



Proceedings of the VLDB Endowment

Volume 4, No. 5 – February 2011

**Proceedings of the 37th International Conference on
Very Large Data Bases, Seattle, WA**

Editor-in-Chief:

H. V. Jagadish

Guest Editors:

José Blakeley, Joseph M. Hellerstein, Nick Koudas, Wolfgang Lehner, Sunita Sarawagi, Uwe Röhm

PVLDB – Proceedings of the VLDB Endowment

Volume 4, No. 5, February 2011.

The 37th International Conference on Very Large Data Bases, Seattle, WA.

Copyright 2011 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.

Volume 4, Number 5: VLDB 2011 Research Track Papers

Pages ii - vi and 267 - 337

ISSN 2150-8097, February 2011.

Additional copies only online at: portal.acm.org, arxiv.org, and www.vldb.org

TABLE OF CONTENTS

Front Matter

Copyright Notice	ii
Table of Contents	iii
PVLDB Review Board	iv

Letters

Letter from the Research Track Co-Chair	<i>Wolfgang Lehner</i>	vi
---	------------------------	----

Research Papers

Human-Assisted Graph Search: It's Okay to Ask Questions	267
<i>Aditya Parameswaran, Anish Das Sarma, Hector Garcia-Molina, Neoklis Polyzotis, Jennifer Widom</i>	
Guided Data Repair	279
<i>..... Mohamed Yakout, Ahmed K. Elmagarmid, Jennifer Neville, Mourad Ouzzani, Ihab F. Ilyas</i>	
Hyper-Local, Directions-Based Ranking of Places	290
<i>..... Petros Venetis, Hector Gonzalez, Christian S. Jensen, Alon Halevy</i>	
Incrementally Maintaining Classification using an RDBMS	302
<i>..... M. Levent Koc, Christopher Ré</i>	
High-Throughput Transaction Executions on Graphics Processors	314
<i>..... Bingsheng Ye, Jeffrey Xu Yu</i>	
Distributed Inference and Query Processing for RFID Tracking and Monitoring	326
<i>..... Zhao Cao, Charles Sutton, Yanlei Diao, Prashant Shenoy</i>	

PVLDB REVIEW BOARD

VLDB 2011 General PC Co-Chairs

José Blakeley, Microsoft

Joe Hellerstein, University of California – Berkeley

VLDB 2011 Research Track Co-Chairs

Nick Koudas, University of Toronto and Sysomos Inc.

Wolfgang Lehner, Dresden University of Technology

Sunita Sarawagi, IIT Bombay

Reviewer

Ashraf Aboulnaga (University of Waterloo)

Sibel Adali (Rensselaer Polytechnic Institute)

Charu Aggarwal (IBM Watson Research Center)

Divyakant Agrawal (Univ. California, Santa Barbara)

Anastasia Ailamaki (EPFL Lausanne)

Gustavo Alonso (ETH Zurich)

Shivnath Babu (Duke University)

Roberto Bayardo (Google)

Elisa Bertino (Purdue University)

Peter Boncz (CWI, Netherlands)

Angela Bonifati (Icar-CNR)

Christof Bornhoevd (SAP Palo Alto)

Mike Cafarella (University of Washington)

K. Selcuk Candan (Arizona State University)

Malu Castellanos (HP Labs)

Tiziana Catarci (University of Rome)

Chee-Yong Chan (National University of Singapore)

Kevin Chang (University of Illinois, Urbana-Champaign)

Surajit Chaudhuri (Microsoft Research)

Rada Chirkova (North Carolina State University)

Jan Chomicki (University at Buffalo)

Chin-Wan Chung (Korea Advanced Institute of SaT)

Chris Clifton (Purdue University)

Christine Collet (Grenoble Institute of Technology)

Graham Cormode (AT&T Labs)

Gautam Das (University of Texas, Arlington)

Anish Das Sarma (Yahoo! Research)

Amol Deshpande (University of Maryland)

AnHai Doan (University of Wisconsin)

Xin Dong (AT&T Labs)

Alexandre Evfimievski (IBM Research)

Wenfei Fan (University of Edinburgh & Bell Labs)

Johann-Christoph Freytag (Humboldt-Universität Berlin)

Johannes Gehrke (Cornell University)

Rainer Gemulla (IBM Almaden Research Center)

Aristides Gionis (Yahoo! Research)

Goetz Graefe (HP Labs)

Torsten Grust (Universität Tübingen, Germany)

Giovanna Guerrini (University of Genova)

Dimitris Gunopulos (University of Athens, Greece)

Theo Haerder (University of Kaiserslautern)

Alon Halevy (Google)

Vagelis Hristidis (Florida International University)

Meichun Hsu (HP Labs, Palo Alto)

Ihab Ilyas (University of Waterloo)

Zachary Ives (University of Pennsylvania)

Dean Jacobs (SAP)

Christian Jensen (Aalborg University)

Chris Jermaine (University of Florida)

Raghav Kaushik (Microsoft Research)

