Effective XML Keyword Search with Relevance Oriented Ranking .....	517
<i>Zhifeng Bao, Tok Wang Ling, Bo Chen, and Jiaheng Lu</i>	
Distributed Structural Relaxation of XPath Queries .....	529
<i>Georgia Koloniari and Evaggelia Pitoura</i>	
Sketch-Based Summarization of Ordered XML Streams .....	541
<i>Veronica Mayorga and Neoklis Polyzotis</i>	
Flexible XML Querying Using Skyline Semantics .....	553
<i>Sara Cohen and Maayan Shiloach</i>	

## Research Session 13: System Performance

Efficient Distribution of Full-Fledged XQuery .....	565
<i>Ying Zhang, Nan Tang, and Peter Boncz</i>	
X-CSR: Dataflow Optimization for Distributed XML Process Pipelines .....	577
<i>Daniel Zinn, Shawn Bowers, Timothy McPhillips, and Bertram Ludäscher</i>	
Improving Transaction-Time DBMS Performance and Functionality .....	581
<i>David B. Lomet and Feifei Li</i>	
Predicting Multiple Metrics for Queries: Better Decisions Enabled by Machine Learning .....	592
<i>Archana Ganapathi, Harumi Kuno, Umeshwar Dayal, Janet L. Wiener, Armando Fox, Michael Jordan, and David Patterson</i>	

## Research Session 14: Query Processing I

Space-Constrained Gram-Based Indexing for Efficient Approximate String Search .....	604
<i>Alexander Behm, Shengyue Ji, Chen Li, and Jiaheng Lu</i>	
Exploring a Few Good Tuples from Text Databases .....	616
<i>Alpa Jain and Divesh Srivastava</i>	

Confidence-Aware Join Algorithms .....	628
<i>Parag Agrawal and Jennifer Widom</i>	
SPROUT: Lazy vs. Eager Query Plans for Tuple-Independent Probabilistic Databases .....	640
<i>Dan Olteanu, Jiewen Huang, and Christoph Koch</i>	

## **Research Session 15: Spatial Queries**

Distance Oracles for Spatial Networks .....	652
<i>Jagan Sankaranarayanan and Hanan Samet</i>	
Reverse Furthest Neighbors in Spatial Databases .....	664
<i>Bin Yao, Feifei Li, and Piyush Kumar</i>	
Spatial Range Querying for Gaussian-Based Imprecise Query Objects .....	676
<i>Yoshiharu Ishikawa, Yuichi Iijima, and Jeffrey Xu Yu</i>	
Keyword Search in Spatial Databases: Towards Searching by Document .....	688
<i>Dongxiang Zhang, Yeow Meng Chee, Anirban Mondal, Anthony K. H. Tung, and Masaru Kitsuregawa</i>	

## **Research Session 16: Social Networking**

Routing Questions to the Right Users in Online Communities .....	700
<i>Yanhong Zhou, Gao Cong, Bin Cui, Christian S. Jensen, and Junjie Yao</i>	
On Efficient Recommendations for Online Exchange Markets .....	712
<i>Zeinab Abbassi and Laks V. S. Lakshmanan</i>	
Querying Communities in Relational Databases .....	724
<i>Lu Qin, Jeffrey Xu Yu, Lijun Chang, and Yufei Tao</i>	
Instant Advertising in Mobile Peer-to-Peer Networks .....	736
<i>Zaiben Chen, Heng Tao Shen, Quanqing Xu, and Xiaofang Zhou</i>	

## **Research Session 17: Query Processing II**

On Efficient Query Processing of Stream Counts on the Cell Processor .....	748
<i>Dina Thomas, Rajesh Bordawekar, Charu C. Aggarwal, and Philip S. Yu</i>	
Parallel Skyline Computation on Multicore Architectures .....	760
<i>Sungwoo Park, Taekyung Kim, Jonghyun Park, Jinha Kim, and Hyeonseung Im</i>	
Minimizing Communication Cost in Distributed Multi-query Processing .....	772
<i>Jian Li, Amol Deshpande, and Samir Khuller</i>	
Sequence Pattern Query Processing over Out-of-Order Event Streams .....	784
<i>Mo Liu, Ming Li, Denis Golovnya, Elke A. Rundensteiner, and Kajal Claypool</i>	

