**GUEST EDITORIAL**

# VLDB SI survey editorial

Lei Chen[1] · Sihem Amer-Yahia[2]

VLDB Journal (VLDB J.) organizes a special issue on surveys of essential and emerging topics in database management. A good survey paper can provide a concise but broad overview of a topic, give a comprehensive review of existing work and compare existing solutions from different angles. Most importantly, a survey paper can provide insights regarding interesting and important new research directions. Therefore, for this special issue, VLDB J. is soliciting excellent survey papers on essential and emerging topics in database management. We have received 28 submissions, out of which 15 survey papers are finally accepted after one- or two-round reviews. The accepted survey papers cover a wide spectrum of topics including trajectory distance measure, in-memory database acceleration, decomposition of networks, data visualization, interactive data systems, regret minimization query and variants, microblogs data management, spatial crowdsourcing, data set search, indexing on flash storage, complex event recognition, graph community search, LSM-based storage, graph edit distance computation and even modeling and mining.

The paper "A survey of trajectory distance measures and performance evaluation" by H. Su, S. Liu, B. Zheng, X. Zhou, K. Zheng, H. Su, S. Liu, B. Zheng, X. Zhou and K. Zheng conducted a comprehensive survey of trajectory distance measures, which is commonly used to determine the similarity of two trajectories. They categorized the existing distance measures and studied their effectiveness in six different trajectory transformations.

The paper "In-memory database acceleration on FPGAs: a survey" by J. Fang, Y. T. B. Mulder, J. Hidders, J. Lee and H. P. Hofstee surveyed using FPGAs to accelerate in-memory database systems. They study the memory-related technology trends and conclude that FPGAs can be considered for integration in database systems. Thus, they summarized the state-of-the-art research on FPGA-accelerated database operators and discussed the major challenges, potential solutions and future directions.

The paper "The core decomposition of networks: theory, algorithms, and applications" by F. D. Malliaros, C. Giatsidis, A. N. Papadopoulos and M. Vazirgiannis surveyed the core decomposition of networks, which measures the connectivity of a node by considering its neighbors. They introduce the theories and algorithms under various settings for core decomposition, as well as a spectrum of the benefited applications.

The paper "Making data visualization more efficient and effective: a survey" by X. Qin, Y. Luo, N. Tang and G. Li surveyed the techniques that enhance the efficiency and effectiveness of data visualization. They categorized the techniques in terms of the optimization directions, i.e., visualization specifications, efficient approaches for data visualization, and data visualization recommendation. The literature of these three directions has been carefully investigated, respectively.

The paper "Evaluating Interactive Data Systems: Survey and Case Studies" by P. Rahman, L. Jiang and A. Nandi studied the important metrics and factors for evaluating interactive data systems. They list the popular metrics and propose complementary ones to fully capture the user behaviors. Three case studies have been carried out to demonstrate the system enhancements gained from analyzing the user behaviors.

The paper "An Experimental Survey of Regret Minimization Query and Variants: Bridging the Best Worlds Between Top-K Query and Skyline Query," by Min Xie, Raymond Chi-Wing Wong and Ashwin Lall, studied different variants of regret minimization query that has garnered considerable attention in recent years and presented some interesting problems that have not yet been addressed in the literature. Specifically, the authors implement 12 state-of-the-art methods published in top-tier venues for obtaining regret minimization sets and give an experimental comparison under various parameter settings on both synthetic and

✉ Lei Chen
leichen@cse.ust.hk

Sihem Amer-Yahia
Sihem.Amer-Yahia@univ-grenoble-alpes.fr

[1] Hong Kong University of Science and Technology, Sai Kung, Hong Kong

[2] CNRS, Univ. Grenoble Alpes, Grenoble, France

real datasets. The evaluation shows that the optimal choice of methods for regret minimization depends on the application demands.

