Advances in Databases and Information Systems (ADBIS) 2016 & 2017

Kjetil Nørvåg[a], Mirjana Ivanovic[b], Marite Kirikova[c], Bernhard Thalheim[d]

[a]*Norwegian University of Science and Technology, Trondheim, Norway*
[b]*University of Novi Sad, Novi Sad, Serbia*
[c]*Riga Technical University, Riga, Latvia*
[d]*Christian-Albrechts-Universität, Kiel, Germany*

Welcome to this special issue dedicated to the best papers presented at the 20th and 21st European Conference on Advances in Databases and Information Systems (ADBIS), held in Prague, Czech Republic, August 28-31, 2016, and in Nicosia, Cyprus, September 24-27, 2017.

The ADBIS series of conferences aims at providing a forum for the dissemination of research accomplishments and at promoting interaction and collaboration between the database and information systems research communities from European countries and the rest of the world. The ADBIS conferences provide an international platform for the presentation of research on database theory, development of advanced DBMS technologies, and their advanced applications. As such, ADBIS has created a tradition: its 20th and 21st editions continued the series held in St. Petersburg (1997), Poznan (1998), Maribor (1999), Prague (2000), Vilnius (2001), Bratislava (2002), Dresden (2003), Budapest (2004), Tallinn (2005), Thessaloniki (2006), Varna (2007), Pori (2008), Riga (2009), Novi Sad (2010), Vienna (2011), Poznan (2012), Genoa (2013), Ohrid (2014), and Poitiers (2015).

Out of the 21 full papers accepted among 85 submitted to ADBIS 2016, and out of 26 full papers accepted among 107 submitted to ADBIS 2017, we selected 8 papers to be invited for this special issue. After a second round of reviews we finally accepted five papers.

The first paper, entitled “Sparse prefix sums: Constant-time range sum queries over sparse multidimensional data cube” by Michael Shekelyan, Anton Dignös, and Johann Gamper deals with prefix sums, a powerful technique to answer range-sum queries over multi-dimensional arrays in $\mathcal{O}(1)$ time by looking up a constant number of values in an array of size $\mathcal{O}(N)$, where N is the number of cells in the multi-dimensional array. However, the technique suffers from $\mathcal{O}(N)$ update and storage costs. Relative prefix sums address the high update costs by partitioning the array into blocks, thereby breaking the dependency between cells. In this paper, the authors present sparse prefix sums that exploit data sparsity to reduce the high storage costs of relative prefix sums. By building upon relative prefix sums, sparse prefix sums achieve the same update complexity as relative prefix sums. The authors of relative prefix sums erroneously claimed that the update complexity is $\mathcal{O}(\sqrt{N})$ for any number of dimensions. The authors of the paper in this special issue show that this claim holds only for two dimensions, whereas the correct complexity for an arbitrary number of d dimensions is $\mathcal{O}(N^{\frac{d-1}{d}})$. To reduce the storage costs, the sparse prefix sums technique exploits sparsity in the data and avoids to materialize prefix sums for empty rows and columns in the data grid; instead, look-up tables are used to preserve constant query time. Sparse prefix sums are the first approach to achieve $\mathcal{O}(1)$ query time with sub-linear storage costs for range-sum queries over sparse low-dimensional arrays. A thorough experimental evaluation shows that the approach works very well in practice. On the tested real-world data sets the storage costs are reduced by an order of magnitude with only a small overhead in query time, thus preserving microsecond-fast query answering.

The second paper, entitled "Automatic assessment of interactive OLAP exploration" by Mahfoud Djedaini, Krista Drushku, Nicolas Labroche, Patrick Marcel, Verónika Peralta, and Willeme Verdeaux considers Interactive Database Exploration (IDE), the process of exploring a database by means of a sequence of queries aiming at answering an often imprecise user information need. In this paper, the authors are interested in the following problem: how to automatically assess the quality of such an exploration. They study this problem under the following angles. First, they formulate the hypothesis that the quality of the exploration can be measured by evaluating the improvement of the skill of writing queries that contribute to the exploration. Second, they restrict to a particular use case of database exploration, namely OLAP explorations of data cubes. Third, they propose to use simple query features to model its contribution to an exploration. The first hypothesis allows to use the Knowledge Tracing, a popular model for skill acquisition, to measure the evolution of the ability to write contributive queries. The restriction to OLAP exploration allows to take advantage of well known OLAP primitives and schema. Finally, using query features allows to apply a supervised learning approach to model query contribution. The authors show on both real and artificial explorations that automatic assessment of OLAP explorations is feasible and is consistent with the users and experts viewpoints.

The third paper, entitled "Interactive and space-efficient multi-dimensional time series subsequence matching" by Danila Piatov, Sven Helmer, Anton Dignös, and Johann Gamper describes a highly efficient access method, called Delta-Top-Index, to answer top-k subsequence matching queries over a multi-dimensional time series data set. Compared to a naïve implementation, this index has a storage cost that is up to two orders of magnitude smaller, while providing answers within microseconds. Additionally, the authors apply cache optimization techniques to speed up the construction of the index. Finally, they demonstrate the efficiency and effectiveness of the technique in an experimental evaluation with real-world data.

The fourth paper, entitled "A generic and efficient framework for flash-aware spatial indexing" by Anderson C. Carniel, Ricardo R. Ciferri, and Cristina D.A. Ciferri considers spatial indexing on Solid State Drives (SSDs). Spatial indexing on flash-based SSDs has become a core aspect in spatial database applications, and has been carried out by flash-aware spatial indices. Although there are some flash-aware spatial indices proposed in the literature, they do not exploit all the benefits of SSDs, leading to loss of efficiency and durability. In this article, the authors propose eFIND, a new generic and efficient framework for flash-aware spatial indexing. eFIND takes into account the intrinsic characteristics of SSDs by employing (i) a write buffer to avoid expensive random writes, (ii) a flushing algorithm that smartly picks modifications to be flushed in batch to the SSD, (iii) a read buffer to decrease the overhead of random reads, (iv) a temporal control to avoid interleaved reads and writes, and (v) a log-structured approach to provide data durability. Performance tests showed the efficiency of eFIND. Compared to the state of the art, eFIND improved the construction of spatial indices from 43% to 77%, and the spatial query processing from 4% to 23%.

The fifth paper, entitled "Instant restore after a media failure" by Caetano Sauer, Theo Härder, and Goetz Graefe deals with media failures. Such failures usually leave database systems unavailable for several hours until recovery is complete, especially in applications with large devices and high transaction volume. Previous work introduced a technique called single-pass restore, which increases restore bandwidth and thus substantially decreases time to repair. Instant restore goes further as it permits read/write access to any data on a device undergoing restore—even data not yet restored—by restoring individual data segments on demand. Thus, the restore process is guided primarily by the needs of applications, and the observed mean time to repair is effectively reduced from several hours to a few seconds. This paper presents an implementation and evaluation of instant restore. The technique is incrementally implemented on a system starting with the traditional ARIES design for logging and recovery. Experiments show that the transaction latency perceived after a media failure can be cut down to less than a second. The net effect is that a few "nines" of availability are added to the system using simple and low-overhead software techniques.

Guest Editors

Kjetil Nørvåg
Norwegian University of Science and Technology, Trondheim, Norway

Mirjana Ivanovic
University of Novi Sad, Novi Sad, Serbia

Marite Kirikova
Riga Technical University, Riga, Latvia

Bernhard Thalheim
Christian-Albrechts-Universität, Kiel, Germany