## **Research Session 18: Scientific and Advanced Applications**

Computing Distance Histograms Efficiently in Scientific Databases .....	796
<i>Yi-Cheng Tu, Shaoping Chen, and Sagar Pandit</i>	
Differencing Provenance in Scientific Workflows .....	808
<i>Zhuowei Bao, Sarah Cohen-Boulakia, Susan B. Davidson, Anat Eyal, and Sanjeev Khanna</i>	
XOntoRank: Ontology-Aware Search of Electronic Medical Records .....	820
<i>Fernando Farfan, Vagelis Hristidis, Anand Ranganathan, and Michael Weiner</i>	

Supporting Database Applications as a Service .....	832
<i>Mei Hui, Dawei Jiang, Guoliang Li, and Yuan Zhou</i>	

## Research Session 19: Data Mining II

GraphSig: A Scalable Approach to Mining Significant Subgraphs in Large Graph Databases .....	844
<i>Sayan Ranu and Ambuj K. Singh</i>	
Context-Aware Object Connection Discovery in Large Graphs .....	856
<i>James Cheng, Yiping Ke, Wilfred Ng, and Jeffrey Xu Yu</i>	
STAR: Steiner-Tree Approximation in Relationship Graphs .....	868
<i>Gjergji Kasneci, Maya Ramanath, Mauro Sozio, Fabian M. Suchanek, and Gerhard Weikum</i>	
A Latent Topic Model for Complete Entity Resolution .....	880
<i>Liangcai Shu, Bo Long, and Weiyi Meng</i>	

## Research Session 20: Query Processing III

Distance-Based Representative Skyline .....	892
<i>Yufei Tao, Ling Ding, Xuemin Lin, and Jian Pei</i>	
Similarity Group-By .....	904
<i>Yasin N. Silva, Walid G. Aref, and Mohamed H. Ali</i>	
Top-k Set Similarity Joins .....	916
<i>Chuan Xiao, Wei Wang, Xuemin Lin, and Haichuan Shang</i>	
Best-Effort Top-k Query Processing Under Budgetary Constraints .....	928
<i>Michal Shmueli-Scheuer, Chen Li, Yosi Mass, Haggai Roitman, Ralf Schenkel, and Gerhard Weikum</i>	

## Research Session 21: Data Integration and Warehousing

Aggregate Query Answering under Uncertain Schema Mappings .....	940
<i>Avigdor Gal, Maria Vanina Martinez, Gerardo I. Simari, and V. S. Subrahmanian</i>	
Large-Scale Deduplication with Constraints Using Dedupalog .....	952
<i>Arvind Arasu, Christopher Ré, and Dan Suciu</i>	
Recommending Join Queries via Query Log Analysis .....	964
<i>Xiaoyan Yang, Cecilia M. Procopiuc, and Divesh Srivastava</i>	
Resolution-Aware Query Answering for Business Intelligence .....	976
<i>Yannis Sismanis, Ling Wang, Ariel Fuxman, Peter J. Haas, and Berthold Reinwald</i>	

## Research Session 22: Data Mining III

Another Outlier Bites the Dust: Computing Meaningful Aggregates in Sensor Networks .....	988
<i>Antonios Deligiannakis, Yannis Kotidis, Vasilis Vassalos, Vassilis Stoumpos, and Alex Delis</i>	
Online Anomaly Prediction for Robust Cluster Systems .....	1000
<i>Xiaohui Gu and Haixun Wang</i>	

Fa: A System for Automating Failure Diagnosis .....	1012
<i>Songyun Duan, Shivnath Babu, and Kamesh Munagala</i>	
Efficient Mining of Closed Repetitive Gapped Subsequences from a Sequence Database .....	1024
<i>Bolin Ding, David Lo, Jiawei Han, and Siau-Cheng Khoo</i>	

