2009 IEEE 25th International Conference on Data Engineering

(ICDE 2009)

Shanghai, China
29 March – 2 April 2009

Pages 1-591
Message from the General Co-chairs
Letter from the Program Chairs
Organization Committee
Program Committee
External Reviewers
Sponsors

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

Research Session 1: Data Privacy
Modeling and Integrating Background Knowledge in Data Anonymization ........................................6
  Tiancheng Li, Ninghui Li, and Jian Zhang
Deriving Private Information from Association Rule Mining Results ......................................................18
  Zutao Zhu, Guan Wang, and Wenliang Du
Light-Weight, Runtime Verification of Query Sources .........................................................................30
  Tingjian Ge and Stan Zdonik
Auditing a Database under Retention Restrictions ..............................................................................42
  Wentian Lu and Gerome Miklau

Research Session 2: Web I
ApproxRank: Estimating Rank for a Subgraph ....................................................................................54
  Yao Wu and Louiza Raschid
BinRank: Scaling Dynamic Authority-Based Search Using Materialized SubGraphs .............................66
  Heasoo Hwang, Andrey Balmin, Berthold Reinwald, and Erik Nijkamp
AJAX Crawl: Making AJAX Applications Searchable .................................................................78
  Cristian Duda, Gianni Frey, Donald Kossmann, Reto Matter, and Chong Zhou

A Data Structure for Sponsored Search ...................................................................................90
  Arnd Christian König, Kenneth Church, and Martin Markov

Research Session 3: Streams I

A Framework for Clustering Massive-Domain Data Streams ......................................................102
  Charu Aggarwal

Self-Tuning, Bandwidth-Aware Monitoring for Dynamic Data Streams .......................................114
  Navendu Jain, Praveen Yalagandula, Mike Dahlin, and Yin Zhang

Towards Efficient Processing of General-Purpose Joins in Sensor Networks ..............................126
  Mirco Stern, Erik Buchmann, and Klemens Böhm

Forward Decay: A Practical Time Decay Model for Streaming Systems ......................................138
  Graham Cormode, Vladislav Shkapenyuk, Divesh Srivastava, and Bojian Xu

Research Session 4: Data Security and Provenance

Unified Declarative Platform for Secure Networked Information Systems ..................................150
  Wenchao Zhou, Yun Mao, Boon Thau Loo, and Martin Abadi

An Architecture for Regulatory Compliant Database Management ...........................................162
  Soumyadeb Mitra, Marianne Winslett, Richard T. Snodgrass, Shashank Yaduvanshi, and
  Sumedh Ambokar

Perm: Processing Provenance and Data on the Same Data Model through Query Rewriting ........174
  Boris Glavic and Gustavo Alonso

Join Optimization of Information Extraction Output: Quality Matters! .......................................186
  Alpa Jain, Panagiotis G. Ipeirotis, AnHai Doan, and Luis Gravano

Research Session 5: Web II

STAIRS: Towards Efficient Full-Text Filtering and Dissemination in a DHT Environment ............198
  Weixiong Rao, Ada Wai-Chee Fu, Lei Chen, and Hanhua Chen

iVA-File: Efficiently Indexing Sparse Wide Tables in Community Systems .............................210
  Boduo Li, Mei Hui, Jianzhong Li, and Hong Gao

SP\(^2\)Bench: A SPARQL Performance Benchmark .....................................................................222
  Michael Schmidt, Thomas Hornung, Georg Lausen, and Christoph Pinkel

Weighted Proximity Best-Joins for Information Retrieval ............................................................234
  Risi Thonangi, Hao He, Anhai Doan, Haixun Wang, and Jun Yang

Research Session 6: Streams II

Access Methods for Markovian Streams ....................................................................................246
  Julie Letchner, Christopher Re, Magdalena Balazinska, and Matthai Philipose

Keyword Search over Dynamic Categorized Information ............................................................258
  Manish Bhide, Venkatesan T. Chakaravarthy, Krithi Ramamritham, and Prasan Roy
Shared Winner Determination in Sponsored Search Auctions .................................................................270
  *David J. Martin and Joseph Y. Halpern*

Deductive Framework for Programming Sensor Networks ....................................................................281
  *Himanshu Gupta, Xianjin Zhu, and Xiang Xu*

