

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

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

Moshe Y. Vardi

Rice University, Houston, TX, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

Abdelkader Hameurlain Josef Küng
Roland Wagner Stephen W. Liddle
Klaus-Dieter Schewe Xiaofang Zhou (Eds.)

Transactions on Large-Scale Data- and Knowledge- Centered Systems X

Special Issue on Database-
and Expert-Systems Applications

Editors-in-Chief

Abdelkader Hameurlain
Paul Sabatier University, IRIT, Toulouse, France
E-mail: hameur@irit.fr

Josef Küng
Roland Wagner
University of Linz, FAW, Austria
E-mail: {jkueng, rrwagner}@faw.at

Guest Editors

Stephen W. Liddle
Brigham Young University, Provo, UT, USA
E-mail: liddle@byu.edu

Klaus-Dieter Schewe
Software Competence Center Hagenberg, Austria
and University of Linz, FAW, Austria
E-mail: kd.schewe@scch.at

Xiaofang Zhou
University of Queensland, Brisbane, QLD, Australia
E-mail: zxf@uq.edu.au

ISSN 0302-9743 (LNCS) e-ISSN 1611-3349 (LNCS)
ISSN 1869-1994 (TLDKS)
ISBN 978-3-642-41220-2 e-ISBN 978-3-642-41221-9
DOI 10.1007/978-3-642-41221-9
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013948623

CR Subject Classification (1998): H.2.8, H.2, I.2.6, H.3, D.2, C.2, F.2, J.1

© 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

Following the 23rd International Conference on Database and Expert Systems Applications (DEXA 2012, proceedings published in LNCS volumes 7446 and 7447) the programme chairs invited the authors of twelve selected papers to submit original, extended and revised papers to a special issue of the Transactions on Large-Scale Data- and Knowledge-Centered Systems (TLDKS). Following these invitations ten papers were submitted, each of which was carefully reviewed by two experts in the field in two reviewing rounds. Finally, seven papers were accepted for this special issue.

In their paper “Stepwise Development of Formal Models for Web Services Compositions” Idir Ait-Sadoune and Yamine Ait-Ameur deal with modelling and property verification for composed web services. In general, the composition process for web services is defined by a choreography and orchestration of atomic services. These compositions are seen as state transition systems describing the communication protocol between the participating services. The description languages for service compositions suffer from a lack of formal semantics and from ambiguities in the definition of their constructors, and the associated tools neglect formal verification and validation of the behaviour and the properties of the composed services. The article focuses on the modelling and verification of composed web services described by BPEL standard using the *Event-B* method. The static and dynamic parts of BPEL are formalised by Event-B, and formal refinement is used to structure the development with BPEL. A one-to-one link is guaranteed between BPEL elements and their Event-B formalisation. This correspondence provides assistance for developers to improve the quality of the obtained BPEL process. Properties are verified by proving theorems, and the whole approach is implemented in the BPEL2B tool.

The article “Computing Skyline Incrementally in Response to Online Preference Modification” by Tassadit Bouadi, Marie-Odile Cordier and René Quiniou is dedicated to skyline queries, which are understood as queries retrieving the most interesting objects from a database with respect to multi-dimensional preferences. The presented EC2Sky approach focuses on how to answer efficiently skyline queries in the presence of dynamic user preferences and large volumes of data. The approach exploits the fact that the skyline associated with any preference on a particular dimension can be computed without domination tests from the skyline points associated with first-order preferences on that particular dimension. This can be supported by IPO-Trees (Implicit Preference Order Trees), a data structure that supports the materialisation of skyline points associated with the most preferred values. However, the size of an IPO-tree grows exponentially with the number of dimensions. Therefore, the authors developed an incremental method for calculating the skyline points related to several dimensions associated with dynamic preferences. For this purpose a materialisation of

linear size is sufficient, which permits greater flexibility for updates of dimension preferences and improves the execution time and storage size of queries. Experiments on synthetic data highlight the relevance of EC2Sky compared with the IPO-Tree method.