## Research Session 23: Query Processing IV

Online Interval Skyline Queries on Time Series .....	1036
<i>Bin Jiang and Jian Pei</i>	
Efficient Processing of Warping Time Series Join of Motion Capture Data .....	1048
<i>Yueguo Chen, Gang Chen, Ke Chen, and Beng Chin Ooi</i>	
Probabilistic Skyline Operator over Sliding Windows .....	1060
<i>Wenjie Zhang, Xuemin Lin, Ying Zhang, Wei Wang, and Jeffrey Xu Yu</i>	
Topologically Sorted Skylines for Partially Ordered Domains .....	1072
<i>Dimitris Sacharidis, Stavros Papadopoulos, and Dimitris Papadias</i>	

## Research Session 24: Mobility

Supporting Generic Cost Models for Wide-Area Stream Processing .....	1084
<i>Olga Papaemmanouil, Uğur Çetintemel, and John Jannotti</i>	
Probabilistic Inference over RFID Streams in Mobile Environments .....	1096
<i>Thanh Tran, Charles Sutton, Richard Cocci, Yanming Nie, Yanlei Diao, and Prashant Shenoy</i>	
Recursive Computation of Regions and Connectivity in Networks .....	1108
<i>Mengmeng Liu, Nicholas E. Taylor, Wenchao Zhou, Zachary G. Ives, and Boon Thau Loo</i>	
Non-Exposure Location Anonymity .....	1120
<i>Haibo Hu and Jianliang Xu</i>	

## Short Research Papers

Privacy Preserving Publishing on Multiple Quasi-identifiers .....	1132
<i>Jian Pei, Yufei Tao, Jiexing Li, and Xiaokui Xiao</i>	
FF-Anonymity: When Quasi-identifiers Are Missing .....	1136
<i>Ke Wang, Yabo Xu, Ada W. C. Fu, and Raymond C. W. Wong</i>	
Outsourcing Search Services on Private Spatial Data .....	1140
<i>Man Lung Yiu, Gabriel Ghinita, Christian S. Jensen, and Panos Kalnis</i>	
Processing Group Nearest Group Query .....	1144
<i>Ke Deng, Hu Xu, Shazia Sadiq, Yansheng Lu, Gabriel Pui Cheong Fung, and Heng Tao Shen</i>	
Separating Authentication from Query Execution in Outsourced Databases .....	1148
<i>Stavros Papadopoulos, Dimitris Papadias, Weiwei Cheng, and Kian-Lee Tan</i>	
On High Dimensional Projected Clustering of Uncertain Data Streams .....	1152
<i>Charu C. Aggarwal</i>	



