

# Lecture Notes in Business Information Processing

162

## Series Editors

Wil van der Aalst

*Eindhoven Technical University, The Netherlands*

John Mylopoulos

*University of Trento, Italy*

Michael Rosemann

*Queensland University of Technology, Brisbane, Qld, Australia*

Michael J. Shaw

*University of Illinois, Urbana-Champaign, IL, USA*

Clemens Szyperski

*Microsoft Research, Redmond, WA, USA*

Philippe Cudre-Mauroux  
Paolo Ceravolo  
Dragan Gašević (Eds.)

# Data-Driven Process Discovery and Analysis

Second IFIP WG 2.6, 2.12 International Symposium  
SIMPDA 2012, Campione d'Italia, Italy, June 18-20, 2012  
Revised Selected Papers



Springer

## Volume Editors

Philippe Cudre-Mauroux  
University of Fribourg, Switzerland  
E-mail: phil@exascale.info

Paolo Ceravolo  
Università degli Studi di Milano, Italy  
E-mail: paolo.ceravolo@unimi.it

Dragan Gašević  
Athabasca University, AB, Canada  
E-mail: dgasevic@acm.org

ISSN 1865-1348

ISBN 978-3-642-40918-9

DOI 10.1007/978-3-642-40919-6

Springer Heidelberg New York Dordrecht London

e-ISSN 1865-1356

e-ISBN 978-3-642-40919-6

Library of Congress Control Number: 2013947989

© IFIP International Federation for Information Processing 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

The rapid growth of organizational and business processes managed via information systems made a big variety of data available that consequently created a high demand for making available data analysis techniques more effective and valuable. The second edition of the International Symposium on Data-Driven Process Discovery and Analysis (SIMPDA 2012) was conceived to offer a forum where researchers from different communities and industry can share their insights in this hot new field. The symposium featured a number of advanced keynotes illustrating new approaches, presentations on recent research, a competitive PhD seminar, and selected research and industrial demonstrations. The goal is to foster exchanges among academic researchers, industry, and a wider audience interested in process discovery and analysis. The event was organized jointly by the IFIP WG 2.6 and W.G 2.12.

Submissions covered theoretical issues related to process representation, discovery, and analysis or provided practical and operational experiences in process discovery and analysis. To improve the quality of the contributions the symposium fostered the discussion during the presentation, giving authors the opportunity to improve their work by extending the presented results. For this reason, authors of accepted papers and keynote speakers were invited to submit extended articles to this post-symposium volume of LNBIP. There 17 submissions and 6 papers accepted for publication.

In the first paper "Lightweight RDF Data Model for Business Processes Analysis," Marcello Leida et al., presents a lightweight data representation model to implement business process monitoring transparently to the data creation process.

The second paper by Santiago Aguirre et al., "Combination of Process Mining and Simulation Techniques for Business Process Redesign: a Methodology Approach," addresses the problem of using simulation for organizing process redesign. In particular the simulation model is constructed based on the discovery analysis and on the waiting times calculated through a statistical analysis of the event log data.

The third paper by Wil van der Aalst et al., "Improving Business Process Models Using Observed Behavior," proposes a technique for aligning reference process models to observed behaviours, introducing five quality dimensions that are balanced and used to introduce enrich the reference model. In particular this work analyses the effect of introducing the similarity dimension in addition to other dimensions previously adopted in the literature.

The fourth paper by Sjoerd van der Spoel et al., "Process Prediction in Noisy Data Sets: A Case Study in a Dutch Hospital," applies classifier algorithms for predicting the outcome and duration of a process with the objective of improving the capability of predicting the cash flow of health care organizations.

The fifth paper by Andreas Wombacher et al., "Towards Automatic Capturing of Semi-Structured Process Provenance", discusses the problem of integrating provenance information with process monitoring systems, improving the automation of the procedure. The authors propose to achieve this result by using the access logs on file used in the execution of the process and provide a demonstration of the analysis that can be performed using that approach.

The sixth paper by Jan Mendling, "Managing Structural and Textual Quality of Business Process Models," gives an overview on how empirical research informs structural and textual quality assurance of process models.

We gratefully acknowledge the strong research community that gathered around the research problems related to process data analysis and the high quality of their research work, which is hopefully reflected in the papers of this issue. We would also like to express our deep appreciation for the referees' hard work and dedication. Above all, thanks are due to the authors for submitting the best results of their work to the symposium on Data-Driven Process Discovery and Analysis.