The third article in this collection by Flavio Ferrarotti, Sven Hartmann, Sebastian Link, Mauricio Marin and Emir Muñoz handles foundations, applications and performance issues of ‘The finite Implication Problem for Expressive XML Keys’. The theoretical contribution is the definition of a new fragment of XML keys that keeps the right balance between expressiveness and efficiency of maintenance. More precisely, they characterise the associated implication problem axiomatically and develop a low-degree polynomial time decision algorithm. In comparison to previous work this new fragment of XML keys enhances the possibility of capturing properties of XML data that are significant for an application at hand. The practical contribution includes an efficient implementation of this decision algorithm and a thorough evaluation of its performance, demonstrating that reasoning about expressive notions of XML keys can be done efficiently in practice, and scales well. The results promote the use of XML keys on real-world XML practice, where a little more semantics makes applications a lot more effective. To exemplify this potential, the decision algorithm is used to calculate non-redundant covers for sets of XML keys. In turn, this permits a significant reduction of the the time required to validate large XML documents against keys from the proposed fragment.

In their article ‘ALACRITY: Analytics-Driven Lossless Data Compression for Rapid In-Situ Indexing, Storing, and Querying’ John Jenkins and his co-authors present ALACRITY, an approach to effectiveness of a fused data and index encoding of scientific, floating-point data in generating lightweight data structures that are amenable to common types of queries used in scientific data analysis. The authors exploit the representation of floating-point values by extracting significant bytes, using the resulting unique values to bin the remaining data along fixed-precision boundaries. To optimise query processing an inverted index is used mapping each generated bin to a list of records contained within, which optimises query processing with attribute range constraints. Overall, the storage footprint for both index and data is shown to be below numerous configurations of bitmap indexing while matching or outperforming query performance.

The article ‘A Declarative Approach to View Selection Modeling’ by Imene Mami, Zohra Bellahsene and Remi Coletta deals with the important view selection problem in database and data warehousing systems. Given a database (or a data warehouse) schema and a query workload, the view selection problem is to choose an appropriate set of views to be materialised such that the total query costs are optimised under constraints such as limited amount of resources and total view maintenance costs. The view selection problem is known to be NP-complete. The new contribution of the authors is a declarative approach that involves a constraint programming technique known for its efficiency for

the resolution of NP-complete problems. The view selection problem is modeled as a constraint satisfaction problem in an easy and declarative way, and its resolution is performed automatically by the constraint solver. The approach guarantees more flexibility and is extensible, as it can model and handle new constraints and new heuristic search strategies to reduce the solution space. The authors show that the performance outperforms genetic algorithms, which are known to provide the best trade-off between quality of solutions in terms of cost saving and execution time.

In the article “A Framework for Modeling, Computing and Presenting Time-Aware Recommendations” Kostas Stefanidis and his co-authors deal with recommendation systems. While many existing approaches recommend items of potential interest to users by completely ignoring the temporal aspects of ratings, the authors argue that time-aware recommendations need to be pushed to the foreground. Therefore, they introduce an extensive model for time-aware recommendations from two perspectives. From a fresh-based perspective, the use of different aging schemes for decreasing the effect of historical ratings and increasing the influence of fresh and novel ratings is proposed. From a context-based perspective, the focus is on the provision of different suggestions under different temporal specifications. In addition, to facilitate user browsing, an effective presentation layer for time-aware recommendations based on user preferences and summaries for the suggested items has been developed.

Finally, the last article in this collection is “Incremental Mining of Top-k Maximal Influential Paths in Network Data” by Enliang Xu, Wynne Hsu, Mong Li Lee and Dhaval Patel dealing with information diffusion, which refers to the spread of abstract ideas and concepts, technical information, and actual practices within a social system. The spread denotes flow or movement from a source to an adopter, typically via communication and influence. While a lot of research on information diffusion analysis has focused on discovering “influential users” and “who influences whom” neglecting the continuity of influence among users, the authors develop a new method for inferring top-k maximal influential paths capturing the continuity of influence. For this a generative influence propagation model is defined based on the independent cascade model and the linear threshold model, which mathematically models the spread of certain information through a network. The top-k maximal influential path inference problem is formalised giving rise to an efficient algorithm called TIP to infer the top-k maximal influential paths. TIP makes use of the properties of top-k maximal influential paths to dynamically increase the support and prune the projected databases. In order to address database evolution over time an incremental mining algorithm IncTIP has been developed in addition to maintain top-k maximal influential paths. Effectiveness and efficiency of both TIP and IncTIP are evaluated by a case study on both synthetic and real-world datasets.

We would like to thank all authors for their contributions to this special issue. We are grateful to all reviewers for their invaluable work in reviewing the papers and ensuring the high quality of this collection of articles. Finally, without the editorial assistance by Gabriela Wagner, who handled all the communication with the authors and the reviewers, this volume would not have been possible.