Expressive Location-Based Continuous Query Evaluation with Binary Decision Diagrams .....	1155
<i>Zhengdao Xu and Hans-Arno Jacobsen</i>	
Self Organizing Semantic Topologies in P2P Data Integration Systems .....	1159
<i>Ami Eyal and Avigdor Gal</i>	
Reachability Indexes for Relational Keyword Search .....	1163
<i>Alexander Markowetz, Yin Yang, and Dimitris Papadias</i>	
XQuery Join Graph Isolation: Celebrating 30+ Years of XQuery Processing Technology .....	1167
<i>Torsten Grust, Manuel Mayr, and Jan Rittinger</i>	
Distinct Counting with a Self-Learning Bitmap .....	1171
<i>Aiyou Chen and Jin Cao</i>	
Holistic Query Transformations for Dynamic Web Applications .....	1175
<i>Amit Manjhi, Charles Garrod, Bruce M. Maggs, Todd C. Mowry, and Anthony Tomasic</i>	
A Concise Representation of Range Queries .....	1179
<i>Ke Yi, Xiang Lian, Feifei Li, and Lei Chen</i>	
Progressive Keyword Search in Relational Databases .....	1183
<i>Guoliang Li, Xiaofang Zhou, Jianhua Feng, and Jianyong Wang</i>	
An Incremental Threshold Method for Continuous Text Search Queries .....	1187
<i>Kyriakos Mouratidis and HweeHwa Pang</i>	
A Subspace Symbolization Approach to Content-Based Video Search .....	1191
<i>Xiangmin Zhou, Xiaofang Zhou, Athman Bouguettaya, and John A. Taylor</i>	
Evaluating TOP-K Queries over Business Processes .....	1195
<i>Daniel Deutch and Tova Milo</i>	
Scalable Keyword Search on Large Data Streams .....	1199
<i>Lu Qin, Jeffrey Xu Yu, Lijun Chang, and Yufei Tao</i>	
Visible Reverse k-Nearest Neighbor Queries .....	1203
<i>Yunjun Gao, Baihua Zheng, Gencai Chen, Wang-Chien Lee, Ken C. K. Lee, and Qing Li</i>	
Scheduling Updates in a Real-Time Stream Warehouse .....	1207
<i>Lukasz Golab, Theodore Johnson, and Vladislav Shkapenyuk</i>	
AURA: Enabling Attribute-Based Spatial Search in RFID Rich Environments .....	1211
<i>Tejas A. Bapat, K. Selçuk Candan, V. Snehith Cherukuri, and Hari Sundaram</i>	
Web Monitoring 2.0: Crossing Streams to Satisfy Complex Data Needs .....	1215
<i>Haggai Roitman, Avigdor Gal, and Louiqa Raschid</i>	
Exploiting Domain Knowledge to Improve Biological Significance of Biclusters with Key Missing Genes .....	1219
<i>Jin Chen, Liping Ji, Wynne Hsu, Kian-Lee Tan, and Seung Y. Rhee</i>	
On the Efficiency of Provenance Queries .....	1223
<i>Anastasios Kementsietsidis and Min Wang</i>	
Using Semantics for Speech Annotation of Images .....	1227
<i>Chaitanya Desai, Dmitri V. Kalashnikov, Sharad Mehrotra, and Nalini Venkatasubramanian</i>	

Discovering Conditional Functional Dependencies .....	1231
<i>Wenfei Fan, Floris Geerts, Laks V. S. Lakshmanan, and Ming Xiong</i>	
Integrating and Ranking Uncertain Scientific Data .....	1235
<i>Landon Detwiler, Wolfgang Gatterbauer, Brent Louie, Dan Suciu, and Peter Tarczy-Hornoch</i>	
Adaptive Multi-join Query Processing in PDBMS .....	1239
<i>Sai Wu, Quang Hieu Vu, Jianzhong Li, and Kian-Lee Tan</i>	
Finding Equivalent Rewritings with Exact Views .....	1243
<i>Michael Compton</i>	
Uncertain Location Based Range Aggregates in a Multi-dimensional Space .....	1247
<i>Ying Zhang, Xuemin Lin, Yufei Tao, and Wenjie Zhang</i>	
Efficient Evaluation of Multiple Preference Queries .....	1251
<i>Leong Hou U, Nikos Mamoulis, and Kyriakos Mouratidis</i>	
iBroker: An Intelligent Broker for Ontology Based Publish/Subscribe Systems .....	1255
<i>Myung-Jae Park and Chin-Wan Chung</i>	
Extracting Web Query Interfaces Based on Form Structures and Semantic Similarity .....	1259
<i>Jun Hong, Zhongtian He, and David A. Bell</i>	
Efficient Indices Using Graph Partitioning in RDF Triple Stores .....	1263
<i>Ying Yan, Chen Wang, Aoying Zhou, Weining Qian, Li Ma, and Yue Pan</i>	
Privacy-Preserving Singular Value Decomposition .....	1267
<i>Shuguo Han, Wee Keong Ng, and Philip S. Yu</i>	
OPAQUE: Protecting Path Privacy in Directions Search .....	1271
<i>Ken C. K. Lee, Wang-Chien Lee, Hong Va Leong, and Baihua Zheng</i>	
Metric Functional Dependencies .....	1275
<i>Nick Koudas, Avishek Saha, Divesh Srivastava, and Suresh Venkatasubramanian</i>	
A General Proximity Privacy Principle .....	1279
<i>Ting Wang, Shicong Meng, Bhuvan Bamba, Ling Liu, and Calton Pu</i>	
Efficient Private Record Linkage .....	1283
<i>Mohamed Yakout, Mikhail J. Atallah, and Ahmed Elmagarmid</i>	
BioNav: Effective Navigation on Query Results of Biomedical Databases .....	1287
<i>Abhijith Kashyap, Vagelis Hristidis, Michalis Petropoulos, and Sotiria Tavoulari</i>	
Efficient Table Anonymization for Aggregate Query Answering .....	1291
<i>Cecilia M. Procopiuc and Divesh Srivastava</i>	
Max Regional Aggregate over Sensor Networks .....	1295
<i>Yongzhen Zhuang and Lei Chen</i>	
Recommendation Diversification Using Explanations .....	1299
<i>Cong Yu, Laks V. S. Lakshmanan, and Sihem Amer-Yahia</i>	
Tree Indexing on Flash Disks .....	1303
<i>Yinan Li, Bingsheng He, Qiong Luo, and Ke Yi</i>	
A Model for Data Leakage Detection .....	1307
<i>Panagiotis Papadimitriou and Hector Garcia-Molina</i>	
Scalability for Virtual Worlds .....	1311
<i>Nitin Gupta, Alan Demers, Johannes Gehrke, Philipp Unterbrunner, and Walker White</i>	

