

# VLDB2013

39<sup>th</sup> International Conference on Very Large Data Bases, Riva del Garda, Trento, Italy



## Proceedings of the VLDB Endowment

Volume 6, No. 3 – January 2013

**Proceedings of the 39th International Conference on  
Very Large Data Bases, Riva del Garda, Trento, Italy**

Editors-in-Chief:

**Michael Böhlen, Christoph Koch**

Associate Editors – Research Track:

**Ashraf Aboulnaga, Sihem Amer-Yahia, Chee Yong Chan, Yanlei Diao, Ada Waichee Fu, Johannes Gehrke,  
Alon Halevy, Jayant Haritsa, Nikos Mamoulis, Thomas Neumann, Dan Olteanu, Divesh Srivastava, Jens Teubner**

Associate Editor – Experiments and Analysis Track:

**Stefan Manegold**

Guest Editors:

**Ada Waichee Fu, Alon Halevy**

Proceedings Editors:

**Peer Kröger, Stratis D. Viglas**

PVLDB – Proceedings of the VLDB Endowment

Volume 6, No. 3, January 2013.

The 39th International Conference on Very Large Data Bases, Riva del Garda, Trento, Italy.

## **Copyright 2013 VLDB Endowment**

Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyright for components of this work owned by others than VLDB Endowment must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists requires prior specific permission and/or a fee. Request permission to republish from PVLDB under email: [info@vldb.org](mailto:info@vldb.org).

Volume 6, Number 3, January 2013: VLDB 2013

Pages ii - x and 157 - 228

ISSN 2150-8097

Additional copies only online at: [portal.acm.org](http://portal.acm.org), [arxiv.org/corr](http://arxiv.org/corr), and [www.vldb.org](http://www.vldb.org)

## TABLE OF CONTENTS

### Front Matter

Copyright Notice .....	ii
Table of Contents .....	iii
VLDB 2013 Organization and Review Board .....	iv

### Letters

Letter from the Guest Editors .....	<i>Ada Waichee Fu, Alon Halevy</i>	x
-------------------------------------	------------------------------------	---

### Research Papers

Lightweight Privacy-Preserving Peer-to-Peer Data Integration .....	157
..... <i>Ye Zhang, Wai-Kit Wong, S. M. Yiu, Nikos Mamoulis, David W. Cheung</i>	
Memory Efficient Minimum Substring Partitioning .....	169
..... <i>Yang Li, Pegah Kamousi, Fangqiu Han, Shengqi Yang, Xifeng Yan, Subhash Suri</i>	
NeMa: Fast Graph Search with Label Similarity.....	181
..... <i>Arijit Khan, Yinghui Wu, Charu C. Aggarwal, Xifeng Yan</i>	
PARAS: A Parameter Space Framework for Online Association Mining .....	193
..... <i>Xika Lin, Abhishek Mukherji, Elke A. Rundensteiner, Carolina Ruiz, Matthew O. Ward</i>	
Actively Soliciting Feedback for Query Answers in Keyword Search-Based Data Integration .....	205
..... <i>Zhepeng Yan, Nan Zheng, Zachary G. Ives, Partha Pratim Talukdar, Cong Yu</i>	
Spatial Keyword Query Processing: An Experimental Evaluation .....	217
..... <i>Lisi Chen, Gao Cong, Christian S. Jensen, Dingming Wu</i>	

## **VLDB 2013 ORGANIZATION AND REVIEW BOARD**

### **General Chairs**

Themis Palpanas, University of Trento  
Yannis Velegarakis, University of Trento

### **Program Chairs**

Michael Böhlen, University of Zurich  
Christoph Koch, EPFL

### **Advisory Board**

Paolo Atzeni, Università Roma Tre  
Stefano Ceri, Politecnico di Milano  
John Mylopoulos, University of Trento

### **Award Committee**

Surajit Chaudhuri, Microsoft (Chair)  
Mike Carey, University of California, Irvine  
Susan Davidson, University of Pennsylvania  
Alon Halevy, Google  
Sunita Sarawagi, IIT Bombay

### **Associate Editors**

Ada Wai-Chee Fu, Chinese University of Hong Kong  
Alon Halevy, Google  
Ashraf Aboulnaga, University of Waterloo  
Chee-Yong Chan, National University of Singapore  
Dan Olteanu, Oxford University  
Divesh Srivastava, AT&T Labs  
Jayant Haritsa, Indian Institute of Science Bangalore  
Jens Teubner, ETH Zurich  
Johannes Gehrke, Cornell University  
Nikos Mamoulis, University of Hong Kong  
Sihem Amer-Yahia, Qatar Computing Research Institute  
Stefan Manegold, CWI  
Thomas Neumann, Technische Universität München  
Yanlei Diao, University of Massachusetts Amherst

