

Proceedings

**13th International Conference on Management of Data
(COMAD 2006)**

December 14-16, 2006, Delhi, India

Editors

Laks V. S. Lakshmanan, Prasan Roy, and Anthony K. H. Tung

GOLD SPONSORS



SILVER SPONSORS



PERSISTENT



SUPPORT

Very
Large
Data
Bases

VLDB Endowment

Table of Contents

| | |
|---|------|
| Message from the Conference Chairs | vii |
| Conference Organization | viii |
| Program Committee | ix |

Keynote Talks

| | |
|--|---|
| Information Search in Peer-to-Peer Systems | 1 |
| <i>Gerhard Weikum (Max-Planck Institute for Informatics, Germany)</i> | |
| StrangerDB: Safe Data Management with Untrusted Servers | 2 |
| <i>Dennis Shasha (Courant Institute - New York University, USA)</i> | |
| Taming the Dynamics of Distributed Data | 3 |
| <i>Krithi Ramamritham (Indian Institute of Technology - Bombay, India)</i> | |

Research Session 1: XML Processing

| | |
|---|----|
| A Concise Labeling Scheme for XML Data | 4 |
| <i>Risi Thonangi (SETLabs, Infosys Technologies Ltd.)</i> | |
| Statistical Analysis of Real XML Data Collections | 15 |
| <i>Irena Mlynkova (Charles University, Czech Republic), Kamil Toman (Charles University, Czech Republic), Jaroslav Pokorny (Charles University, Czech Republic)</i> | |
| Semantic and Structure Based XML Similarity: An Integrated Approach | 27 |
| <i>Joe Tekli (University of Bourgogne, France), Richard Chbeir (University of Bourgogne, France), Kokou Yetongnon (University of Bourgogne, France)</i> | |

Research Session 2: Indexing and Similarity Search

| | |
|--|----|
| Using Relations to Index Biological Document Repositories for Efficient Searching | 39 |
| <i>Rohit Goyal (IIT Delhi, India), Lipika Dey (IIT Delhi, India)</i> | |
| Efficient Similarity Retrieval In Music Databases | 48 |
| <i>Maria M. Ruxanda (Aalborg University, Denmark), Christian S. Jensen (Aalborg University, Denmark)</i> | |
| Supporting Approximate Similarity Queries with Quality Guarantees in P2P Systems | 60 |
| <i>Qi Zhong (Microsoft Corporation, USA), Iosif Lazaridis (UC Irvine, USA), Mayur Deshpande (UC Irvine, USA), Chen Li (UC Irvine, USA), Sharad Mehrotra (UC Irvine, USA), Hal Stern (UC Irvine, USA)</i> | |

Research Session 3: Potpourri

| | |
|--|----|
| Preserving Obliviousness Characteristic of Honeypot database | 72 |
| <i>S. K. Gupta (IIT Delhi, India), Anand Gupta (NSIT Delhi, India), Renu Damor (IIT Delhi, India), Vikram Goyal (IIT Delhi, India), Sangeeta Sabharwal (NSIT Delhi, India)</i> | |
| Towards Kernel Density Estimation over Streaming Data | 80 |
| <i>Christoph Heinz (University of Marburg, Germany), Bernhard Seeger (University of Marburg, Germany)</i> | |
| Genea: Schema-Aware Mapping of Ontologies into Relational Databases | 92 |
| <i>Tim Kraska (University of Muenster, Germany), Uwe Roehm (University of Sydney, Australia)</i> | |

Research Session 4: Web and Distributed Data

| | |
|---|-----|
| Fault-Tolerant Queries over Sensor Data | 104 |
| <i>Iosif Lazaridis (UC Irvine, USA), Qi Han (Colorado School of Mines, USA), Sharad Mehrotra (UC Irvine, USA), Nalini Venkatasubramanian (UC Irvine, USA)</i> | |
| EcoRep: An Economic Model for efficient dynamic replication in Mobile-P2P networks | 116 |

| | |
|--|-----|
| <i>Anirban Mondal (University of Tokyo, Japan), Sanjay Madria (University of Missouri-Rolla, USA), Masaru Kitsuregawa (University of Tokyo, Japan)</i> | |
| Unsupervised Learning from URL Corpora..... | 128 |
| <i>Deepak P (IBM India Research Lab-SIRC, India), Deepak Khemani (IIT Madras, India)</i> | |
| Measures of Ignorance on the Web | 140 |
| <i>Siddhartha Reddy (IIIT Bangalore, India), Srinath Srinivasa (IIIT Bangalore, India), Mandar Mutalikdesai (IIIT Bangalore, India)</i> | |

