GUEST EDITORIAL PREFACE

Special Issue on ADBIS 2010: Advances in Databases and Information Systems

Barbara Catania, Department of Informatics, Bioengineering, Robotics, and System Engineering, University of Genoa, Genoa, Italy

Mirjana Ivanovic, Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Novi Sad, Serbia

Yannis Manolopoulos, Department of Informatics, Aristotle University, Thessaloniki, Greece

Bernhard Thalheim, Computer Science Institute, Christian-Albrechts-University Kiel, Kiel, Germany

The 14th East-European Conference on Advances in Databases and Information Systems, in short ADBIS, took place in the town of Novi Sad, Serbia from September 20 to 24, 2010. The main objective of the ADBIS series of conferences is to provide a forum for the dissemination of research accomplishments and promote interaction and collaboration between the database and information systems research communities from Central and East European countries and the rest of the world. As such, ADBIS has meanwhile created a tradition: its 2010 edition continued the conferences held in St. Petersburg (1997), Poznan (1998), Maribor (1999), Prague (2000), and Vilnius (2001), Bratislava (2002), Dresden (2003), Budapest (2004), Tallinn (2005), Thessaloniki (2006), Varna (2007), Pori (2008) and Riga (2009). In the past, such special issues have appeared in the Information Systems journal in 2000 and 2004, with selected papers from the 1999 and 2002 ADBIS conferences, respectively. In the present special issue, 5 papers out of the 50 (36 long and 14 short), which had been presented in Novi Sad, are included. The valuable space of the present journal is greatly acknowledged.

The first paper by Philipp Rösch and Wolfgang Lehner is entitled "Optimizing Sample Design for Approximate Query Processing" and considers sampling methods since the rapid increase of data volumes makes it a crucial component of modern dbms's, whereas the problem of automatically determining the optimal sample for a given query is almost unaddressed. To tackle this problem the authors propose a sample advisor based on a novel cost

model. Primarily designed for advising samples of a few queries specified by an expert, they additionally propose two extensions of the sample advisor. The first extension enhances the applicability by utilizing recorded workload information and taking memory bounds into account. The second extension increases the effectiveness by merging samples in case of overlapping pieces of sample advice. For both extensions, they present exact and heuristic solutions. The authors analyze the cost model properties and demonstrate the effectiveness and the efficiency of the heuristic solutions with a variety of experiments.

The second paper by Yi Ou, Peiquan Jin and Theo Härder is entitled "Flash-Aware Buffer Management for Database Systems". Classical buffer replacement policies, e. g., LRU, are suboptimal for dbms's with persistent flash disks, because they *ignore* the distinguished characteristics of flash-based storage devices. The authors introduce the basic principles of buffer management for such devices and present two efficient buffer algorithms applying these principles. As confirmed experimentally, these algorithms significantly improve the performance of flash-based databases.

The third paper by Boris Novikov, Alice Pigul and Anna Yarygina is entitled "A Performance Analysis of Semantic Caching for XML Query Processing". As well known, caching is important to achieve high system performance. Semantic caching tries to benefit from the certain knowledge of data semantics. The authors expect that this information might enable reuse of semantically close data rather than exactly equal to cached data in the traditional system. However, the major obstacle for extensive application of semantic caching is the computational complexity of the query containment problem, which is undecidable, in general. The authors introduce and compare three approximate conservative query matching algorithms for semantic caching of semi-structured queries. They analyze their applicability for distributed query processing and outline few scenarios where semantic caching can be beneficial for query processing in a distributed system of heterogeneous semi-structured information resources.

The fourth paper is entitled "On Estimating the Maximum Domination Value and the Skyline Cardinality of Multi-Dimensional Data Sets" and is co-authored by Eleftherios Tiakas, Apostolos Papadopoulos and Yannis Manolopoulos. The paper deals with skyline and top-k dominating queries, which consider the dominance relationship between objects to select the most promising ones, based on user preferences. To enable query optimization, the expected number of skyline objects as well as an object's maximum domination value has to be estimated. The authors estimate the maximum domination value under the *uniformity* and attribute *independence* assumptions. For this purpose, they provide three different methodologies and study their performance and accuracy. Among the proposed estimation methods, the so-called *Estimation with Roots* method prevails and returns the most accurate results. They also introduce the *eliminating dimension*, i.e. the dimension beyond which all domination values become equal to zero, providing an efficient estimation of that dimension. Finally, they provide an accurate estimation of the skyline cardinality.

The last paper by Michail Kazimianec and Nikolaus Augsten, entitled "Clustering with Proximity Graphs: Exact and Efficient Algorithms", considers a string clustering algorithm: Graph Proximity Cleansing (GPC). According to GPC, for each potential cluster a so-called proximity graph is computed. Since this computation is expensive, the state-of-the-art GPC algorithms only approximate the proximity graph by using sampling. Further, the quality of GPC clusters has never been compared to standard clustering techniques like -means, density-based, or hierarchical clustering. The authors propose two efficient algorithms, PG-DS and PG-SM, for the exact computation of proximity graphs. They show experimentally that their solutions are faster even for very small sample sizes. They provide an experimental

evaluation of GPC and conclude that it is very efficient showing good clustering quality in comparison to the standard techniques. These results open a new perspective on string clustering in settings, where no knowledge about the input data is available.

These five papers were put through the standard review process as specified by the reviewing journal. To this end, we acknowledge the help of Giovanna Guerrini (Italy), Matteo Golfarelli (Italy), Sergei Kuznetsov (Russia), Marco Mesiti (Italy), Tadeusz Morzy (Poland), Seon-Yeong Park (Republic of Korea), Stefano Rovetta (Italy), Dragan Stojanovic (Serbia), Athena Vakali (Greece), and Panos Vassiliadis (Greece) in putting together this special issue.

Barbara Catania
Mirjana Ivanovic
Yannis Manolopoulos
Bernhard Thalheim
Guest Editors
IJKBO