### **Experiments and Analysis Track Associate Editor**

Stefan Manegold, CWI

### **Industrial and Applications Track Associate Editors**

Min Wang, HP Labs China

Cong Yu, Google Research

### **Demonstration Chairs**

Jun Yang, Duke University

Dimitrios Gunopulos, University of Athens

Letizia Tanca, Politecnico di Milano

### **Reproducibility Chairs**

Philippe Bonnet, IT University of Copenhagen

Juliana Freire, New York University

Dennis Shasha, New York University

### **Research Track Review Board**

Karl Aberer, EPFL, Switzerland

Foto Afrati, NTU Athens

Charu Aggarwal, IBM T. J. Watson Research Center

Yanif Ahmad, JHU

Jose-Luis Ambite, University of Southern California

Walid Aref, Purdue University

Magdalena Balazinska, University of Washington

Srikanta Bedathur, IIIT Delhi

Peter Boncz, CWI

Nico Bruno, Microsoft

Randal Burns, JHU

Andrea Cali, University of London, Birkbeck College

Carlos Castillo, Yahoo!

Gang Chen, Zhejiang University

Lei Chen, Hong Kong University of Science and Technology

Shimin Chen, HP Labs China

James Cheng, CUHK

Reynold Cheng, University of Hong Kong

Gao Cong, Nanyang Technological University

Brian Cooper, Google

Bin Cui, Peking University

Carlo Curino, MIT

Sudipto Das, Microsoft Research

Anish Das Sarma, Google Research

Atish Das Sarma, eBay Research Labs

Antonios Deligiannakis, Technical University of Crete

Amol Deshpande, University of Maryland

Xin Luna Dong, AT&T Labs-Research

Sameh Elnikety, Microsoft Research

Mohamed Eltabakh, Worcester Polytechnic Institute

Alan Fekete, University of Sydney

Hakan Ferhatosmanoglu, Bilkent University

Alvaro Fernandes, U. of Manchester

Juliana Freire, New York University

Benjamin C. M. Fung, Concordia University

Fabien Gandon, INRIA

Minos Garofalakis, Technical University of Crete, Greece

Buğra Gedik, Bilkent University

Rainer Gemulla, Max-Planck-Institut Saarbrücken  
Gabriel Ghinita, University of Massachusetts Boston  
Parke Godfrey, York University  
Michaela Goetz, Cornell University  
Lukasz Golab, University of Waterloo  
Sergio Greco, University of Calabria  
Le Gruenwald, University of Oklahoma  
Krishna Gummadi, MPI  
Haryadi Gunawi, University of California, Berkeley  
Rahul Gupta, IIT Bombay  
Marios Hadjieleftheriou, AT&T labs  
Kuno Harumi, HP Labs  
Michael Hay, Cornell  
Bingsheng He, NTU Singapore  
Sven Helmer, Free University of Bozen-Bolzano  
Howard Ho, IBM Almaden Research  
Katja Hose, Aalborg University  
Bill Howe, University of Washington  
Jeong-Hyon Hwang, State University of New York, Albany  
Stratos Idreos, CWI  
Hans-Arno Jacobsen, University of Toronto  
Ricardo Jimenez-Peris, Technical University of Madrid  
Ruoming Jin, Kent State University  
Ryan Johnson, University of Toronto  
Vanja Josifovski, Yahoo Inc.  
Panos Kalnis, King Abdullah University of Science and Technology  
Vana Kalogeraki, Athens Univ. of Econ. and Business  
Carl-Christian Kanne, University of Mannheim  
Hillol Kargupta, University of Maryland Baltimore County  
Yiping Ke, Institute of High Performance Computing  
Anne-Marie Kermarrec, INRIA  
Daniel Kifer, PSU  
Changkyu Kim, Intel  
George Kollios, Boston University  
Christian König, Microsoft Research  
Laks V. S. Lakshmanan, University of British Columbia