We are very grateful to the Università degli Studi di Milano and to IFIP for their financial support, and to the University of Fribourg, and the University of Athabasca.

August 2013

Paolo Ceravolo  
Philippe Cudre-Mauroux  
Dragan Gasevic

# Organization

## Conference Co-Chairs

Paolo Ceravolo	Università degli Studi di Milano, Italy
Philippe Cudre-Mauroux	University of Fribourg, Switzerland
Dragan Gasevic	Athabasca University, Canada

## Advisory Board

Tharam Dillon	DEBII, Curtin University, Australia
Ernesto Damiani	Università degli Studi di Milano, Italy
Moataz A. Ahmed	KFUPM, Saudi Arabia
Elizabeth Chang	DEBII, Curtin University, Australia
Erich Neuhold	University of Vienna, Austria
Karl Aberer	EPFL, Switzerland

## Demonstration and Showcase Committee

Elizabeth Chang	DEBII, Curtin University, Australia
Christian Guetl	Graz University of Technology, Austria

## Publicity Chair

Matthew Smith	Leibniz University Hannover, Germany
---------------	--------------------------------------

## PhD. Award Committee

Gregorio Piccoli	Zucchetti spa, Italy
Paolo Ceravolo	Università degli Studi di Milano, Italy
Farookh Hussain	Curtin University, Australia

## Web Chair

Fulvio Frati	Università degli Studi di Milano, Italy
--------------	---

## Program Committee

Peter Spyns	Free University of Brussels, Belgium
Rafael Accorsi	University of Freiburg, Germany

VIII Organization

Irene Vanderfeesten	Eindhoven University of Technology, The Netherlands
Daniele Bonetta	Università della Svizzera Italiana, Switzerland
Etienne Rivière	Université de Neuchatel, Switzerland
Sylvain Hallé	Université du Québec à Chicoutimi, Canada
Ioana Georgiana	Free University of Brussels, Belgium
Schahram Dustdar	Vienna University of Technology, Austria
Hong-Linh Truong	Vienna University of Technology, Austria
Ebrahim Bagheri	Athabasca University, Canada
Mustafa Jarrar	Birzeit University, Palestinian Territory
Hamid Motahari	HP Labs, USA
Valentina Emilia	Balas, University of Arad, Romania
George Spanoudakis	City University London, UK
Gregorio Martinez Perez	University of Murcia, Spain
Wei-Chiang Hong	Oriental Institute of Technology, Taiwan (China)
Mohamed Mosbah	University of Bordeaux, France
Jerzy Korczak	Wroclaw University of Economics, Poland
Jan Mendling	Wirtschaftsuniversitat Wien, Austria
Maurice van Keulen	University of Twente, The Netherlands
Bernhard Bauer	University of Augsburg, Germany
Christos Kloukinas	City University London, UK
Gabriele Ruffatti	Engineering Group, Italy
Alessandra Toninelli	Engineering Group, Italy
Eduardo Fernandez-Medina	University of Castilla-La Mancha, Spain
Chi Hung	Tsinghua University, China
Nora Cuppens	Telecom Bretagne, France
Debasis Giri	Haldia Institute of Technology, India
Wil Van der Aalst	Technische Universiteit Eindhoven, The Netherlands
Antonio Mana Gomez	University of Malaga, Spain
Davide Storelli	Università del Salento, Italy
Jose M. Alcaraz Calero	Hewlett-Packard Labs, UK

# Table of Contents

A Lightweight RDF Data Model for Business Process Analysis . . . . .	1
<i>Marcello Leida, Basim Majeed, Maurizio Colombo, and Andrej Chu</i>	
Combination of Process Mining and Simulation Techniques for Business Process Redesign: A Methodological Approach . . . . .	24
<i>Santiago Aguirre, Carlos Parra, and Jorge Alvarado</i>	
Improving Business Process Models Using Observed Behavior . . . . .	44
<i>J.C.A.M. Buijs, M. La Rosa, H.A. Reijers, B.F. van Dongen, and W.M.P. van der Aalst</i>	
Process Prediction in Noisy Data Sets: A Case Study in a Dutch Hospital . . . . .	60
<i>Sjoerd van der Spoel, Maurice van Keulen, and Chintan Amrit</i>	
Towards Automatic Capturing of Semi-structured Process Provenance . . . . .	84
<i>Andreas Wombacher and Mohammad Rezwatul Huq</i>	
Managing Structural and Textual Quality of Business Process Models . . . . .	100
<i>Jan Mendling</i>	
<b>Author Index . . . . .</b>	<b>113</b>