GUEST EDITORIAL PREFACE

Special Issue on ADBIS'2014 Conference

Margita Kon-Popovska, Saints Cyril and Methodius University of Skopje, Skopje, Macedonia

Yannis Manolopoulos, Aristotle University of Thessaloniki, Thessaloniki, Greece

Goce Trajcevski, Northwestern University, Evanston, IL, USA

This special issue of JDM consists of selected papers from the ADBIS'2014 Conference (18th East-European Conference on Advances in Databases and Information Systems), which took place at Lake Ohrid during September of 2014. The main objective of the conference was to provide a forum, where researchers and practitioners from broad range of specializations in Databases and Information Systems can gather and interact, exchange ideas and present their accomplishments and vision. The acceptance rate was 31%, whereas the number of presented papers was 27. Invitations for extended versions and possible inclusion in the present special issue have been sent to 8 out of these 27 papers, based on the overall evaluation of the original submissions during the review process, along with the evaluation of the respective presentations by anonymous jury at the conference. Finally, 4 papers were accepted for inclusion in this issue.

The articles appearing in this issue cover a range of interesting and timely topics – from algorithms and structures for efficient data mining on GPU architecture, through parallel sky-line query processing; designing temporal operators for NoSQL databases; graph databases and efficient processing of top-k queries; to addressing modelling and design issues such as using weighted constraints to resolve entities and flexibility of schemas in relational models. These, in many aspects, render this special issue quite along the lines of the mission of JDM – to publish original research in a broad range of database design, analysis and management.

Broadly speaking, the papers in this issue can be classified in two categories: (1) works tackling aspects of efficient data mining and query processing in various settings; and (2) works related to modelling and representation issues, and their impact on various queries' processing.

The following two papers illustrate the different challenges and perspectives related to efficient queries processing.

The paper *A Parallel Algorithm for Building iCPI-trees* by Witold Andrzejewski and Pawel Boinski addresses the problem of *collocation pattern discovery* in the settings of spatial data-

bases. Specifically, mining in this context amounts to identifying types of spatial objects that are frequently located together in a spatial neighborhood. Given the vast amount of location-in-time and other contextual data generated in a streaming fashion and the large size of the corresponding datasets the paper introduces a GPU-based version of iCPI-tree generation algorithm for the collocation pattern discovery problem. In addition to the analysis of the improvements over the state of the art, the paper presents experimental evaluations demonstrating the speed-ups achieved with the proposed solution.

The paper *Top-k Differential Queries in Graph Databases* by Elena Vasilyeva, Maik Thiele, Christof Bornhövd and Wolfgang Lehner focuses on a specific context of query processing – the one arising in graph databases where the users, upon formulating the queries of interest, may face a situation in which the processing "fails" – detected by empty answers being returned. However, the "failure" of the processing is often due to the huge volume of the data (and, possibly due to the complex schemas). To overcome this, the authors advocate the concept of differential queries to bridge the gap between an unexpected result (e.g. an empty result set) and the query intention of users. Whenever a query subgraph is missing from the answer, that information may help users in understanding which vertices and edges are the reasons for queries that unexpectedly return empty answers. This, in turn, will give them an opportunity to re-formulate the query and extract a better-quality answer. To better cope with the magnitude-problem, the concept of differential queries is refined to top-k differential queries that calculate the ranking based on users' preferences and therefore significantly support the users' understanding of query database management systems.

Modelling challenges and representations enabling richer contexts of applications of database technologies are addressed by the following two papers.

The paper *Defining Temporal Operators for Column Oriented NoSQL Databases* by Yong Hu and Stefan Dessloch moves the settings to Column-oriented NoSQL databases (CoNoSQLDBs). Recalling that in such settings, it is common to store multiple data versions with timestamp information, the work identifies the subtle problem that such temporal interval representations may yield incorrect results when processing particular queries. To remedy these potential inconsistencies, the authors propose that the original CoNoSQLDB tables are transformed into two alternative representations: - explicit history representation (EHR); - tuple time-stamping representation (TTR) in which each tuple (data version) has an explicit temporal interval. In addition, an extension is proposed to temporal relational algebra to TTRO operator model with minor modifications to the model. Similarly, a novel temporal operator is proposed for HER representation. Lastly, the work introduces a collection of corresponding temporal data processing operators (*Union*, *Difference*, *Intersection*, *Project*, *Filter*, *Cartesian product* and *Theta-Join*).

The last paper in this special issue, *Scaling with Confidence: Entity Resolution under Weighted Constraints* by Zeyu Shen and Qing Wang addresses the challenge of effective specification and efficient use of constraints during the stage of ER generation. The work is based on the observation that some constraints are more effective or robust than others, and proposes weights to express the "confidences" on constraints. Specifically, the authors study the problem of entity resolution (ER) – which is, determining the records in a given database which refer to the same entities, and in the presence of weighted constraints. A framework is introduced that can interweave positive and negative constraints into the ER process, and investigate how effectively and efficiently weighted constraints can be used for generating ER clustering results.

Each of the articles has been extended from their respective conference versions and was subjected to an additional round of reviews. The authors were provided with suggestions for improvement of their work and were additionally given instructions for final improvements of the articles, in response to which the camera-ready versions appearing in this special issue were

assembled. We would like to thank the authors, the reviewers and the editorial board of JDM for all the efforts in making this special issue possible.

Margita Kon-Popovska
Yannis Manolopoulos
Goce Trajcevski
Guest Editors
JDM