Efficient Query Evaluation over Temporally Correlated Probabilistic Streams .....	1315
<i>Bhargav Kanagal and Amol Deshpande</i>	
Temporal Outlier Detection in Vehicle Traffic Data .....	1319
<i>Xiaolei Li, Zhenhui Li, Jiawei Han, and Jae-Gil Lee</i>	
CoTS: A Scalable Framework for Parallelizing Frequency Counting over Data Streams .....	1323
<i>Sudipto Das, Shyam Antony, Divyakant Agrawal, and Amr El Abbadi</i>	
Concept Clustering of Evolving Data .....	1327
<i>Shixi Chen, Haixun Wang, and Shuigeng Zhou</i>	
Dynamic Approaches to In-network Aggregation .....	1331
<i>Oliver Kennedy, Christoph Koch, and Al Demers</i>	
Chameleon: Context-Awareness inside DBMSs .....	1335
<i>Hicham G. Elmongui, Walid G. Aref, and Mohamed F. Mokbel</i>	

## **Industrial Session 1**

Query Rewrites with Views for XML in DB2 .....	1339
<i>P. Godfrey, J. Gryz, A. Hoppe, W. Ma, and C. Zuzarte</i>	
A Decade of XML Data Management: An Industrial Experience Report from Oracle .....	1351
<i>Zhen Hua Liu and Ravi Murthy</i>	
Oracle Streams: A High Performance Implementation for Near Real Time Asynchronous Replication .....	1363
<i>Lik Wong, Nimar S. Arora, Lei Gao, Thuvan Hoang, and Jingwei Wu</i>	
Scale-Up Strategies for Processing High-Rate Data Streams in System S .....	1375
<i>Henrique Andrade, Buğra Gedik, Kun-Lung Wu, and Philip S. Yu</i>	

## **Industrial Session 2**

Design and Evaluation of the iMed Intelligent Medical Search Engine .....	1379
<i>Gang Luo</i>	
Business Intelligence from Voice of Customer .....	1391
<i>L. Venkata Subramaniam, Tanveer A. Faruque, Shajith Iqbal, Shantanu Godbole, and Mukesh K. Mohania</i>	
A Static Analysis Framework for Database Applications .....	1403
<i>Arjun Dasgupta, Vivek Narasayya, and Manoj Syamala</i>	
An Incremental Knowledge Acquisition Method for Improving Duplicate Invoices Detection .....	1415
<i>Van Hai Ho, Paul Compton, Boualem Benatallah, Julien Vayssière, Lucio Menzel, and Hartmut Vogler</i>	

## **Industrial Session 3**

Finding Actionable Knowledge via Automated Comparison .....	1419
<i>Lei Zhang, Bing Liu, Jeffrey Benkler, and Chi Zhou</i>	

