VLDB SI 2018 editorial

Sihem Amer-Yahia[1] · Jian Pei[2]

The 2018 VLDB Conference was held in Rio de Janeiro, Brazil, during August 27–31, 2018. We received 741 research submissions, of which 18.35% were accepted for presentation by the Program Committee. From these high-quality manuscripts, a Best Papers Committee consisting of Luc Bouganim, Johannes Gehrke, Ioana Manolescu, Renee Miller, Mohamed Mokbel, and Srinivasan Parthasarathy selected six outstanding papers and also picked the best paper award for the conference. We invited the authors of the six selected papers to submit an extended version for consideration in this special issue, and they all accepted our call. The reviewers for the manuscripts were a mix of those who had reviewed the conference versions, as well as additional experts who reviewed only the journal submissions. After two rounds of reviewing, all six papers were accepted for publication in this issue, covering a diverse spectrum of topics ranging from core database engines to knowledge management.

The paper "The Ubiquity of Large Graphs and Surprising Challenges of Graph Processing" by S. Sahu, A. Mhedhbi, S. Salihoglu, J. Lin, and T. Özsu obtained the Best Paper Award at the conference. The authors surveyed 89 users and reviewed user emails and code repositories of 22 software products. The participants' responses provided useful insights into the types of graphs users have, the software and computations users use, and the major challenges users face when processing their graphs.

The paper "General Dynamic Yannakakis: Conjunctive Queries with Theta Joins Under Updates" by M. Idris, M. Ugarte, S. Vansummeren, H. Voigt, and W. Lehner presents a new approach for dynamically evaluating queries with multi-way θ-joins under updates that is effective in avoiding both materialization and recomputation of results, while supporting a wide range of applications. To do that, the authors generalized Dynamic Yannakakis, an algorithm for dynamically processing acyclic equijoin queries. They also generalized the notions of acyclicity and free-connexity to arbitrary θ-joins and show how to compute corresponding join trees.

The paper "An Analytical Study of Large SPARQL Query Logs" by A. Bonifati, W. Martens, and T. Timm conducted an extensive analytical study on a large corpus of real SPARQL query logs. Their corpus is inherently heterogeneous and consists of a majority of DBpedia query logs along with query logs on biological datasets (namely BioPortal and BioMed datasets) and geological datasets (LGD), query logs on bibliographic data (SWDF), and query logs from a museum's SPARQL endpoint (British Museum). The authors completed this corpus with the example queries from Wikidata (Feb. 2017), which are cherry picked from real SPARQL queries on this data source.

The paper "Scalable Algorithms for Signal Reconstruction by Leveraging Similarity Joins" by A. Asudeh, A. Nazi, J. Augustine, S. Thirumuruganathan, N. Zhang, G. Das, and D. Srivastava investigated how a wide ranging problem of large-scale signal reconstruction can benefit from techniques developed by the database community. Efficiently solving SRP has a number of applications in diverse domains including network traffic engineering, astronomy, medical imaging, etc. The authors proposed an algorithm based on the Lagrangian dual form of SRP. They identified a number of computational bottlenecks and evaluated the use of database techniques such as sampling and similarity joins for speeding them up without much loss in accuracy.

The paper "Snorkel: Rapid Training Data Creation with Weak Supervision" by A. Ratner, S. H. Bach, H. Ehrenberg, J. Fries. S. Wu, C. Ré develops Snorkel, a new paradigm for soliciting and managing weak supervision to create training datasets. Users provide higher-level supervision in the form of labeling functions that capture domain knowledge and resources, without having to carefully manage the noise and conflicts inherent in combining weak supervision sources. The evaluations demonstrate that Snorkel significantly

✉ Sihem Amer-Yahia
  sihem.amer-yahia@univ-grenoble-alpes.fr

  Jian Pei
  jpei@cs.sfu.ca

[1] CNRS, University of Grenoble Alpes, Grenoble, France

[2] Simon Fraser University, Burnaby, Canada

reduces the cost and difficulty of training powerful machine learning models while exceeding prior weak supervision methods and approaching the quality of large, hand-labeled training sets.

The paper "Morton Filters: Fast, Compressed Sparse Cuckoo Filters" by A. D. Breslow and N. S. Jayasena presented a high-throughput filter that supports improved throughput for lookups, insertions, and deletions without increasing memory usage. Most notably, an MF's insertion throughput is about $3\times$ to $15\times$ higher than a comparable CF for load factors at or above 0.75. Further, the authors showed that lookup and deletion throughput are up to $2.5\times$ and $1.6\times$ faster, respectively.

In closing, we thank all the authors for their diligent efforts in substantively extending their conference manuscripts and for submitting them in a timely manner. We also thank our team of reviewers for their insightful comments and suggestions, which enhanced the final quality of the papers. Finally, we reiterate our profound gratitude to the Best Papers Committee of VLDB 2018 for tackling the formidable task of selecting these six papers from a large and competitive group of conference publications.

PVLDB Volume 11 Editors-in-Chief

VLDB 2018 Program Committee Chairs