**Research Session 7: Data Uncertainty**

Histograms and Wavelets on Probabilistic Data ..................................................................................293
  *Graham Cormode and Minos Garofalakis*

Semantics of Ranking Queries for Probabilistic Data and Expected Ranks .............................................305
  *Graham Cormode, Feifei Li, and Ke Yi*

Ranking with Uncertain Scores .............................................................................................................317
  *Mohamed A. Soliman and Ihab F. Ilyas*

Leveraging COUNT Information in Sampling Hidden Databases .......................................................329
  *Arjun Dasgupta, Nan Zhang, and Gautam Das*

**Research Session 8: Transactions**

A Robust Technique to Ensure Serializable Executions with Snapshot Isolation DBMS ....................341
  *Mohammad Alomari, Alan Fekete, and Uwe Röhm*

Transaction Support for Log-Based Middleware Server Recovery ....................................................353
  *Rui Wang, Betty Salzberg, and David Lomet*

Adaptive Scheduling of Web Transactions ..........................................................................................357
  *Shenoda Guirguis, Mohamed A. Sharaf, Panos K. Chrysanthis, Alexandros Labrinidis, and Kirkl Pruhns*

BP-Wrapper: A System Framework Making Any Replacement Algorithms (Almost) .......................369
  *Xiaoning Ding, Song Jiang, and Xiaodong Zhang*

**Research Session 9: Streams and Interface**

Sketching Sampled Data Streams ........................................................................................................381
  *Florin Rusu and Alin Dobra*

Continuous Subgraph Pattern Search over Graph Streams ....................................................................393
  *Changliang Wang and Lei Chen*

Top-k Exploration of Query Candidates for Efficient Keyword Search on Graph-Shaped (RDF) Data .........................................................................................................................405
  *Thanh Tran, Haofen Wang, Sebastian Rudolph, and Philipp Cimiano*

A Spreadsheet Algebra for a Direct Data Manipulation Query Interface ............................................417
  *Bin Liu and H. V. Jagadish*

**Research Session 10: Data Mining I**

Using Anonymized Data for Classification ............................................................................................429
  *Ali Inan, Murat Kantarcioğlu, and Elisa Bertino*

Decision Trees for Uncertain Data ........................................................................................................441
  *Smith Tsang, Ben Kao, Kevin Y. Yip, Wai-Shing Ho, and Sau Dan Lee*
Research Session 11: Query Optimization

Power Hints for Query Optimization ................................................................. 469
Nicolas Bruno, Surajit Chaudhuri, and Ravi Ramamurthy

Double Index Nested-Loop Reactive Join for Result Rate Optimization ............. 481
Mihaela A. Bornea, Vasilis Vassalos, Yannis Kotidis, and Antonios Deligiannakis

Join Reordering by Join Simulation .................................................................... 493
Chaitanya Mishra and Nick Koudas

PSALM: Cardinality Estimation in the Presence of Fine-Grained Access Controls ................................................................. 505
Huaxin Zhang, Ihab F. Ilyas, and Kenneth Salem

Research Session 12: XML

Effective XML Keyword Search with Relevance Oriented Ranking ......................... 517
Zhifeng Bao, Tok Wang Ling, Bo Chen, and Jiaheng Lu

Distributed Structural Relaxation of XPath Queries .............................................. 529
Georgia Koloniari and Evaggelia Pitoura

Sketch-Based Summarization of Ordered XML Streams ....................................... 541
Veronica Mayorga and Neoklis Polyzotis

Flexible XML Querying Using Skyline Semantics .............................................. 553
Sara Cohen and Maayan Shiloach

Research Session 13: System Performance

Efficient Distribution of Full-Fledged XQuery ..................................................... 565
Ying Zhang, Nan Tang, and Peter Boncz

X-CSR: Dataflow Optimization for Distributed XML Process Pipelines .................... 577
Daniel Zinn, Shawn Bowers, Timothy McPhillips, and Bertram Ludäscher

Improving Transaction-Time DBMS Performance and Functionality .................. 581
David B. Lomet and Feifei Li

Predicting Multiple Metrics for Queries: Better Decisions Enabled by Machine Learning .............................................................................................................. 592
Archana Ganapathi, Harumi Kuno, Umeshwar Dayal, Janet L. Wiener, Armando Fox,
Michael Jordan, and David Patterson

