



Advances in Databases and Information Systems

Ladjel Bellatreche, Patrick Valduriez, Tadeuz Morzy

► **To cite this version:**

Ladjel Bellatreche, Patrick Valduriez, Tadeuz Morzy. Advances in Databases and Information Systems. Information Systems, 70, 2017, 10.1016/j.is.2017.08.003 . lirmm-01630719

HAL Id: lirmm-01630719

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-01630719>

Submitted on 8 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Advances in Databases and Information Systems

Ladjel Bellatreche, LIAS/ISAE-ENSMA – Poitiers University, France. E-mail: bellatreche@ensma.fr

Patrick Valduriez, INRIA and LIRMM, Montpellier, France. E-mail: Patrick.Valduriez@inria.fr

Tadeuz Morzy, Institute of Computing Science, Poznan University of Technology, Poznan, Poland.
E-mail: tadeusz.morzy@put.poznan.pl

Welcome to this special issue dedicated to the best papers presented at the 19th East-European Conference on Advances in Databases and Information Systems (ADBIS) and its satellite events that was held in Poitiers, France, September 8 - 11, 2015 (<https://adbis2015.ensma.fr/>).

The ADBIS conference aims at providing a forum where researchers and practitioners in the fields of databases and information systems can interact, exchange ideas and disseminate their accomplishments and visions. Inaugurated 19 years ago, ADBIS originally included communities from Central and Eastern Europe, however, throughout its lifetime it has spread and grown to include participants from many other countries throughout the world. This edition is special, as it is the first time that ADBIS takes place in France.

Out of the 31 full papers accepted among 135 submitted to ADBIS 2015, we selected only 5 papers to be invited for this special issue. The special issue also received 18 papers from the open call for papers. After a second round of reviews we finally accepted four papers (three from ADBIS conference and one from the open call), which represents a selective acceptance rate (17%). These 4 papers represent innovative and high-quality research. The topics of these accepted papers are very timely and include: graph sequences, evolving graphs, ETL (Extract, Transform, Load), NoSQL implementations, data integration, distributed and parallel processing, clustering, privacy, k-segment problem.

The first paper, entitled “A Scalable Dynamic Programming Scheme for the Computation of Optimal k-Segments for Ordered Data” by Giovanni Mählknecht, Anton Dignös, and Johann Gamper deals with the optimality of k-segments of an ordered dataset of size n . More concretely, its goal is to find k tuples that are obtained by merging consecutive tuples such that a given error metric is minimized. Several instances of this problem exist in the literature, such as piecewise-constant approximation (PCA), parsimonious temporal aggregation (PTA), and the construction of V -optimal histograms (VOH). The authors give a nice formalization of the problem of optimal k-segments and describe various optimization techniques that improve previous results on the subject. All the proposed techniques are well presented and are shown to be effective through detailed experiments.

The second paper, entitled “Efficient Processing of Shortest Path Queries in Evolving Graph Sequences”, by Chenghui Ren, Eric Lo, Ben Kao, Xinjie Zhu, Reynold Cheng, David W. Cheung, considers a timely topic related to the usage of graphs to represent data. It addresses the problem of answering shortest path queries on time evolving graph sequences. A rich state of art on

efficient algorithms and data structures for evaluating distance-based queries and reachability queries on very large graphs is given. The authors provide the find-verify-fix (FVF) framework and discuss two solutions, the FVF-F based on flat clustering, and FVF-H based on hierarchical clustering. The paper presents an extension of a previous VLDB'11 paper by the authors. To validate the efficiency of their proposal, the authors evaluate the FVF framework on both real and synthetic databases for Evolving Graph Sequences (EGS). Three real EGS datasets are used: Internet, Flickr, and FBFriend, which are available at: <http://socialnetworks.mpi-sws.org/datasets.html>.

The third paper, entitled “Inference Control of Open Relational Queries under Closed-World Semantics based on Theorem Proving”, by Joachim Biskup, Martin Bring and Michael Bulinski deals with the problem of privacy preservation in relational databases using inference control procedures performed by theorem provers. The standard theorem provers are not efficient to deal with large databases, because of the large number of prover calls (among others). The authors propose some strategies to improve the performance of inference control, particularly by: 1) enhancing theorem proving by transformation of completeness sentences capturing the closed world assumption occurring as an input of the prover; 2) reducing the number of pertinent prover calls by a divide-and-conquer approach; 3) dedicated direct calculation of positions of candidate tuples which can be excluded from probing during the search for a most informative query answer that does not violate the confidentiality requirements. Worst-case complexity, practical feasibility and scalability issues are described.

The fourth paper, entitled “Distributed Snapshot Maintenance in Wide-Column NoSQL Databases Using Partitioned Incremental ETL Pipelines”, by Weiping Qu and Stefan Desseloch proposes the HBelt system which integrates the HBase data store with a distributed, pipelined data integration engine (based on Kettle ETL engine) to deliver distributed snapshot maintenance for real-time analytics. The paper addresses the issue of data freshness of the data buffered in HBase using incremental ETL flows. Two data parallelism techniques are used (data partitioning and data pipelining). A consistency model is defined for incoming queries for distributed snapshot maintenance. Intensive experiments are conducted to evaluate the performance of HBelt deployed on a 6-node cluster with a dataset of TPC-DS benchmark and to compare its efficiency against the sequential execution mode. The obtained results show that HBelt contributes in reducing the maintenance overhead and increasing the request throughput for real-time analytics in HBase.