Paul Larson, Microsoft  
Mong-Li Lee, National University of Singapore  
Wang-Chien Lee, Penn State University  
Wolfgang Lehner, Technische Universität Dresden  
Chengkai Li, The University of Texas at Arlington  
Cuiping Li, Renmin University of China  
Feifei Li, University of Utah  
Guoliang Li, Tsinghua University  
Lipyeow Lim, University of Hawaii at Manoa  
Xuemin Lin, University of New South Wales  
Eric Lo, The Hong Kong Polytechnic University  
Boon Thau Loo, University of Pennsylvania  
Qiong Luo, Hong Kong University of Science and Technology  
Ashwin Machanavajjhala, Duke University  
Sanjay Madria, University of Missouri-Rolla  
Amélie Marian, Rutgers University  
Frank McSherry, Microsoft  
Sharad Mehrotra, University of California, Irvine  
Poess Meikel, Oracle  
Mohamed Mokbel, University of Minnesota  
Bongki Moon, University of Arizona  
Kyriakos Mouratidis, Singapore Management University  
Gero Muhl, University of Rostock  
Karin Murthy, IBM Research  
Suman Nath, MSR  
Wolfgang Nejdl, University of Hannover  
Sylvia Nittel, University of Maine  
Beng Chin Ooi, National University of Singapore  
Tamer Ozsu, University of Waterloo  
Esther Pacitti, University of Montpellier  
Ippokratis Pandis, IBM Almaden  
Olga Papaemmanouil, Brandeis University  
Srinivasan Parthasarathy, The Ohio State University  
Jignesh Patel, University of Wisconsin  
Peter Pietzuc, Imperial College London  
Neoklis Polyzotis, University of California, Santa Cruz  
Lucian Popa, IBM Research

Bordawekar Rajesh, IBM T.J. Watson  
Vibhor Rastogi, Yahoo  
Christopher Re, University of Wisconsin, Madison  
Matthias Renz, Ludwig-Maximilians University Munich, Germany  
Marie-Christine Rousset, IMAG  
Sourav S. Bhowmick, Nayang Technological University  
Dimitris Sacharidis, IMIS Athena, Greece  
Kenneth Salem, University of Waterloo  
Maria Sapino, University of Torino  
Monica Scannapieco, Istat  
Bernhard Seeger, Philipps-Universität Marburg  
Pierre Senellart, Télécom ParisTech  
Cyrus Shahabi, USC  
Lidan Shou, Zhejiang University  
Adam Silberstein, Trifacta  
Radu Sion, Stony Brook University  
Yannis Sismanis, IBM, USA  
Mohamed Soliman, University of Waterloo  
Julia Stoyanovich, University of Pennsylvania  
Yufei Tao, Chinese University of Hong Kong  
Sandeep Tata, IBM Research  
Nesime Tatbul, ETH Zurich

#### **Demonstration Program Committee**

Anastasia Ailamaki, EPFL  
Sihem Amer-Yahia, Qatar Computing Research Institute  
Leopoldo Bertossi, University of Carleton  
Francois Bry, University of Munich  
Chee-Yong Chan, National University of Singapore  
Kevin Chang, UIUC  
Chin-Wan Chung, Korea Advanced Institute of SaT  
Gautam Das, University of Texas, Arlington  
Aris Gkoulalas-Divanis, IBM Research Ireland  
Torsten Grust, Universität Tübingen  
Herodotos Herodotou, Microsoft Research  
Yoshiharu Ishikawa, Nagoya University  
Flip Korn, AT&T Labs

Evimaria Terzi, University of Boston  
Martin Theobald, Max Planck Institute, Germany  
Anthony Tung, National University of Singapore  
Kostas Tzoumas, Technical University of Berlin  
Sergei Vassilvitskii, Google  
Stratis D. Viglas, University of Edinburgh  
Ke Wang, Simon Fraser University  
Ingmar Weber, Yahoo!  
Raymond Chi-Wing Wong, Hong Kong University of Science and Technology  
Xiaokui Xiao, NTU  
Dong Xin, Google  
Xifeng Yan, University of Santa Barbara  
Jiong Yang, Case Western Reserve University  
Ke Yi, Hong Kong University of Science and Technology  
Man Lung Yiu, Hong Kong Polytechnic University  
Cong Yu, Google Research  
Ge Yu, Northeastern University, China  
Jeffrey Yu, Chinese University of Hong Kong  
Wenjie Zhang, UNSW Australia  
Baihua Zheng, Singapore Management University  
Aoying Zhou, East China Normal University  
Xiaofang Zhou, University of Queensland

Nick Koudas, University of Toronto  
Nikos Mamoulis, University of Hong Kong  
Giansalvatore Mecca, Università della Basilicata  
Alexandra Meliou, University of Washington  
Rachel Pottinger, University of British Columbia  
Rajeev Rastogi, Yahoo! India  
Bernhard Seeger, University of Marburg  
Ambuj Singh, University of California, Santa Barbara  
Jens Teubner, ETH Zurich  
Wei Wang, University of New South Wales  
Li Xiong, Emory University  
Jia Yuan Yu, IBM Research  
Demetris Zeinalipour, University of Cyprus  
Shuigeng Zhou, Fudan University

