

CONTENTS

Invited reproducibility paper

Fernando Chirigati, Rebecca Capone, Rémi Rampin, Juliana Freire and Dennis Shasha 95 A collaborative approach to computational reproducibility

Andreas Wolke, Martin Bichler, Fernando Chirigati and Victoria Steeves 98 Reproducible experiments on dynamic resource allocation in cloud data centers

Special section on The 7th International Conference on Similarity Search and Applications (SISAP 2014) Edited by Agma Juci Machado Traina and Caetano Traina Junior

Agma Juci Machado Traina and Caetano Traina Junior 1 Preface

Michael E. Houle, Xiguo Ma, Vincent Oria and Jichao Sun 2 Efficient similarity search within user-specified projective subspaces

Edgar Chávez, Verónica Ludueña, Nora Reyes and Patricia Roggero 15 Faster proximity searching with the distal SAT

Gonzalo Navarro and Nora Reyes 48 New dynamic metric indices for secondary memory

Wadha J. Al Marri, Qutaibah Malluhi, Mourad Ouzzani, Mingjie Tang and Walid G. Aref 79 The similarity-aware relational database set operators



Information Systems

DATABASES: CREATION, MANAGEMENT AND UTILIZATION

EUROPEAN EDITORS
GOTTFRIED VOSSEN,
FELIX NAUMANN
GERMANY

U.S. EDITOR
DENNIS SHASHA
U.S.A.



AN INTERNATIONAL JOURNAL

ELECTRONIC ACCESS
www.elsevier.com/locate/is

INFORMATION SYSTEMS

DATABASE: CREATION, MANAGEMENT AND UTILIZATION

European Editor

Gottfried Vossen
European Research Center for Information Systems
University of Muenster
Leonardo-Campus 3
48149 Muenster, Germany
Tel: +49-251-833-8150
e-mail: vossen@uni-muenster.de

Felix Naumann
Hasso Plattner Institute
Prof.-Dr.-Helmert-Str. 2-3
D-14482 Potsdam, Germany
e-mail: felix.naumann@hpi.de
Keywords: Data integration, data quality, data profiling, databases

Area Editors

Amer-Yahia
National de Recherche Scientifique (CNRS)
Laboratoire d'Informatique de Grenoble (LIG)
681 rue de la passerelle
Saint Martin D'Hères, 38401 France
e-mail: Sihem.Amer-Yahia@imag.fr
Social search and recommendations, explanations, serendipity, propagation and influence in social media

Carsten Binnig
Baden-Wuerttemberg Cooperative State University, Mannheim
Coblitzallee 1-9, D-68163 Mannheim, Germany
e-mail: carsten.binnig@dhbw-mannheim.de
System implementation, parallel and distributed data management, main-memory databases, visual and interactive data exploration

Rainer Böhme
Institute of Computer Science, University of Innsbruck
Technikerstrasse 21A, 6020 Innsbruck, Austria
e-mail: rainer.boehme@uibk.ac.at
Security and Privacy

Philippe Bonnet
IT University of Copenhagen
Rued Langaard Vej 7, 2300 Copenhagen, Denmark
e-mail: phbo@itu.dk
Systems issues, performance, sensors, flash databases

Felipe Carino Jr
Carino Consulting, 5 East Longden Avenue
Arcadia, CA 91006, USA
e-mail: fcarino@pacbell.net
Techniques for, and experiences with, industrial strength data warehouses or data mining, multimedia and object/relational warehouses

Marlon Dumas
http://kodu.ut.ee/~dumas/
University of Tartu
J Liivi 2, 50409 Tartu, Estonia
Tel.: +372 7375473, Fax: +372 7375468
e-mail: marlon.dumas@ut.ee

Mathias Fischer
Information Systems and IT-Security Group Westfälische
Wilhelms-Universität Münster Leonards-Campus 11
48149 Münster, Germany
Tel.: +49 251 83-38230, Fax: +49 251 83-38119
e-mail: mathias.fischer@uni-muenster.de
http://www.erc.is/p/mathias.fischer

Lukasz Golab
University of Waterloo
Waterloo, Ontario, Canada N2L3G1
e-mail: lgolab@uwaterloo.ca
Data stream management, data quality, data exploration

Ming Hua
Facebook, 1 Hacker Way
Menlo Park, CA 94025, USA
e-mail: arceehua@fb.com
Social media and social networks

Yannis Ioannidis
Department of Informatics
University of Athens, Panepistimioupolis TYP A Bldg.,
157-71 Ilisia, Athens, Greece
e-mail: yannis@di.uoa.gr
Query optimization, scientific databases, data, visualization and database user interfaces

Ioannis Katakis
http://www.katakis.eu
Department of Informatics and Telecommunications
National and Kapodistrian University of Athens
Panepistimiopolis, Ilisia, 157 84 Athens, Greece
Tel.: +30 6974331270
e-mail: katak@di.uoa.gr

Bettina Kemme
School of Computer Science, McGill University,
3480 University St. Montreal P2, Canada H3A 2A7
e-mail: kemme@cs.mcgill.ca
Transaction processing, distributed database systems and fault-tolerance