Research Session 14: Query Processing I

Space-Constrained Gram-Based Indexing for Efficient Approximate String Search ................................................................. 604
Alexander Behm, Shengyue Ji, Chen Li, and Jiaheng Lu

Exploring a Few Good Tuples from Text Databases ........................................... 616
Alpa Jain and Divesh Srivastava
Confidence-Aware Join Algorithms .................................................................628
   Parag Agrawal and Jennifer Widom
SPROUT: Lazy vs. Eager Query Plans for Tuple-Independent Probabilistic Databases ........................................640
   Dan Olteanu, Jiwen Huang, and Christoph Koch

Research Session 15: Spatial Queries

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

Research Session 16: Social Networking

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

Research Session 17: Query Processing II

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

Research Session 18: Scientific and Advanced Applications

Computing Distance Histograms Efficiently in Scientific Databases ...............796
   Yi-Cheng Tu, Shaoping Chen, and Sagar Pandit
Differencing Provenance in Scientific Workflows .......................................808
   Zhuowei Bao, Sarah Cohen-Boulakia, Susan B. Davidson, Anat Eyal, and Sanjeev Khanna
XOntoRank: Ontology-Aware Search of Electronic Medical Records .............820
   Fernando Farfan, Vagelis Hristidis, Anand Ranganathan, and Michael Weiner
Supporting Database Applications as a Service ................................................................. 832

Mei Hui, Dawei Jiang, Guoliang Li, and Yuan Zhou

Research Session 19: Data Mining II

GraphSig: A Scalable Approach to Mining Significant Subgraphs in Large Graph Databases ................................................................................................................................. 844

Sayan Ranu and Ambuj K. Singh

Context-Aware Object Connection Discovery in Large Graphs ............................................. 856

James Cheng, Yiping Ke, Wilfred Ng, and Jeffrey Xu Yu

STAR: Steiner-Tree Approximation in Relationship Graphs .................................................. 868

Gjergji Kasneci, Maya Ramanath, Mauro Sozio, Fabian M. Suchanek, and Gerhard Weikum

A Latent Topic Model for Complete Entity Resolution ......................................................... 880

Liangcai Shu, Bo Long, and Weiyi Meng

Research Session 20: Query Processing III

Distance-Based Representative Skyline .................................................................................. 892

Yufei Tao, Ling Ding, Xuemin Lin, and Jian Pei

Similarity Group-By .................................................................................................................. 904

Yasin N. Silva, Walid G. Aref, and Mohamed H. Ali

Top-k Set Similarity Joins ........................................................................................................ 916

Chuan Xiao, Wei Wang, Xuemin Lin, and Haichuan Shang

Best-Effort Top-k Query Processing Under Budgetary Constraints .................................... 928

Michal Shmueli-Scheuer, Chen Li, Yosi Mass, Haggai Roitman, Ralf Schenkel, and Gerhard Weikum

Research Session 21: Data Integration and Warehousing

Aggregate Query Answering under Uncertain Schema Mappings .......................................... 940

Avigdor Gal, Maria Yanina Martinez, Gerardo I. Simari, and V. S. Subrahmanian

Large-Scale Deduplication with Constraints Using Dedupalog ............................................ 952

Arvind Arasu, Christopher Ré, and Dan Suciu

Recommending Join Queries via Query Log Analysis ............................................................ 964

Xiaoyan Yang, Cecilia M. Procopiuc, and Divesh Srivastava

Resolution-Aware Query Answering for Business Intelligence ............................................. 976

Yannis Sismanis, Ling Wang, Ariel Fuxman, Peter J. Haas, and Berthold Reinwald

Research Session 22: Data Mining III

Another Outlier Bites the Dust: Computing Meaningful Aggregates in Sensor Networks .......................................................... 988

Antonios Deligiannakis, Yannis Kotidis, Vasilis Vassalos, Vassilis Stoumpos, and Alex Delis

Online Anomaly Prediction for Robust Cluster Systems ....................................................... 1000

Xiaohui Gu and Haixun Wang
Fa: A System for Automating Failure Diagnosis .................................................................1012
Songyun Duan, Shivnath Babu, and Kamesh Munagala

