Lecture Notes in Computer Science 7968

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison
Lancaster University, UK

Takeo Kanade
Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler
University of Surrey, Guildford, UK

Jon M. Kleinberg
Cornell University, Ithaca, NY, USA

Alfred Kobsa
University of California, Irvine, CA, USA

Friedemann Mattern
ETH Zurich, Switzerland

John C. Mitchell
Stanford University, CA, USA

Moni Naor
Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz
University of Bern, Switzerland

C. Pandu Rangan
Indian Institute of Technology, Madras, India

Bernhard Steffen
TU Dortmund University, Germany

Madhu Sudan
Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos
University of California, Los Angeles, CA, USA

Doug Tygar
University of California, Berkeley, CA, USA

Gerhard Weikum
Max Planck Institute for Informatics, Saarbruecken, Germany

Georg Gottlob Giovanni Grasso
Dan Olteanu Christian Schallhart (Eds.)

Big Data

29th British National Conference
on Databases, BNCOD 2013
Oxford, UK, July 8-10, 2013
Proceedings

Springer

Volume Editors

Georg Gottlob
Giovanni Grasso
Dan Olteanu
Christian Schallhart

University of Oxford
Wolfson Building, Parks Road
Oxford, OX1 3QD, UK

E-mail:
{georg.gottlob; giovanni.grasso; dan.olteanu; christian.schallhart}@cs.ox.ac.uk

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-39466-9 e-ISBN 978-3-642-39467-6
DOI 10.1007/978-3-642-39467-6
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2013942451

CR Subject Classification (1998): H.4, H.3, H.2.8, H.2, I.2.4, I.2.6

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

© Springer-Verlag Berlin Heidelberg 2013
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This volume contains the papers presented at BNCOD 2013: 29th British National Conference on Databases held during July 7–9, 2013, in Oxford.

The BNCOD Conference is a venue for the presentation and discussion of research papers on a broad range of topics related to data-centric computation. For some years, every edition of BNCOD has centered around a main theme, acting as a focal point for keynote addresses, tutorials, and research papers. The theme of BNCOD 2013 is *Big Data*. It encompases a growing need to manage data that is too big, too fast, or too hard for the existing technology.

This year, BNCOD attracted 42 complete submissions from 14 different African, European, South and North American countries. Each submission was reviewed by three Program Committee members. The committee decided to accept 20 papers on such topics as query and update processing, relational storage, benchmarking, XML query processing, Big Data, spatial data, indexing, data extraction and social networks. The conference program also included three keynote talks, two tutorials, and one panel session.

We would like to thank the authors for supporting BNCOD by submitting their work to the conference, the Program Committee members for their help in shaping an excellent conference program, and the distinguished speakers for accepting our invitation. Thanks also go to Elizabeth Walsh, Karen Barnes, and Andrea Pilot for their involvement in the local organization of the conference.

May 2013

Georg Gottlob
Giovanni Grasso
Dan Olteanu
Christian Schallhart

Organization

Program Committee

Marcelo Arenas	Pontificia Universidad Catolica de Chile, Chile
François Bry	Ludwig Maximilian University, Munich, Germany
Andrea Calì	Birkbeck College, University of London, UK
James Cheney	University of Edinburgh, UK
Thomas Eiter	TU Vienna, Austria
Tim Furche	University of Oxford, UK
Floris Geerts	University of Antwerp, Belgium
Jun Hong	Queen's University Belfast, UK
Stratos Idreos	CWI Amsterdam, The Netherlands
Mike Jackson	Birmingham City University, UK
Anne James	Coventry University, UK
Georg Lausen	Albert Ludwig University, Freiburg, Germany
Lachlan Mackinnon	University of Greenwich, UK
Sebastian Maneth	University of Oxford, UK
Peter McBrien	Imperial College, London, UK
David Nelson	University of Sunderland, UK
Werner Nutt	Free University of Bozen-Bolzano, Italy
Dan Olteanu	University of Oxford, UK
Norman Paton	University of Manchester, UK
Reinhard Pichler	TU Vienna, Austria
Mark Roantree	Dublin City University, Ireland
Florin Rusu	University of California, Merced, USA
Sandra Sampaio	University of Manchester, UK
Pierre Senellart	Telecom Paris Tech, France
Letizia Tanca	Politecnico di Milano, Italy
Andy Twigg	University of Oxford, UK
Stratis Viglas	University of Edinburgh, UK
John Wilson	University of Strathclyde, UK
Peter Wood	Birkbeck College, University of London, UK