Filip Korn
Google Research, New York,
NY 10011, USA
e-mail: filip@google.com
Stream processing, data mining and spatial indexing

Laks V.S. Lakshmanan
Department of Computer Science
University of British Columbia
2366 Main Mall, Vancouver, BC, Canada V6T 1Z4
e-mail: laks@cs.ubc.ca
Social Networks and Media, Recommender Systems, Data Cleaning, Top-k Processing

Maurizio Lenzerini
Dipartimento di Ingegneria Informatica,
Automatica e Gestionale Antonio Ruberti
Università di Roma La Sapienza
Via Ariosto 25, I-00185 Roma, Italy
e-mail: lenzerini@dis.uniroma1.it
Data modelling, knowledge representation and reasoning techniques

Peri Loucopoulos
Manchester Business School
The University of Manchester
Manchester, M60 1QD, UK
e-mail: pericles.loucopoulos@manchester.ac.uk
Business applications, information systems engineering, requirements engineering, conceptual modelling, CASE technology

Amélie Marian
Rutgers University
Computer Science, 110 Frelinghuysen Rd,
Piscataway, USA
e-mail: amelian@cs.rutgers.edu
Top-k query processing, data corroboration, recommendation systems, XML and Web data management

Torben Bach Pedersen
Aalborg University, Dept. of Computer Science
Sølvgaardsvej 300,
9220 Aalborg, Denmark
e-mail: tbp@cs.aau.dk
Data warehouses, business intelligence, big data analytics

Manfred Reichert
http://www.uni-ulm.de/en/in/dbis/team/manfred-reichert.html
Ulm University, Institute of Databases and
Information Systems
89069 Ulm, Germany
Phone: +49 731 50 24135, Fax: +49 731 50 24134
e-mail: manfred.reichert@uni-ulm.de

U.S. Editor

Dennis Shasha
Department of Computer Science
Courant Institute of Mathematical Science
New York University, 251 Mercer Street
New York, NY 10012-1185, U.S.A.
Tel: +1-212-998-3086
e-mail: shasha@courant.nyu.edu

Reproducibility Editor

Fernando Chirigati
New York University
New York, USA
fchirigati@nyu.edu

Ralf Schenkel
University of Passau
Lehrstuhl für Informationsmanagement
Innstraße 43, 94032 Passau, Germany
e-mail: ralf.schenkel@uni-passau.de
Information extraction, mining and retrieval

Julia Stoyanovich
Assistant Professor of Computer Science
The College of Computing & Informatics
Drexel University
e-mail: stoyanovich@drexel.edu

Dan Suciu
Computer Science & Engineering, University of Washington,
Seattle, WA 98195-2350, USA
e-mail: suciu@cs.washington.edu
Probabilistic databases and applications of theory to novel tasks

Jens Teubner
TU Dortmund University, Department of Computer Science
Otto-Hahn-Strasse 14, 44227 Dortmund, Germany
e-mail: jens.teubner@cs.tu-dortmund.de
Database system implementation, databases on modern hardware

Martin Theobald
Institute of Databases and Information Systems (DBIS)
James-Franck-Ring 027/R523
University of Ulm, 89069 Ulm, Germany
e-mail: martin.theobald@uni-ulm.de
Information retrieval, linked data, RDF data management, probabilistic/temporal data models

Jan Van den Bussche
Universiteit Hasselt, Agoralaan building D,
3590 Diepenbeek, Belgium
e-mail: jan.vandenbussche@uhasselt.be
Database theory; logic; query languages; data mining; bioinformatics

Matthias Weidlich
http://www.matthiasweidlich.net
Humboldt-Universität zu Berlin
Unter den Linden 6,
10099 Berlin, Germany
Tel.: +49 (0) 30 2093 3143, Fax: +49 (0) 30 2093 5484
e-mail: matthias.weidlich@hu-berlin.de

Limsoon Wong
National University of Singapore
School of Computing,
13 Computing Drive,
Singapore 117417
e-mail: wongls@comp.nus.edu.sg
Database theory, data integration, data mining, bioinformatics, computational biology

Xifeng Yan
Computer Science Department,
University of California at Santa Barbara
Rm 1111, Harold Frank Hall,
University of California,
Santa Barbara,
CA 93106-5110, USA
e-mail: xyan@cs.ucsb.edu
Data mining, intelligent algorithms, scalable systems

INFORMATION SYSTEMS

DATABASE: CREATION, MANAGEMENT AND UTILIZATION

GUIDE FOR AUTHORS

For a full and complete Guide for Authors, please go to: <http://www.elsevier.com/locate/is>

Author inquiries

You can track your submitted article at http://help.elsevier.com/app/answers/detail/a_id/89/p/8045/. You can track your accepted article at <http://www.elsevier.com/trackarticle>. You are also welcome to contact Customer Support via <http://support.elsevier.com>.

Author benefits: (1) e-offprint, (2) 30% discount on all Elsevier books, and (3) Information Systems has no page charges.