Updates in the AquaLogic Data Services Platform .....	1431
<i>Michael Blow, Vinayak Borkar, Michael Carey, Christopher Hillery,</i>	
<i>Alexander Kotopoulos, Dmitry Lychagin, Radu Preotiuc-Pietro, Panagiotis Reveliotis,</i>	
<i>Joshua Spiegel, and Till Westmann</i>	
Web Query Recommendation via Sequential Query Prediction .....	1443
<i>Qi He, Daxin Jiang, Zhen Liao, Steven C. H. Hoi, Kuiyu Chang, Ee-Peng Lim, and Hang Li</i>	
rFEED: A Mixed Workload Scheduler for Enterprise Data Warehouses .....	1455
<i>Abhay Mehta, Chetan Gupta, Song Wang, and Umesh Dayal</i>	

## **Demostration Session 1**

An Internet-Scale Service for Publishing and Locating XML Documents .....	1459
<i>Praveen Rao and Bongki Moon</i>	
DynaCet: Building Dynamic Faceted Search Systems over Databases .....	1463
<i>Senjuti Basu Roy, Haidong Wang, Ullas Nambiar, Gautam Das, and Mukesh Mohania</i>	
Flexible Recommendations for Course Planning .....	1467
<i>Georgia Koutrika, Benjamin Bercovitz, Robert Ikeda, Filip Kaliszan, Henry Liou,</i>	
<i>and Hector Garcia-Molina</i>	
GuruMine: A Pattern Mining System for Discovering Leaders and Tribes .....	1471
<i>Amit Goyal, Byung-Won On, Francesco Bonchi, and Laks V. S. Lakshmanan</i>	
I3DC: Interactive Three-Dimensional Cubes .....	1475
<i>Ke Yang, Yinan Li, Qiong Luo, Pedro V. Sander, and Jiaoying Shi</i>	
MatchUp: Autocompletion for Mashups .....	1479
<i>Serge Abiteboul, Ohad Greenshpan, Tova Milo, and Neoklis Polyzotis</i>	
SmallBlue: Social Network Analysis for Expertise Search and Collective Intelligence .....	1483
<i>Ching-Yung Lin, Nan Cao, Shi Xia Liu, Spiros Papadimitriou, Jimeng Sun, and Xifeng Yan</i>	
WEST: Modern Technologies for Web People Search .....	1487
<i>Dmitri V. Kalashnikov, Zhaoqi Chen, Rabia Nuray-Turan, Sharad Mehrotra,</i>	
<i>and Zheng Zhang</i>	
WISE: A Workflow Information Search Engine .....	1491
<i>Qihong Shao, Peng Sun, and Yi Chen</i>	

## **Demostration Session 2**

ACStream: Enforcing Access Control over Data Streams .....	1495
<i>Jianneng Cao, Barbara Carminati, Elena Ferrari, and Kian-Lee Tan</i>	
Automated Diagnosis of System Failures with Fa .....	1499
<i>Songyun Duan and Shivnath Babu</i>	
KSpot: Effectively Monitoring the K Most Important Events in a Wireless Sensor Network .....	1503
<i>Panayiotis Andreou, Demetrios Zeinalipour-Yazti, Martha Vassiliadou,</i>	
<i>Panos K. Chrysanthis, and George Samaras</i>	

Environmental Monitoring 2.0 .....	1507
<i>Sebastian Michel, Ali Salehi, Liqian Luo, Nicholas Dawes, Karl Aberer,</i>	
<i>Guillermo Barrenetxea, Mathias Bavay, Aman Kansal, K. Ashwin Kumar, Suman Nath,</i>	
<i>Marc Parlange, Stewart Tansley, Catharine van Ingen, Feng Zhao, and Yongluan Zhou</i>	
Online Near-Duplicate Video Clip Detection and Retrieval: An Accurate and Fast System .....	1511
<i>Zi Huang, Liping Wang, Heng Tao Shen, Jie Shao, and Xiaofang Zhou</i>	
PLUS: A Message-Efficient Prototype for Location-Based Applications .....	1515
<i>Yu-Ling Hsueh, Roger Zimmermann, Wei-Shinn Ku, Haojun Wang, and Chung-Dau Wang</i>	
V*-kNN: An Efficient Algorithm for Moving k Nearest Neighbor Queries .....	1519
<i>Sarana Nutanong, Rui Zhang, Egemen Tanin, and Lars Kulik</i>	