Efficient Mining of Closed Repetitive Gapped Subsequences from a Sequence Database ...........................................................................................................................................................................1024
Bolin Ding, David Lo, Jiawei Han, and Siau-Cheng Khoo

Research Session 23: Query Processing IV

Online Interval Skyline Queries on Time Series ........................................................................1036
Bin Jiang and Jian Pei

Efficient Processing of Warping Time Series Join of Motion Capture Data ................................1048
Yueguo Chen, Gang Chen, Ke Chen, and Beng Chin Ooi

Probabilistic Skyline Operator over Sliding Windows ................................................................1060
Wenjie Zhang, Xuemin Lin, Ying Zhang, Wei Wang, and Jeffrey Xu Yu

Topologically Sorted Skylines for Partially Ordered Domains ..................................................1072
Dimitris Sacharidis, Stavros Papadopoulos, and Dimitris Papadis

Research Session 24: Mobility

Supporting Generic Cost Models for Wide-Area Stream Processing ........................................1084
Olga Papaemmanouil, Uğur Çetintemel, and John Jannotti

Probabilistic Inference over RFID Streams in Mobile Environments ....................................1096
Thanh Tran, Charles Sutton, Richard Cocci, Yanming Nie, Yanlei Diao, and Prashant Shenoy

Recursive Computation of Regions and Connectivity in Networks ..........................................1108
Mengmeng Liu, Nicholas E. Taylor, Wenchao Zhou, Zachary G. Ives, and Boon Thau Loo

Non-Exposure Location Anonymity ..........................................................................................1120
Haibo Hu and Jianliang Xu

Short Research Papers

Privacy Preserving Publishing on Multiple Quasi-identifiers ......................................................1132
Jian Pei, Yufei Tao, Jiexing Li, and Xiaokui Xiao

FF-Anonymity: When Quasi-identifiers Are Missing .................................................................1136
Ke Wang, Yabo Xu, Ada W. C. Fu, and Raymond C. W. Wong

Outsourcing Search Services on Private Spatial Data .................................................................1140
Man Lung Yu, Gabriel Ghinita, Christian S. Jensen, and Panos Kalnis

Processing Group Nearest Group Query ..................................................................................1144
Ke Deng, Hu Xu, Shazia Sadiq, Yansheng Lu, Gabriel Pui Cheong Fung, and Heng Tao Shen

Separating Authentication from Query Execution in Outsourced Databases .............................1148
Stavros Papadopoulos, Dimitris Papadis, Weiwei Cheng, and Kian-Lee Tan

On High Dimensional Projected Clustering of Uncertain Data Streams .....................................1152
Charu C. Aggarwal
Discovering Conditional Functional Dependencies .................................................................1231
    Wenfei Fan, Floris Geerts, Laks V. S. Lakshmanan, and Ming Xiong

Integrating and Ranking Uncertain Scientific Data ...............................................................1235
    Landon Detwiler, Wolfgang Gatterbauer, Brent Louie, Dan Suciu, and Peter Tarczy-Hornoch

Adaptive Multi-join Query Processing in PDBMS .................................................................1239
    Sai Wu, Quang Hieu Vu, Jianzhong Li, and Kian-Lee Tan

Finding Equivalent Rewritings with Exact Views ...............................................................1243
    Michael Compton

Uncertain Location Based Range Aggregates in a Multi-dimensional Space ......................1247
    Ying Zhang, Xuemin Lin, Yufei Tao, and Wenjie Zhang

Efficient Evaluation of Multiple Preference Queries .......................................................1251
    Leong Hou U, Nikos Mamoulis, and Kyriakos Mouratidis

iBroker: An Intelligent Broker for Ontology Based Publish/Subscribe Systems ..................1255
    Myung-Jae Park and Chin-Wan Chung

Extracting Web Query Interfaces Based on Form Structures and Semantic Similarity ..........1259
    Jun Hong, Zhongtian He, and David A. Bell

Efficient Indices Using Graph Partitioning in RDF Triple Stores ......................................1263
    Ying Yan, Chen Wang, Aoying Zhou, Weining Qian, Li Ma, and Yue Pan