Bettina Kemme (McGill University)
Eamonn Keogh (University of California, Riverside)
Martin Kersten (CWI)
Christoph Koch (Cornell University)
Flip Korn (AT&T Labs)
Donald Kossmann (ETH Zurich)
Alberto Laender (Federal University of Minas Gerais)
Dongwon Lee (Penn State University)
Kristen Lefevre (University of Michigan)
Chen Li (University of California, Irvine)
Bin Liu (University of Michigan)
David Lomet (Microsoft Research)
Samuel Madden (MIT)
Nikos Mamoulis (University of Hong Kong)
Ioana Manolescu (INRIA)
Claudia Medeiros (University of Campinas)
Sergey Melnik (Google)
Marco Mesiti (Universita degli Studi di Milano)
Chaitanya Mishra (Facebook Inc.)
Felix Naumann (University of Potsdam)
Raymond Ng (University of British Columbia)
Christopher Olston (Yahoo! Research)
Themis Palpanas (University of Trento)
Dimitris Papadias (Hong Kong University of SaT)
Stavros Papadopoulos (Chinese University of Hong Kong)
Stefano Paraboschi (University of Bergamo)
Jian Pei (Simon Fraser University)
Rachel Pottinger (University of British Columbia)
Vijayshankar Raman (IBM Almaden Research Centre)
Prakash Ramanan (Wichita State University)

PVLDB Information Director

Gerald Weber (University of Auckland)

Steering Committee

Serge Abiteboul, Peter Apers, Philip Bernstein, Elisa Bertino, Peter Buneman, Martin Kersten, Z. Meral Ozsoyuglu

Louiq Raschid (University of Maryland)
Kenneth Ross (Columbia University)
Elke Rundensteiner (Worcester Polytechnic Institute)
Yehoshua Sagiv (Hebrew University, Jerusalem)
Ken Salem (University of Waterloo)
Kai-Uwe Sattler (Ilmenau University of Technology)
Bernhard Seeger (University of Marburg)
Jayavel Shanmugasundaram (Yahoo! Research)
Kyuseok Shim (Seoul National University)
Divesh Srivastava (AT&T Labs)
Dan Suciu (University of Washington)
S. Sudarshan (IIT Bombay)
Kian-Lee Tan (National University of Singapore)
Val Tannen (University of Pennsylvania)
Jens Teubner (ETH Zurich)
Martin Theobald (Max-Planck-Institut für Informatik)
Frank Tompa (University of Waterloo)
Anthony Tung (National University of Singapore)
Patrick Valduriez (INRIA)
Wie Wang (University of North Carolina)
Gerhard Weikum (Max Planck Institute, Germany)
Yuqing Wu (Indiana University)
Fei Xu (Microsoft Search)
Sihem Yahia (Yahoo! Research)
Jun Yang (Duke University)
Cong Yu (Yahoo! Research)
Jefferey Yu (Chinese University of Hong Kong)
Ting Yu (North Carolina State University)
Xiaohui Yu (York University)
Justin Zobel (University of Melbourne)

VLDB 2011 Proceedings Chair

Uwe Roehm (University of Sydney)

LETTER FROM THE RESEARCH TRACK CO-CHAIR

It is my pleasure to present the 5th issue of the PVLDB Journal Volume 4 as part of the 37th International Conference on Very Large Data Bases (VLDB 2011) to take place in Seattle (WA) from Aug 29th to Sept 3rd. This issue comprises six excellent papers accepted within the ongoing review process with monthly deadlines and extremely challenging review cycles. This set of papers covers a wide range of database research topics, from Web Search Improvement to Transaction Processing on GPUs.

In more detail, the contribution "Human-Assisted Graph Search: It's Okay to Ask Questions" by Aditya Parameswaran et al. addresses application domains such as debugging workflows, image segmentation and categorization, and interactive search and filter synthesis, and proposes a method to leverage human intelligence when searching for target nodes in directed acyclic graphs. Experiments demonstrate the superiority of their approach in the realms of webpage categorization on a real taxonomy. The second paper of this PVLDB issue, "Guided Data Repair" by Mohamed Yakout et al., presents a framework to incorporate user feedback into a data cleaning process in order to enhance and accelerate existing automatic repair techniques. Active learning techniques are applied to keep the user involvement as low as possible. The next paper, "Hyper-Local, Directions-Based Ranking of Places" by Petros Venetis et al., addresses the problem of determining the importance of points of interest, or places, in local-search results. The authors propose a framework that enables a range of aspects of directions queries to be exploited for the ranking of places, including the frequency of references to certain places in directions queries. An empirical study shows that the approach is feasible for real Web search engines.

The research paper "Incrementally Maintaining Classification using an RDBMS" by M. Levent Koc et al. discusses techniques to incremental maintenance for model-based classification views in the presence of updates. The authors present a new incremental view maintenance strategy and a hybrid index structure that they show to be more efficient than existing maintenance approaches on a variety of data by an order of magnitude. The next paper called "High-Throughput Transaction Executions on Graphics Processors", written by Bingsheng He et al., tackles the problem of running OLTP workloads on GPU environments. The paper proposes a grouping of different transactions and the concurrent execution of the bulk of transactions on the GPU. Comparisons with CPU-based setups show a 4-10 times higher throughput for the proposed architecture. Finally, the last paper of this issue, "Distributed Inference and Query Processing for RFID Tracking and Monitoring" by Zhao Cao et al., proposes an architecture to combine location and containment inference with stream query processing for a scalable and distributed RFID tracking and monitoring solution. The authors' contributions include novel inference techniques to provide accurate estimates of object locations and inter-object relationships as well as distributed inference and query processing techniques that minimize the computation state transferred across local databases.

In summary, we are proud to present this collection of top-notch research results within the PVLDB series. Thanks again to the authors and to all the reviewers for their hard work making this innovative process possible. I am looking forward to seeing you all in Seattle!

Wolfgang Lehner, Dresden University of Technology, Germany
VLDB 2011 Research Track Co-Chair