### **Demonstration Session 3**

The PRISM Workbench: Database Schema Evolution without Tears .....	1523
<i>Carlo A. Curino, Hyun J. Moon, MyungWon Ham, and Carlo Zaniolo</i>	
SbQA: A Self-Adaptable Query Allocation Process .....	1527
<i>Jorge-Arnulfo Quiané-Ruiz, Philippe Lamarre, and Patrick Valduriez</i>	
FRISK: Keyword Query Cleaning and Processing in Action .....	1531
<i>Ken Q. Pu and Xiaohui Yu</i>	
XBLEND: Visual XML Query Formulation Meets Query Processing .....	1535
<i>Zhou Yong, Sourav S. Bhowmick, Erwin Leonardi, and K. G. Widjanarko</i>	
Shaman: A Self-Healing Database System .....	1539
<i>Songyun Duan, Peter Franklin, Vamsidhar Thummala, Dongdong Zhao,</i>	
<i>and Shivnath Babu</i>	
QuickStart: An Upfront Client-Based Design Advisor for Parallel Data Warehouses .....	1543
<i>Malu Castellanos, Ivo Jimenez, Neal Coddington, Hans Zeller, Steven Whang,</i>	
<i>and Umeshwar Dayal</i>	
Efficient and Robust Database Support for Data-Intensive Applications in Dynamic Environments .....	1547
<i>Jon Olav Hauglid, Kjetil Nørvåg, and Norvald H. Ryeng</i>	
Galaxy: Encouraging Data Sharing among Sources with Schema Variants .....	1551
<i>Peter Mork, Len Seligman, Arnon Rosenthal, Michael Morse, Chris Wolf, Jeff Hoyt,</i>	
<i>and Ken Smith</i>	
Hippocratic PostgreSQL .....	1555
<i>Jalaja Padma, Yasin N. Silva, Muhammad U. Arshad, and Walid G. Aref</i>	

### **ICDE 2009 Workshops**

## **DBRank: Third International Workshop on Ranking in Databases**

Weighted Set Similarity: Queries and Updates .....	1559
<i>Divesh Srivastava</i>	
Incremental Reverse Nearest Neighbor Ranking .....	1560
<i>Hans-Peter Kriegel, Peer Kröger, Matthias Renz, Andreas Züfle, and Alexander Katzdobler</i>	
Continuous Skylining on Volatile Moving Data .....	1568
<i>Mu-Woong Lee and Seung-won Hwang</i>	
Visualized Elucidations of Ranking by Exploiting Object Relations .....	1576
<i>Xinpeng Zhang, Yasuhito Asano, and Masatoshi Yoshikawa</i>	
Ranking of Object Summaries .....	1580
<i>Georgios John Fakas and Zhi Cai</i>	
Improving the Effectiveness of XML Retrieval with User Navigation Models .....	1584
<i>M. S. Ali, Mariano P. Consens, and Bassam Helou</i>	

## **M3SN: Modeling, Managing, and Mining of Evolving Social Networks**

Jelly: A Language for Building Community-Centric Information Exploration Applications .....	1588
<i>Sihem Amer-Yahia, Jian Huang, and Cong Yu</i>	
SoQL: A Language for Querying and Creating Data in Social Networks .....	1595
<i>Royi Ronen and Oded Shmueli</i>	
On Protecting Private Information in Social Networks: A Proposal .....	1603
<i>Bo Luo and Dongwon Lee</i>	
Contents-Based Analysis of Community Formation and Evolution in Blogspace .....	1607
<i>Seok-Chul Baek, Sukwon Kang, Hyung Noh, and Sang-Wook Kim</i>	
Actively Building Private Recommender Networks for Evolving Reliable Relationships .....	1611
<i>Ira Assent</i>	
Social Streams Blog Crawler .....	1615
<i>Matthew Hurst and Alexey Maykov</i>	