Privacy-Preserving Singular Value Decomposition ...........................................................1267
    Shuguo Han, Wee Keong Ng, and Philip S. Yu

OPAQUE: Protecting Path Privacy in Directions Search ....................................................1271
    Ken C. K. Lee, Wang-Chien Lee, Hong Ya Leong, and Baihua Zheng

Metric Functional Dependencies .........................................................................................1275
    Nick Koudas, Avishek Saha, Divesh Srivastava, and Suresh Venkatasubramanian

A General Proximity Privacy Principle ............................................................................1279
    Ting Wang, Shicong Meng, Bhuvan Bamba, Ling Liu, and Calton Pu

Efficient Private Record Linkage .......................................................................................1283
    Mohamed Yakout, Mikhail J. Atallah, and Ahmed Elmagarmid

BioNav: Effective Navigation on Query Results of Biomedical Databases .........................1287
    Abhijith Kashyap, Vagelis Hristidis, Michalis Petropoulos, and Sotiria Tavoulari

Efficient Table Anonymization for Aggregate Query Answering .....................................1291
    Cecilia M. Procopiuc and Divesh Srivastava

Max Regional Aggregate over Sensor Networks ..................................................................1295
    Yongzhen Zhuang and Lei Chen

Recommendation Diversification Using Explanations .......................................................1299
    Cong Yu, Laks V. S. Lakshmanan, and Sihem Amer-Yahia

Tree Indexing on Flash Disks .............................................................................................1303
    Yinan Li, Bingsheng He, Qiong Luo, and Ke Yi

A Model for Data Leakage Detection ..................................................................................1307
    Panagiotis Papadimitriou and Hector Garcia-Molina

Scalability for Virtual Worlds .............................................................................................1311
    Nitin Gupta, Alan Demers, Johannes Gehrke, Philipp Unterbrunner, and Walker White
Efficient Query Evaluation over Temporally Correlated Probabilistic Streams .......................................................... 1315
  Bhargav Kanagal and Amol Deshpande
Temporal Outlier Detection in Vehicle Traffic Data ................................................................................................. 1319
  Xiaolei Li, Zhenhui Li, Jiawei Han, and Jae-Gil Lee
CoTS: A Scalable Framework for Parallelizing Frequency Counting over Data Streams ........................................ 1323
  Sudipto Das, Shyam Antony, Divyakant Agrawal, and Amr El Abbadi
Concept Clustering of Evolving Data ...................................................................................................................... 1327
  Shixi Chen, Haixun Wang, and Shuigeng Zhou
Dynamic Approaches to In-network Aggregation ....................................................................................................... 1331
  Oliver Kennedy, Christoph Koch, and Al Demers
Chameleon: Context-Awareness inside DBMSs ........................................................................................................... 1335
  Hicham G. Elmongui, Walid G. Aref, and Mohamed F. Mokbel

**Industrial Session 1**

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

**Industrial Session 2**

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

**Industrial Session 3**

Finding Actionable Knowledge via Automated Comparison ....................................................................................... 1419
  Lei Zhang, Bing Liu, Jeffrey Benkler, and Chi Zhou
### Updates in the AquaLogic Data Services Platform

*Micahel Blow, Vinayak Borkar, Michael Carey, Christopher Hillery, Alexander Kotopoulis, Dmitry Lychagin, Radu Preotiuc-Pietro, Panagiotis Reveliotis, Joshua Spiegel, and Till Westmann*

Web Query Recommendation via Sequential Query Prediction

*Qi He, Daxin Jiang, Zhen Liao, Steven C. H. Ho, Kuivy Chang, Ee-Peng Lim, and Hang Li*

rFEED: A Mixed Workload Scheduler for Enterprise Data Warehouses

*Abhay Mehta, Chetan Gupta, Song Wang, and Umesh Dayal*

---

### Demonstration Session 1

An Internet-Scale Service for Publishing and Locating XML Documents

*Praveen Rao and Bongki Moon*

DynaCet: Building Dynamic Faceted Search Systems over Databases

*Senjuti Basu Roy, Haidong Wang, Ullas Nambiar, Gautam Das, and Mukesh Mohania*

Flexible Recommendations for Course Planning

*Georgia Koutrika, Benjamin Bercovitz, Robert Ikeda, Filip Kaliszczan, Henry Liou, and Hector Garcia-Molina*