### **Industrial Track Committee**

Michael Brodie, Verizon  
Alejandro Buchmann, Technische Universität  
Darmstadt  
Shimin Chen, HP Labs China  
Umeshwar Dayal, HP Labs  
Shel Finkelstein, SAP  
Dieter Gawlick, Oracle  
Tasos Kementsietsidis, T.J. Watson Research Center  
Tim Kraska, Brown University  
Yue Lu, twitter  
Arnab Nandi, The Ohio State University

Felix Naumann, University of Potsdam  
Fatma Ozcan, IBM Research  
Radu Popescu-Zeletin, Fraunhofer-Institut für Offene  
Kommunikationssysteme  
Raghu Ramakrishnan, Microsoft  
Jun Rao, LinkedIn  
Len Seligman, MITRE  
Eric Simon, SAP  
Haixun Wang, Microsoft Research  
Fei Wu, Google Research  
Jackie Xiang, Foursquare

### **Reproducibility Committee**

Matias Bjørling, IT University of Copenhagen  
Wei Cao, Remnin University  
Stratos Idreos, Centrum Wiskunde & Informatica  
Ryan Johnson, University of Toronto  
Martin Kaufmann, ETH Zurich  
David Koop, University of Utah  
Lucja Kot, Cornell University  
Willis Lang, University of Wisconsin

Mian Lu, Hong Kong University of Science and  
Technology  
Dan Olteanu, University of Oxford  
Paolo Papotti, Qatar Computing Research Institute  
(QCRI)  
Ben Sowell, Cornell University  
Radu Stoica, EPFL - Ecole Polytechnique Federale de  
Lausanne  
Dimitris Tsirogiannis, Microsoft Jim Gray Systems Lab

### **PhD Workshop Chairs**

Angela Bonifati, Icar-CNR  
Sanjay Chawla, University of Sydney  
Chris Jermaine, Rice University

### **Tutorial Chairs**

Serge Abiteboul, INRIA  
Gianni Mecca, Universita della Basilicata  
Haixun Wang, Microsoft Research Asia

### **Panel Chairs**

Shivnath Babu, Duke University  
Stavros Harizopoulos, Nou Data  
Ihab Ilyas, Qatar Computing Research Institute

### **Sponsorship Chairs**

Sam Madden, Massachusetts Institute of Technology  
Vassilis Vassalos, Athens Univ. of Econ. and Business  
Paolo Meriardo, Universita Roma Tre

### **Publicity Chair**

Tasos Kementsietsidis, IBM T.J. Watson Research Center

### **Proceedings Chairs**

Peer Kröger, Ludwig-Maximilians University, Munich  
Stratis D. Viglas, University of Edinburgh

### **Web Management Chair**

Francesco Guerra, University of Modena and Reggio Emilia

### **Treasury Chair**

Marios Hadjieleftheriou, AT&T Labs Research



**Local Administration**

Ufficio Convegni and dbTrento Group, University of Trento

**Logo Design**

Sakis Palpanas

**PVLDB Information Director**

Gerald Weber, University of Auckland

**PVLDB Advisory Committee**

Philip Bernstein, Michael Böhlen, Peter Buneman, Susan Davidson, Z. Meral Ozsoyoglu, S. Sudarshan, Gerhard Weikum

## LETTER FROM THE GUEST EDITORS

This is the third issue of Proceedings of the VLDB Endowment (PVLDB) Volume 6. There are six papers in this issue of PVLDB, accepted as a result of a journal-style reviewing process. The papers will be presented at the VLDB 2013 Conference to be held in Riva Del Garda, Italy, in August 2013.

The papers in this issue cover a broad range of topics from privacy preserving peer-to-peer data integration, substring partitioning, graph search, association mining, to soliciting feedback in keyword search-based data integration, and spatial keyword query processing. We would like to thank the authors for their high-quality submissions and the refinement or revision of their papers over the past few months. Thanks also to our PC members and the PC chairs for their hard work in the review process.

We hope you will find the papers in this issue interesting and inspiring, and we hope that you will continue to support us and submit your good work to our future issues. We look forward to seeing you in Riva Del Garda.

---

Ada Waichee Fu, Chinese University of Hong Kong  
Alon Halevy, Google  
Associate Editors, PVLDB 2013