## **MOUND: Management and Mining of Uncertain Data**

Preprocessing Uncertain User Profile Data: Inferring User's Actual Age from Ages of the User's Neighbors .....	1619
<i>Sung Hyuk Park, Sang Pil Han, Soon Young Huh, and Hojin Lee</i>	
Clustering Uncertain Data with Possible Worlds .....	1625
<i>Peter Benjamin Volk, Frank Rosenthal, Martin Hahmann, Dirk Habich, and Wolfgang Lehner</i>	
A Rule-Based Classification Algorithm for Uncertain Data .....	1633
<i>Biao Qin, Yuni Xia, Sunil Prabhakar, and Yicheng Tu</i>	

Tracking High Quality Clusters over Uncertain Data Streams .....	1641
<i>Chen Zhang, Ming Gao, and Aoying Zhou</i>	
Holistically Twig Matching in Probabilistic XML .....	1649
<i>Yawen Li, Guoren Wang, Junchang Xin, Ende Zhang, and Zeling Qiu</i>	
Effective Feature Selection on Data with Uncertain Labels .....	1657
<i>Bo Wang, Yan Jia, Yi Han, and Weihong Han</i>	
Mining of Frequent Itemsets from Streams of Uncertain Data .....	1663
<i>Carson Kai-Sang Leung and Boyu Hao</i>	

## **SMDB 2009: Fourth International Workshop on Self-Managing Database Systems**

An Integrated Approach to Performance Monitoring for Autonomous Tuning .....	1671
<i>Alexander Thiem and Kai-Uwe Sattler</i>	
Online Tuning of Aggregation Tables for OLAP .....	1679
<i>Katja Hose, Daniel Klan, and Kai-Uwe Sattler</i>	
An Economic Model for Self-Tuned Cloud Caching .....	1687
<i>Debabrata Dash, Verena Kantere, and Anastasia Ailamaki</i>	
Self-Tuning for SQL Performance in Oracle Database 11g .....	1694
<i>Peter Belknap, Benoit Dageville, Karl Dias, and Khaled Yagoub</i>	
A Benchmark for Online Index Selection .....	1701
<i>Karl Schnaitter and Neoklis Polyzotis</i>	

## **WISS'09: First IEEE Workshop on Information & Software as Services**

### **WISS'09: Keynote Paper**

Database Management as a Service: Challenges and Opportunities .....	1709
<i>Divyakant Agrawal, Amr El Abbadi, Fatih Emekci, and Ahmed Metwally</i>	

### **WISS'09: Regular Papers**

Support Multi-version Applications in SaaS via Progressive Schema Evolution .....	1717
<i>Jianfeng Yan and Bo Zhang</i>	
Adaptive Parallelization of Queries over Dependent Web Service Calls .....	1725
<i>Manivasakan Sabesan and Tore Risch</i>	
Towards Composition as a Service - A Quality of Service Driven Approach .....	1733
<i>Florian Rosenberg, Philipp Leitner, Anton Michlmayr, Predrag Celikovic, and Schahram Dustdar</i>	
Universal Resource Lifecycle Management .....	1741
<i>Marcos Báez, Fabio Casati, and Maurizio Marchese</i>	
Secure Enterprise Services Consumption for SaaS Technology Platforms .....	1749
<i>Yuecel Karabulut and Ike Nassi</i>	

## **WISS'09: Case Study/Panel - Extended Abstracts**

Study of Web Delivered Services Support Platform .....	1757
<i>Changjie Guo</i>	
Design Patterns for Internet-Scale Services .....	1759
<i>Jinquan Dai</i>	
SAP Business ByDesign .....	1760
<i>Yuebo Zhou</i>	

## **WISS'09: Workshop Theme/Area Survey Paper**

Frontiers in Information and Software as Services .....	1761
<i>K. Selçuk Candan, Wen-Syan Li, Thomas Phan, and Minqi Zhou</i>	

### **Author Index**