Lecture Notes in Computer Science 12732

More information about this subseries at http://www.springer.com/series/7408

Christian Attiogbé · Sadok Ben Yahia (Eds.)

Model and Data Engineering

10th International Conference, MEDI 2021
Tallinn, Estonia, June 21–23, 2021
Proceedings

Springer

Editors
Christian Attiogbé
University of Nantes
Nantes, France

Sadok Ben Yahia
Tallinn University of Technology
Tallinn, Estonia

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-030-78427-0 ISBN 978-3-030-78428-7 (eBook)
https://doi.org/10.1007/978-3-030-78428-7

LNCS Sublibrary: SL2 – Programming and Software Engineering

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The International Conference on Model and Data Engineering (MEDI) is an international forum for the dissemination of research accomplishments on modeling and data management. Especially, MEDI provides a thriving environment for the presentation of research on data models, database theory, data processing, database systems technology, and advanced database applications. This international scientific event, ushered by researchers from Euro-Mediterranean countries, also aims to promote the creation of North-South scientific networks, projects, and faculty/student exchanges.

In recent years, MEDI has taken place in Toulouse, France (2019), Marrakesh, Morocco (2018), Barcelona, Spain (2017), Almerìa, Spain (2016), Rhodes, Greece (2015), Larnaca, Cyprus (2014), Armantea, Italy (2013), Poitiers, France (2012), and Obidos, Portugal (2011). The tenth edition of MEDI was going to take place in Tallinn, Estonia, during June 21–23, 2021. Unfortunately, the third wave of the COVID-19 pandemic broke out this year and we had to move to a virtual conference without the opportunity to carry out face-to-face interchange with the MEDI community.

The Program Committee (PC) received 47 submissions from authors in 22 countries around the world. The selection process was rigorous, where each paper received at least three reviews. The PC, after painstaking discussions, decided to accept 16 full papers (34%), and eight short papers. Accepted papers covered broad research areas on theoretical, systems, and practical aspects. Some trends evident in accepted papers included mining complex databases, concurrent systems, machine learning, swarm optimization, query processing, semantic web, graph databases, formal methods, model-driven engineering, blockchain, cyber-physical systems, IoT applications, and smart systems.

Three keynotes were presented at MEDI 2021. Prof. Torben Bach Pedersen, from Aalborg University, Denmark, gave a keynote titled "*On the Extreme-Scale Model-Based Time Series Management with ModelarDB*" in which he presented concepts, techniques, and algorithms from model-based time series management and the implementation of these in the open-source Time Series Management System (TSMS) ModelarDB. Prof. Ali Mili, from the New Jersey Institute of Technology, USA, presented an invited talk titled "*Assume()*, *Capture()*, *Verify()*, and *Establish(): Ingredients for Scalable Program Analysis*". Prof. Mili discussed the foundations and design of a programming environment that helps the programmer analyze her/his program by querying it through four orthogonal functions *Assume()*, *Capture()*, *Verify()*, and *Establish()*. The keynote by Prof. Marlon Dumas, from the University of Tartu, Estonia, was titled "*From Process Mining to Automated Process Improvement*", and was dedicated to the presentation of an overview of four key pillars of Process Mining 2.0: (*i*) predictive process monitoring; (*ii*) robotic process mining; (*iii*) causal process mining; and (*iv*) search-based process optimization.

In these troubled times, we are grateful to all the authors who submitted their work to MEDI 2021. In addition, we would like to thank the PC members and the external

reviewers who diligently reviewed all submissions and selected the best contributions. Finally, we extend our special thanks to the Local Organizing Committee members who contributed to the success of MEDI 2021, even though they could not take pride in hosting the physical conference this year.

The EasyChair conference management system was set up for MEDI 2021, and we appreciated its ease of use for all tasks in the conference organizing workflow.

May 2021

Christian Attiogbé
Sadok Ben Yahia

Organization

General Chairs

Ladjel Bellatreche	ENSMA Poitiers, France
Jüri Vain	TalTech, Estonia

Program Committee Co-chairs

Christian Attiogbé	University of Nantes, France
Sadok Ben Yahia	TalTech, Estonia

Workshop Chairs

George Chernishev	Saint Petersburg State University, Russia
Jüri Vain	TalTech, Estonia

Program Committee

El-Hassan Abdelwahed	Cadi Ayyad University, Morocco
Alberto Abello	Universitat Politòcnica de Catalunya, Spain
Yamine Aït Ameur	ENSEEIHT, France
Idir Aït Sadoune	Centrale Supélec, France
Christian Attiogbé	University of Nantes, France
Ahmed Awad	University of Tartu, Estonia
Ladjel Bellatreche	ENSMA, France
Orlando Belo	University of Minho, Portugal
Sadok Ben Yahia	Taltech, Estonia
Sidi Mohamed Benslimane	University of Sidi Bel Abbes, Algeria
Jorge Bernardino	Polytechnic Institute of Coimbra, Portugal
Drazen Brdjanin	University of Banja Luka, Bosnia and Herzegovina
Francesco Buccafurri	UNIRC, Italy
Antonio Corral	University of Almerià, Spain
Florian Daniel	Politechnico di Milano, Italy
Patricia Derler	Kontrol GmbH, Austria
Stefania Dumbrava	É.N.S.I. pour l'Industrie et l'Entreprise, France
Radwa Elshawi	University of Tartu, Estonia
Georgios Evangelidis	University of Macedonia, Greece
Alfio Ferrara	University of Milan, Italy
Flavio Ferrarotti	Software Competence Center Hagenberg, Austria
Mamoun Filali-Amine	CNRS, France
Enrico Gallinucci	University of Bologna, Italy
Matteo Golfarelli	University of Bologna, Italy