GuruMine: A Pattern Mining System for Discovering Leaders and Tribes

*Amit Goyal, Byung-Won On, Francesco Bonchi, and Laks V. S. Lakshmanan*

I3DC: Interactive Three-Dimensional Cubes

*Ke Yang, Yinan Li, Qiong Luo, Pedro V. Sander, and Jiaoying Shi*

MatchUp: Autocompletion for Mashups

*Serge Abiteboul, Ohad Greenshpan, Tova Milo, and Neoklis Polyzotis*

SmallBlue: Social Network Analysis for Expertise Search and Collective Intelligence

*Ching-Yung Lin, Nan Cao, Shi Xia Liu, Spiros Papadimitriou, Jimeng Sun, and Xifeng Yan*

WEST: Modern Technologies for Web People Search

*Dmitri V. Kalashnikov, Zhaoqi Chen, Rabia Nuray-Turan, Sharad Mehrotra, and Zheng Zhang*

WISE: A Workflow Information Search Engine

*Qihong Shao, Peng Sun, and Yi Chen*

---

### Demonstration Session 2

ACStream: Enforcing Access Control over Data Streams

*Jianneng Cao, Barbara Carminati, Elena Ferrari, and Kian-Lee Tan*

Automated Diagnosis of System Failures with Fa

*Songyun Duan and Shivnath Babu*

KSpot: Effectively Monitoring the K Most Important Events in a Wireless Sensor Network

*Panayiotis Andreou, Demetrios Zeinalipour-Yazti, Martha Vassiliadou, Panos K. Chrysanthis, and George Samaras*
Online Near-Duplicate Video Clip Detection and Retrieval: An Accurate and Fast System
Zi Huang, Liping Wang, Heng Tao Shen, Jie Shao, and Xiaofang Zhou
PLUS: A Message-Efficient Prototype for Location-Based Applications
Yu-Ling Hsueh, Roger Zimmermann, Wei-Shinn Ku, Haojun Wang, and Chung-Dau Wang
V*-kNN: An Efficient Algorithm for Moving k Nearest Neighbor Queries
Sarana Nutanong, Rui Zhang, Egemen Tanin, and Lars Kulik

Demonstration Session 3
The PRISM Workbench: Database Schema Evolution without Tears
Carlo A. Curino, Hyun J. Moon, MyungWoon Ham, and Carlo Zaniolo
SbQA: A Self-Adaptable Query Allocation Process
Jorge-Arnulfo Quiñé-Ruiz, Philippe Lamarre, and Patrick Valduriez
FRISK: Keyword Query Cleaning and Processing in Action
Ken Q. Pu and Xiaohui Yu
XBLEND: Visual XML Query Formulation Meets Query Processing
Zhou Yong, Sourav S. Bhowmick, Erwin Leonardi, and K. G. Widjanarko
Shaman: A Self-Healing Database System
Songyun Duan, Peter Franklin, Vamsidhar Thummala, Dongdong Zhao, and Shivnath Babu
QuickStart: An Upfront Client-Based Design Advisor for Parallel Data Warehouses
Malu Castellanos, Ivo Jimenez, Neal Coddington, Hans Zeller, Steven Whang, and Umeshwar Dayal
Efficient and Robust Database Support for Data-Intensive Applications in Dynamic Environments
Jon Olav Hauglid, Kjetil Norvåg, and Norvald H. Ryeng
Galaxy: Encouraging Data Sharing among Sources with Schema Variants
Peter Mork, Len Seligman, Arnon Rosenthal, Michael Morse, Chris Wolf, Jeff Hoyt, and Ken Smith
Hippocratic PostgreSQL
Jalaja Padma, Yasin N. Silva, Muhammad U. Arshad, and Walid G. Aref

ICDE 2009 Workshops
DBRank: Third International Workshop on Ranking in Databases

Weighted Set Similarity: Queries and Updates ...............................................................1559
Divesh Srivastava

Incremental Reverse Nearest Neighbor Ranking .........................................................1560
Hans-Peter Kriegel, Peer Kröger, Matthias Renz, Andreas Züfle, and Alexander Katzdobler

Continuous Skylining on Volatile Moving Data ..........................................................1568
Mu-Woong Lee and Seung-won Hwang