Poster Session

| | |
|--|-----|
| Scheduling and Caching in MultiQuery Optimization | 150 |
| <i>Dilys Thomas (Stanford University, USA), A. A. Diwan (IIT Bombay, India), S. Sudarshan (IIT Bombay, India)</i> | |
| Efficient Algorithm for Hierarchical Online Mining of Association Rules | 154 |
| <i>Kishore B. Kumar (MindTree Consulting Private Limited, India), Naresh Jotwani (DA-IICT, India)</i> | |
| A Scalable Replica Management Method in Peer-to-Peer Distributed Storage Systems..... | 158 |
| <i>Jing Zhou (National U. of Defence Tech, China), Yijie Wang (National U. of Defence Tech, China), Sikun Li (National U. of Defence Tech, China)</i> | |
| Prefix Tree with Encryption of Data and Itemsets | 162 |
| <i>Ramkishore Bhattacharyya (Jadavpur University, India)</i> | |
| Anomaly Detection In Labeled Data..... | 166 |
| <i>Rohit Kelkar (Tata Research Development and Design Centre, India), Girish Palshikar (Tata Research Development and Design Centre, India)</i> | |
| Using Domain Ontologies for Efficient Information Retrieval..... | 170 |
| <i>Sandhya Revuri (IIT Madras, India), Sujatha R Upadhyaya (IIT Madras, India), P Sreenivasa Kumar (IIT Madras, India)</i> | |
| Managing XML data with Evolving Schema | 174 |
| <i>B. V. N. Prashant (IIT Madras, India), P. Sreenivasa Kumar (IIT Madras, India)</i> | |
| BasisGraph: Combining Storage and Structure Index for Similarity Search in Graph DBs | 178 |
| <i>Mistry Harjinder Singh (IIIT Bangalore, India), Srinath Srinivasa (IIIT Bangalore, India)</i> | |
| Algebra-Based Optimization of XML-Extended OLAP Queries..... | 182 |
| <i>Xuepeng Yin (Aalborg University, Denmark), Torben Bach Pedersen (Aalborg University, Denmark)</i> | |
| Symbiosis in the Intranet: How Document Retrieval Benefits from Database Information | 186 |
| <i>Christoph Mangold (Universitaet Stuttgart, Germany), Holger Schwarz (Universitaet Stuttgart, Germany), Bernhard Mitschang (Universitaet Stuttgart, Germany)</i> | |

Demonstration Session

| | |
|---|-----|
| Visual Clue Based Extraction of Web Data from Flat and Nested Data Records..... | 190 |
| <i>Siddu P Algur (SDM College of Engg, Dharwad, India), P S Hiremath (Gulbarga University, India)</i> | |
| A Query Interface for Ubiquitous Access to Database Resources | 194 |
| <i>Subhash Bhalla (University of Aizu, Japan), Masaki Hasegawa (University of Aizu, Japan)</i> | |
| UnURL: Unsupervised Learning from URLs | 198 |
| <i>Deepak P (IBM IRL, India), Deepak Khemani (IIT Madras, India)</i> | |
| OCHD: Preserving Obliviousness Characteristic of Honeypot Database | 202 |
| <i>S. K. Gupta (IIT Delhi, India), Renu Damor (IIT Delhi, India), Anand Gupta (NSIT, Delhi, India), Vikram Goyal (IIT Delhi, India)</i> | |
| Stream Mining via Density Estimators: A Concrete Application | 206 |
| <i>Christoph Heinz (University of Marburg, Germany), Bernhard Seeger (University of Marburg, Germany)</i> | |
| Documents meet Databases: A System for Intranet Search | 210 |
| <i>Christoph Mangold (University of Stuttgart, Germany), Holger Schwarz (University of Stuttgart, Germany)</i> | |
| Discovery Services-Enabling RFID Traceability in EPCglobal Networks | 214 |

| | |
|---|-----|
| <i>Steve Beier (IBM SVL, USA), Tyrone Grandison (IBM Almaden, USA), Karin Kailing (IBM Almaden, USA), Ralf Rantzau (IBM SVL, USA)</i> | |
| Semantic and Structure Based XML Similarity: The XS3 Prototype | 218 |
| <i>Joe Tekli (University of Bourgogne, France), Richard Chbeir (University of Bourgogne, France), Kokou Yetongnon (University of Bourgogne, France)</i> | |
| The OLAP-XML Federation System | 222 |
| <i>Xuepeng Yin (Aalborg University, Denmark), Torben Bach Pedersen (Aalborg University, Denmark)</i> | |
| BlogHarvest: Blog Mining and Search Framework | 226 |
| <i>Mukul Joshi (Great Software Lab, India), Nikhil Belsare (Tech Mahindra Ltd., India)</i> | |

Applications and Industrial Session

| | |
|--|-----|
| Efficient Detection of Distributed Constraint Violations | 230 |
| <i>Shipra Agrawal (Stanford University, USA), Supratim Deb (Bell Labs Research, India), K. V. M. Naidu (Bell Labs Research, India), Rajeev Rastogi (Bell Labs Research, India)</i> | |
| On Engineering Web-based Enterprise Applications | 240 |
| <i>Srinivasa Narayanan (Tavant Technologies, USA), Subbu N. Subramanian (Tavant Technologies, USA), Manish Arya (Tavant Technologies, USA)</i> | |

Tutorials

| | |
|--|-----|
| Privacy Preserving Data Publication: From Generalization to Anatomy | 245 |
| <i>Yufei Tao (Chinese University of Hong Kong, China)</i> | |
| Multilingual Database Systems | 246 |
| <i>Jayant Haritsa (Indian Institute of Science - Bangalore, India)</i> | |
| Secure Data Outsourcing..... | 247 |
| <i>Radu Sion (Stony Brook University, USA)</i> | |
| High Performance Data Mining: Consolidation and Renewed Bearing..... | 248 |
| <i>Srinivasan Parthasarathy (Ohio State University, USA)</i> | |
| Scalable Information Extraction and Integration..... | 249 |
| <i>Sunita Sarawagi (Indian Institute of Technology - Bombay, India)</i> | |
| Data Grid Management | 250 |
| <i>Arun Jagadeesan (SDSC - University of California at San Diego, USA)</i> | |

| | |
|---------------------------|-----|
| Author Index | 251 |
|---------------------------|-----|