Anastasios Gounaris	Aristotle University of Thessaloniki, Greece
Marc Gyssens	Hasselt University, Belgium
Raju Halder	Indian Institute of Technology Patna, India
Irena Holubova	Charles University, Czech Republic
Luis Iribarne	University of Almerìa, Spain
Mirjana Ivanovic	University of Novi Sad, Serbia
Nadjet Kamel	University of Science and Technology Houari Boumediene, Algeria
Selma Khouri	ESI, Algeria
Adamantios Koumpis	University of Passau, Germany
Régine Laleau	Paris-Est Creteil University, France
Yves Ledru	Grenoble Alpes University, France
Sebastian Link	University of Auckland, New Zealand
Zhiming Liu	Southwest University, China
Oscar Pastor Lopez	Universitat Politècnica de Valencia, Spain
Ivan Luković	University of Novi Sad, Serbia
Hui Ma	Victoria University of Wellington, New Zealand
Yannis Manolopoulos	Aristotle University of Thessaloniki, Greece
Patrick Marcel	Université François Rabelais Tours, France
Dominique Méry	Université de Lorraine, France
Mohamed Mosbah	Bordeaux INP, France
Chokri Mraidha	CEA LIST, France
Yassine Ouhammou	ENSMA, France
Marc Pantel	ENSEEIHT, France
Jaroslav Pokorny	Charles University, Czech Republic
Giuseppe Polese	University of Salerno, Italy
S. Ramesh	General Motors, USA
Elvinia Riccobene	University of Milan, Italy
Jérôme Rocheteau	ICAM, France
Oscar Romero	Universitat Politècnica de Catalunya, Spain
Manoranjan Satpathy	Indian Institute of Technology Bhubaneswar, India
Milos Savic	University of Novi Sad, Serbia
Klaus-Dieter Schewe	Zhejiang University, China
Timos Sellis	Facebook, USA
Giovanni Simonini	Università di Modena e Reggio Emilia, Italy
Neeraj Kumar Singh	ENSEEIHT, France
Gàbor Szàrnyas	Budapest University of Technology and Economics, Hungary
Kuldar Taveter	University of Tartu, Estonia
Riccardo Tomassini	University of Tartu, Estonia
Riccardo Torlone	Roma Tre University, Italy
Goce Trajcevski	Northwestern University, USA
Javier Tuya	University of Oviedo, Spai
Michael Vassilakopoulos	University of Thessaly, Greece
Panos Vassiliadis	University of Ioannina, Greece

Organization Committee

Monika Perkman	TalTech, Estonia
Wissem Inoubli	TalTech, Estonia
Imen Ben Sassi	TalTech, Estonia
Chahinez Ounoughi	TalTech, Estonia
Siim Salin	TalTech, Estonia

Abstracts of Invited Talks

Extreme-Scale Model-Based Time Series Management with ModelarDB

Torben Bach Pedersen

Aalborg University, Denmark
tbp@cs.aau.dk

Abstract. To monitor critical industrial devices such as wind turbines, high quality sensors sampled at a high frequency are increasingly used. Current technology does not handle these extreme-scale time series well, so only simple aggregates are traditionally stored, removing outliers and fluctuations that could indicate problems.

As a remedy, we present a model-based approach for managing extreme-scale time series that approximates the time series values using mathematical functions (models) and stores only model coefficients rather than data values.

Compression is done both for individual time series and for correlated groups of time series. The keynote will present concepts, techniques, and algorithms from model-based time series management and our implementation of these in the open source Time Series Management System (TSMS) ModelarDB. Furthermore, it will present our experimental evaluation of ModelarDB on extreme-scale real-world time series, which shows that that compared to widely used Big Data formats, ModelarDB provides up to 14x faster ingestion due to high compression, 113x better compression due to its adaptability, 573x faster aggregatation by using models, and close to linear scale-out scalability.

Bio: Torben Bach Pedersen is a Professor of Computer Science at Aalborg University, Denmark. His research interests include Extreme-Scale Data Analytics, Data warehouses and Data Lakes, Predictive and Prescriptive Analytics, with a focus on technologies for "Big Multidimensional Data" - the integration and analysis of large amounts of complex and highly dynamic multidimensional data. His major application domain is Digital Energy, where he focuses on energy flexibility and analytics on extreme-scale energy time series. He is an ACM Distinguished Scientist, and a member of the Danish Academy of Technical Sciences, the SSTD Endowment, and the SSDBM Steering Committee. He has served as Area Editor for IEEE Transactions on Big Data, Information Systems and Springer EDBS, PC (Co-)Chair for DaWaK, DOLAP, SSDBM, and DASFAA, and regularly serves on the PCs of the major database conferences like SIGMOD, PVLDB, ICDE and EDBT. He received Best Paper/Demo awards from ACM e-Energy and WWW. He is co-founder of the spin-out companies FlexShape and ModelarData.