Visualized Elucidations of Ranking by Exploiting Object Relations ........................1576
Xinpeng Zhang, Yasuhiro Asano, and Masatoshi Yoshikawa

Ranking of Object Summaries .....................................................................................1580
Georgios John Fakas and Zhi Cai

Improving the Effectiveness of XML Retrieval with User Navigation Models ..........1584
M. S. Ali, Mariano P. Consens, and Bassam Helou

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

Jelly: A Language for Building Community-Centric Information Exploration
Applications ..................................................................................................................1588
Sihem Amer-Yahia, Jian Huang, and Cong Yu

SoQL: A Language for Querying and Creating Data in Social Networks ..................1595
Roiy Ronen and Oded Shmueli

On Protecting Private Information in Social Networks: A Proposal .......................1603
Bo Luo and Dongwon Lee

Contents-Based Analysis of Community Formation and Evolution in Blogspace ........1607
Seok-Chul Baek, Sukwon Kang, Hyung Noh, and Sang-Wook Kim

Actively Building Private Recommender Networks for Evolving Reliable Relationships .................................................................................................1611
Ira Assent

Social Streams Blog Crawler .......................................................................................1615
Matthew Hurst and Alexey Maykov

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
Sung Hyuk Park, Sang Pil Han, Soon Young Huh, and Hojin Lee

Clustering Uncertain Data with Possible Worlds .......................................................1625
Peter Benjamin Volk, Frank Rosenthal, Martin Hahmann, Dirk Habich, and Wolfgang Lehner

A Rule-Based Classification Algorithm for Uncertain Data .......................................1633
Biao Qin, Yuni Xia, Sunil Prabhakar, and Yicheng Tu
Tracking High Quality Clusters over Uncertain Data Streams .................................................................1641
  Chen Zhang, Ming Gao, and Aoying Zhou

Holistically Twig Matching in Probabilistic XML .........................................................................................1649
  Yawen Li, Guoren Wang, Junchang Xin, Ende Zhang, and Zeling Qiu

Effective Feature Selection on Data with Uncertain Labels ...........................................................................1657
  Bo Wang, Yan Jia, Yi Han, and Weihong Han

Mining of Frequent Itemsets from Streams of Uncertain Data .................................................................1663
  Carson Kai-Sang Leung and Boyu Hao

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

An Integrated Approach to Performance Monitoring for Autonomous Tuning ..............................................1671
  Alexander Thiem and Kai-Uwe Sattler

Online Tuning of Aggregation Tables for OLAP ..............................................................................................1679
  Katja Hose, Daniel Klan, and Kai-Uwe Sattler

An Economic Model for Self-Tuned Cloud Caching ......................................................................................1687
  Debabrata Dash, Verena Kantere, and Anastasia Ailamaki

Self-Tuning for SQL Performance in Oracle Database 11g ........................................................................1694
  Peter Belknap, Benoit Dageville, Karl Dias, and Khaled Yagoub

A Benchmark for Online Index Selection .......................................................................................................1701
  Karl Schnaitter and Neoklis Polyzotis

WISS’09: First IEEE Workshop on Information & Software as Services

WISS’09: Keynote Paper

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

WISS’09: Regular Papers

Support Multi-version Applications in SaaS via Progressive Schema Evolution ........................................1717
  Jianfeng Yan and Bo Zhang

Adaptive Parallelization of Queries over Dependent Web Service Calls ..................................................1725
  Manivasakan Sabesan and Tore Risch

Towards Composition as a Service - A Quality of Service Driven Approach ..........................................1733
  Florian Rosenberg, Philipp Letimer, Anton Michlmayr, Predrag Celikovic, and Schahram Dustdar

Universal Resource Lifecycle Management ....................................................................................................1741
  Marcos Báez, Fabio Casati, and Maurizio Marchese

Secure Enterprise Services Consumption for SaaS Technology Platforms .............................................1749
  Yuecel Karabulut and Ike Nassi
Study of Web Delivered Services Support Platform .................................................................1757
   Changjie Guo

Design Patterns for Internet-Scale Services ............................................................................1759
   Jinquan Dai

SAP Business ByDesign .............................................................................................................1760
   Yuebo Zhou

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

Author Index