July 2013

Stephen W. Liddle
Klaus-Dieter Schewe
Xiaofang Zhou

Editorial Board

Reza Akbarinia	INRIA, France
Stéphane Bressan	National University of Singapore, Singapore
Francesco Buccafurri	Università Mediterranea di Reggio Calabria, Italy
Yuhan Cai	A9.com, USA
Qiming Chen	HP-Lab, USA
Tommaso Di Noia	Politecnico di Bari, Italy
Dirk Draheim	University of Innsbruck, Austria
Johann Eder	Alpen Adria University Klagenfurt, Austria
Stefan Fenz	Vienna University of Technology, Austria
Georg Gottlob	Oxford University, UK
Anastasios Gounaris	Aristotle University of Thessaloniki, Greece
Theo Härder	Technical University of Kaiserslautern, Germany
Dieter Kranzlmüller	Ludwig-Maximilians-Universität München, Germany
Philippe Lamarre	University of Nantes, France
Lenka Lhotská	Technical University of Prague, Czech Republic
Vladimir Marik	Technical University of Prague, Czech Republic
Mukesh Mohania	IBM India, India
Tetsuya Murai	Hokkaido University, Japan
Gultekin Ozsoyoglu	Case Western Reserve University, USA
Torben Bach Pedersen	Aalborg University, Denmark
Günther Pernul	University of Regensburg, Germany
Klaus-Dieter Schewe	University of Linz, Austria
Makoto Takizawa	Seikei University Tokyo, Japan
David Taniar	Monash University, Australia
A. Min Tjoa	Vienna University of Technology, Austria
Chao Wang	Oak Ridge National Laboratory, USA

Reviewers

Annalisa Appice	Università degli Studi di Bari Aldo Moro, Italy
Stéphane Bressan	National University of Singapore, Singapore
Silvana Castano	Università degli Studi di Milano, Italy
Max Chevalier	IRIT - SIG, Université de Toulouse, France
David Embley	Brigham Young University, USA
Flavio Ferrarotti	Victoria University of Wellington, New Zealand
Bernhard Freudenthaler	Software Competence Center Hagenberg, Austria

Theo Härder	Technical University of Kaiserslautern, Germany
Yu Hua	Huazhong University of Science and Technology, China
Gabriele Kern-Isberner	Technical University of Dortmund, Germany
Sang-Wook Kim	Hanyang University, South Korea
Meike Klettke	University of Rostock, Germany
Sebastian Link	The University of Auckland, New Zealand
Hui Ma	Victoria University of Wellington, New Zealand
Sofian Maabout	Université de Bordeaux, France
Atif Mashkoor	Software Competence Center Hagenberg, Austria
Jovan Pehcevski	European University, Macedonia
Elvinia Riccobene	Università degli Studi di Milano, Italy
Leonid Sokolinsky	South Ural State University, Russia
Qing Wang	The Australian National University, Australia

Table of Contents

Stepwise Development of Formal Models for Web Services Compositions: Modelling and Property Verification	1
<i>Idir Ait-Sadoune and Yamine Ait-Ameur</i>	
Computing Skyline Incrementally in Response to Online Preference Modification	34
<i>Tassadit Bouadi, Marie-Odile Cordier, and René Quiniou</i>	
The Finite Implication Problem for Expressive XML Keys: Foundations, Applications, and Performance Evaluation	60
<i>Flavio Ferrarotti, Sven Hartmann, Sebastian Link, Mauricio Marin, and Emir Muñoz</i>	
ALACRITY: Analytics-Driven Lossless Data Compression for Rapid In-Situ Indexing, Storing, and Querying	95
<i>John Jenkins, Isha Arkatkar, Sriram Lakshminarasimhan, David A. Boyuka II, Eric R. Schendel, Neil Shah, Stephane Ethier, Choong-Seock Chang, Jackie Chen, Hemanth Kolla, Scott Klasky, Robert Ross, and Nagiza F. Samatova</i>	
A Declarative Approach to View Selection Modeling	115
<i>Imene Mami, Zohra Bellahsene, and Remi Coletta</i>	
A Framework for Modeling, Computing and Presenting Time-Aware Recommendations	146
<i>Kostas Stefanidis, Eirini Ntoutsis, Mihalis Petropoulos, Kjetil Nørsvåg, and Hans-Peter Kriegel</i>	
Incremental Mining of Top-k Maximal Influential Paths in Network Data	173
<i>Enliang Xu, Wynne Hsu, Mong Li Lee, and Dhaval Patel</i>	
Author Index	201