Assume(), Capture(), Verify(), Establish(): Ingredients for Scalable Program Analysis

Ali Mili

Ying College of Computing, New Jersey Institute of Technology, USA
ali.mili@njit.edu

Abstract. Despite more than half a century of research and development, the routine correctness verification of software remains an elusive goal, and software is often delivered with known failures but undiagnosed/unrepaired faults. Some of the reasons for this shortcoming include: inadequate training, lack of effective and efficient tools, lack of valid formal specifications, and the usual obstacles of scale and complexity. In this talk we discuss the foundations and design of a programming environment that helps the programmer analyze her/his program by querying it through four orthogonal functions: `Assume()`, which enables the user to make assumptions about program states or program functions; `Capture()`, which enables the user to retrieve properties of program states or program functions; `Verify()`, which enables the user to verify some properties of program states or program functions; and `Establish()`, which enables the user to alter the code to satisfy a requirement about program states or program functions. We envision that a user may start a session with vague/incomplete specifications and an incorrect program, and conclude, through successive iterations, with a valid specification and a correct program.

Bio: Ali Mili is Professor and Associate Dean at the Ying College of Computing, New Jersey Institute of Technology, and a member of the Tunisian Academy of Sciences, Letters and Arts. He holds a *Doctorat de Troisieme Cycle* from the *Institut Polytechnique de Grenoble*, a PhD from the University of Illinois in Urbana Champaign, and a *Doctorat es-Sciences d'Etat* from the University of Grenoble. His research interests lie in software engineering.

From Process Mining to Automated Process Improvement

Marlon Dumas

University of Tartu, Estonia
marlon.dumas@ut.ee

In the past decade, process mining has become a mainstream approach to analyze and improve business processes. The key idea of process mining is to extract event logs capturing the execution of business processes on top of enterprise systems, and to use these event logs to produce visual representations of the performance of the process. These representations allow analysts to understand how a given business process is actually executed, if and how its execution deviates with respect to expected or normative pathways, and what factors contribute to poor process performance or undesirable outcomes. For example, process mining techniques allow analysts to discover process models from execution data, to enhance process models with performance metrics, and to identify factors that influence long waiting times or rework in a business process. With hundreds of documented case studies and probably thousands of other success stories, process mining is now an integral part of the Business Process Management (BPM) discipline.

The first generation of process mining methods (Process Mining 1.0) focuses on descriptive analytics. It allows managers and analysts to get detailed pictures of their processes and to manually spot bottlenecks and sources of inefficiencies and defects. We are now witnessing the emergence of Process Mining 2.0: A new generation of process mining methods focused on automation and prescriptive analytics. In other words, we are moving from process mining as a way to produce insights to process mining as an instrument to recommend actions that drive continuous adaptation and improvement and improvement.

This talk will give an overview of four key pillars of Process Mining 2.0: (1) predictive process monitoring; (2) robotic process mining; (3) causal process mining; and (4) search-based process optimization.

Predictive process monitoring methods allow us to analyze ongoing executions of a process in order to predict future states and undesirable outcomes at runtime. These predictions can be used to trigger interventions in order to maximize a given reward function, for example by generating alerts or making recommendations to process workers. The talk will provide a taxonomy of the state of the art in this field, as well as open questions and possible research directions.

Robotic process mining seeks to analyze logs generated by user interactions in order to discover repetitive routines (e.g. clerical routines) that are fully deterministic and can therefore be automated via Robotic Process Automation (RPA) scripts. These scripts are executed by software bots, with minimal user supervision, thus relieving

workers from tedious and error-prone work. The talk will present initial results in the field of robotic process mining and discuss challenges and opportunities.

Causal Process Mining is an emerging sub-field of process mining that seeks to discover and quantify cause-effect relations from business process execution logs. Such cause-effect relations may help process managers to identify business process improvement opportunities. For example, given an event log of an order-to-cash process, a typical causal process mining technique allows us to discover that when a customer is from Southeast-Asia, then assigning activity A to worker X or executing activity A before activity B (rather than the other way around) increases the probability that this customer will be satisfied by 10%. The goal of causal process mining is to identify interventions that make a difference in terms of specific performance metrics. The talk will show how causal machine learning methods are being adapted and extended to attain this goal.

Finally, the talk will introduce a gestating family of methods for search-based process optimization. These techniques rely on multi-objective optimization algorithms in conjunction with data-driven process simulation, in order to discover sets of changes to one or more business processes, which maximize one or more performance measures. The talk will present a framework for search-based process optimization and will sketch approaches that could be explored to realize the vision of a recommender system for process improvement.

Acknowledgments. The research reported in this talk is supported by the European Research Council (PIX project).

Contents

Machine Learning

Data Management - Blockchains

Databases and Ontologies