Language (Usage and Editing services)

Please write your text in good English (American or British usage is accepted, but not a mixture of these). Authors who feel their English language manuscript may require editing to eliminate possible grammatical or spelling errors and to conform to correct scientific English may wish to use the English Language Editing service available from Elsevier's WebShop <http://webshop.elsevier.com/languageediting/> or visit our customer support site <http://support.elsevier.com> for more information.

Funding body agreements and policies

Elsevier has established agreements and developed policies to allow authors whose articles appear in journals published by Elsevier, to comply with potential manuscript archiving requirements as specified as conditions of their grant awards. To learn more about existing agreements and policies please visit <http://www.elsevier.com/fundingbodies>

Illustration services

Elsevier's WebShop (<http://webshop.elsevier.com/illustration-services/>) offers Illustration Services to authors preparing to submit a manuscript but concerned about the quality of the images accompanying their article. Elsevier's expert illustrators can produce scientific, technical and medical-style images, as well as a full range of charts, tables and graphs. Image 'polishing' is also available, where our illustrators take your image(s) and improve them to a professional standard. Please visit the website to find out more.

Publication information: *Information Systems* (ISSN 0306-4379) For 2016, volumes 55 to 62 (8 issues) are scheduled for publication. Subscription prices are available upon request from the Publisher or from the Elsevier Customer Service Department nearest you or from this journal's website (<http://www.elsevier.com/locate/is>). Further information is available on this journal and other Elsevier products through Elsevier's website (<http://www.elsevier.com>). Subscriptions are accepted on a prepaid basis only and are entered on a calendar year basis. Issues are sent by standard mail (surface within Europe, air delivery outside Europe). Priority rates are available upon request. Claims for missing issues should be made within six months of the date of dispatch.

Advertising information: If you are interested in advertising or other commercial opportunities please e-mail Commercialsales@elsevier.com and your inquiry will be passed to the correct person who will respond to you within 48 hours.

Orders, claims, and journal inquiries: please contact the Elsevier Customer Service Department nearest you:

St. Louis: Elsevier Customer Service Department, 3251 Riverport Lane, Maryland Heights, MO 63043, USA; phone: (877) 8397126 [toll free within the USA]; (+1) (314) 4478878 [outside the USA]; fax: (+1) (314) 4478077; e-mail: JournalCustomerService-usa@elsevier.com

Oxford: Elsevier Customer Service Department, The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK; phone: (+44) (1865) 843434; fax: (+44) (1865) 843970; e-mail: JournalsCustomerServiceEMEA@elsevier.com

Tokyo: Elsevier Customer Service Department, 4F Higashi-Azabu, 1-Chome Bldg, 1-9-15 Higashi-Azabu, Minato-ku, Tokyo 106-0044, Japan; Phone: (+81) (3) 5561 5037; fax: (+81) (3) 5561 5047; e-mail: JournalsCustomerServiceJapan@elsevier.com

Singapore: Elsevier Customer Service Department, 3 Killiney Road, #08-01 Winsland House I, Singapore 239519; phone: (+65) 6349 0222; fax: (+65) 6733 1510; e-mail: JournalsCustomerServiceAPAC@elsevier.com.

Frequency: Published 8 issues per annum in March, April, May, June, July, September, November and December

Aims and Scope

Information systems are the software and hardware systems that support data-intensive applications. The journal *Information Systems* publishes articles concerning the design and implementation of languages, data models, process models, algorithms, software and hardware for information systems. Subject areas include data management issues as presented in the principal international database conferences (e.g. ACM SIGMOD, ACM PODS, VLDB, ICDE and ICDE/EDBT) as well as data-related issues from the fields of data mining, information retrieval, internet and cloud data management, web semantics, visual and audio information systems, scientific computing, and organisational behaviour. Implementation papers having to do with massively parallel data management, fault tolerance in practice, and special purpose hardware for data-intensive systems are also welcome.

All papers should motivate the problems they address with compelling examples from real or potential applications. Systems papers must be serious about experimentation either on real systems or simulations based on traces from real systems. Papers from industrial organisations are welcome. Theoretical papers should have a clear motivation from applications. They should either break significant new ground or unify and extend existing algorithms. Such papers should clearly state which ideas have potentially wide applicability.

In addition to publishing submitted articles, the Editors-in-Chief will invite retrospective articles that describe significant projects by the principal architects of those projects. Authors of such articles should write in the first person, tracing the social as well as technical history of their projects, describing the evolution of ideas, mistakes made, and reality tests. Technical results should be explained in a uniform notation with the emphasis on clarity and on ideas that may have applications outside of the environment of that research. Particularly complex details may be summarised with references to previously published papers.

The new Reproducibility Section of Information Systems publishes articles and experiments that make it possible for readers to reproduce the experimental results presented in Information Systems articles. An article describing the reproducibility effort will be co-authored by the original authors and the reproducibility reviewer(s). Submission to the Reproducibility Section of Information Systems is by invitation only.

We will make every effort to allow authors the right to republish papers appearing in Information Systems in their own books and monographs.