Very Large Data 2011 Bases

VLDB2011

37th International Conference on Verly Large Data Bases, Seattle, WA

Proceedings of the VLDB Endowment

Volume 4, No. 6 – March 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. 6, March 2011.
The 37th International Conference on Very Large Data Bases, Seattle, WA.

Volume 4, Number 6: VLDB 2011 Research Track Papers
Pages ii – vi and 338 – 408
ISSN 2150-8097, March 2011.

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

TABLE OF CONTENTS

Front Matter

Letters

Research Track Papers

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)

Louiqa 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)

PVLDB Information Director

Gerald Weber (University of Auckland)

VLDB 2011 Proceedings Chair

Uwe Roehm (University of Sydney)

Steering Committee

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

LETTER FROM THE RESEARCH TRACK CO-CHAIR

I am happy to present the sixth issue of PVLDB comprising of six papers accepted as part of the monthly review cycle and to be presented at the VLDB 2011 conference.

In line with the VLDB tradition, the papers in this issue cover a broad set of topics including information retrieval, text joins, index tuning, statistical inference, map reduce, and social networking. The paper "Similarity Join Size Estimation using Locality Sensitive Hashing" shows how stratified sampling on the hash index created from locality preserving indices can be used for estimating join size. The paper "Query Expansion Based on Clustered Results" is about information retrieval, specifically about choosing multiple keyword suggestions so that each covers a diverse topic. The paper "CoPhy: A Scalable, Portable, and Interactive Index Advisor for Large Workloads" proposes a novel Binary Integer Programming based formulation of the clas- sical index selection problem in relational databases.

Markov logic networks are powerful statistical modeling tools that are increasingly seeing many applications. A well-known limitation of these networks is their inference speed. The paper "Tuffy: Scaling up Statistical Inference in Markov Logic Networks using an RDBMS" addresses this important challenge by leveraging the flexibility of today's relational database engine. The paper "Automatic Optimization for MapReduce Programs" covers the hot topic of optimizing database performance in MapReduce settings. The key idea in this paper is to exploit static analysis of a database program that accesses a database and use that to optimize the Map phase through automatic index creation, compression, or early projections. Finally, we include "On Social-Temporal Group Query with Acquaintance Constraint" which is about finding the set of people to invite for a meeting so as to satisfy both social relationship constraints and specified timing constraint. The paper presents a very useful pruning strategy for an otherwise difficult problem.

I sincerely hope that you find these articles enriching. I thank the authors and the reviewers for their effort in realizing this issue.

Sunita Sarawagi, IIT Bombay
VLDB 2011 Research Track Co-Chair