Additional Reviewers

Antova, Lyublena	Schallhart, Christian
Fink, Robert	Schneider, Patrik
Mazuran, Mirjana	Simkus, Mantas
Pieris, Andreas	Spina, Cinzia Incoronata
Quintarelli, Elisa	Xiao, Guohui
Savenkov, Vadim	Zavodny, Jakub

Table of Contents

Keynotes

Big Data Begets Big Database Theory 1
Dan Suciu

Compilation and Synthesis in Big Data Analytics 6
Christoph Koch

The Providence of Provenance 7
Peter Buneman

Tutorials

A Tutorial on Trained Systems: A New Generation of Data Management Systems? 13
Christopher Ré

Querying Big Social Data 14
Wenfei Fan

Panel

Database Research Challenges and Opportunities of Big Graph Data . . . 29
Alexandra Poulovassilis

Query and Update Processing

Adapting to Node Failure in Sensor Network Query Processing 33
Alan B. Stokes, Alvaro A.A. Fernandes, and Norman W. Paton

Loop Elimination for Database Updates 48
Vadim Savenkov, Reinhard Pichler, and Christoph Koch

Relational Storage

Append Storage in Multi-Version Databases on Flash 62
Robert Gottstein, Ilia Petrov, and Alejandro Buchmann

Lossless Horizontal Decomposition with Domain Constraints on Interpreted Attributes 77
Ingo Feinerer, Enrico Franconi, and Paolo Guagliardo

Benchmarking

MatchBench: Benchmarking Schema Matching Algorithms for Schematic Correspondences 92
Chenjuan Guo, Cornelia Hedeler, Norman W. Paton, and Alvaro A.A. Fernandes

Towards Performance Evaluation of Semantic Databases Management Systems 107
Bery Mbaiossoum, Ladjel Bellatreche, and Stéphane Jean

XML Query Processing

On the Connections between Relational and XML Probabilistic Data Models 121
Antoine Amarilli and Pierre Senellart

On Bridging Relational and Document-Centric Data Stores 135
John Roijackers and George H.L. Fletcher

Fast Multi-update Operations on Compressed XML Data 149
Stefan Böttcher, Rita Hartel, and Thomas Jacobs

A Practical Approach to Holistic B-Twig Pattern Matching for Efficient XML Query Processing 165
Dabin Ding, Dunren Che, Fei Cao, and Wen-Chi Hou

Big Data

Representing MapReduce Optimisations in the Nested Relational Calculus 175
Marek Grabowski, Jan Hidders, and Jacek Sroka

Bisimulation Reduction of Big Graphs on MapReduce 189
Yongming Luo, Yannick de Lange, George H.L. Fletcher, Paul De Bra, Jan Hidders, and Yuqing Wu

Sampling Estimators for Parallel Online Aggregation 204
Chengjie Qin and Florin Rusu

The Business Network Data Management Platform 218
Daniel Ritter

Spatial Data and Indexing

Assessing the Completeness of Geographical Data 228
Simon Razniewski and Werner Nutt

A Comprehensive Study of iDistance Partitioning Strategies for kNN Queries and High-Dimensional Data Indexing 238
Michael A. Schuh, Tim Wylie, Juan M. Banda, and Rafal A. Angryk

Extending High-Dimensional Indexing Techniques Pyramid and iMinMax(θ): Lessons Learned 253
Karthik Ganesan Pillai, Liessman Sturlaugson, Juan M. Banda, and Rafal A. Angryk

Data Extraction and Social Networks

A Game Theory Based Approach for Community Detection in Social Networks 268
Lihua Zhou, Kevin Lü, Chao Cheng, and Hongmei Chen

A Learning Classifier-Based Approach to Aligning Data Items and Labels 282
Neil Anderson and Jun Hong

Self-supervised Automated Wrapper Generation for Weblog Data Extraction 292
George Gkotsis, Karen Stepanyan, Alexandra I. Cristea, and Mike Joy

Author Index 303