The paper "Microblogs Data Management: A Survey," by Amr Magdy, Laila Abdelhafeez, Yunfan Kang, Eric Ong and Mohamed F. Mokbel, gives a comprehensive review of major research and system work in microblogs data management. The paper categorized the literature into three parts: data indexing and querying, data management systems, and data analysis. The authors also classify the studies based on the major component of the framework of each part, to provide better understanding for different techniques and highlight existing challenges and future opportunities in this research direction.

The paper "Spatial Crowdsourcing: A Survey," by Yongxin Tong, Zimu Zhou, Yuxiang Zeng, Lei Chen and Cyrus Shahabi, dives deep into the challenges and techniques brought by the unique characteristics of spatial crowdsourcing. This paper reviews the existing research on four core algorithmic issues in spatial crowdsourcing: (1) task assignment, (2) quality control, (3) incentive mechanism design and (4) privacy protection. The authors also investigate representative spatial crowdsourcing applications and discuss open questions for future spatial crowdsourcing research.

The paper "Dataset Search: A Survey," by Adriane Chapman, Elena Simperl, Laura Koesten, George Konstantinidis, Luis-Daniel Ibáñez, Emilia Kacprzak and Paul Groth, highlights that dataset search can be achieved through two different mechanisms: (1) issue query, return dataset; (2) issue query, build dataset. The authors investigate the approaches and implementations from related areas dataset search, including information retrieval, databases, entity-centric and tabular search, in order to identify possible paths to tackle these questions as well as immediate next steps that will take the field forward.

The paper "Indexing in Flash Storage Devices: A Survey on Challenges, Current Approaches, and Future Trends," by Athanasios Fevgas, Leonidas Akritidis, Panayiotis Bozanis and Yannis Manolopoulos, reviews a broad range of indexes on raw and block flash devices, describing various employed techniques, such as buffering of updates or logging, either in main or in the flash memory, utilization of fat or overflow nodes. The authors broadly survey 62 flash-aware indexes for various data types, analyze the main techniques they employ, and comment on their main advantages and disadvantages.

The paper "Complex event recognition in the Big Data era: a survey" by N. Giatrakos, E. Alevizos, A. Artikis, A. Deligiannakis and M. Garofalakis conducted a comprehensive survey on Complex Event Recognition (CER) from three major points including the formal methods for CER developed in current systems, parallel CER systems based on Big Data platforms and CER applications. The authors also pointed out promising future research possibilities of CER.

The paper "A survey of community search over big graphs" by Y. Fang, X. Huang, L. Qin, Y. Zhang, W. Zhang, R. Cheng and X. Lin thoroughly reviewed existing community search (CS) studies. They categorized previous works based on different community quality model. A comparative analysis on effectiveness and efficiency is performed to give users a comprehensive understanding on various CS algorithms.

The paper "LSM-based storage techniques: a survey" by C. Luo and M. Carey presented a survey on the state-of-the-art log-structured merge-tree (LSM-tree) techniques which are widely used in NoSQL systems. The authors summarized the major design issues of LSM-tree into six categories to provide a taxonomy on existing LSM-trees. Strengths, insufficiencies and trade-offs of different designs are analyzed compared. They also summarized representative LSM-based systems.

The paper "Comparing heuristics for graph edit distance computation" by D. Blumenthal, N. Boria, J. Gamper, S. Bougleux and L. Brun summarized the current heuristics of the calculation of graph edit distance (GED), which is NP-hard, and conducted an experimental study on previous works. The authors pointed out that though the hardness is intrinsically high, the state-of-the-art heuristics achieve small gap between the lower and upper bounds. The authors suggested the future research should focus on further speeding up existing heuristics with tight enough bounds.

The paper "Event modeling and mining: a long journey toward explainable events" by X. Chen and Q. Li proposed a detailed survey of event modeling and mining techniques. The authors introduce four event forms to distinguish different event types, and then, they compare and analyze different event mining techniques used for each event form. Furthermore, unsolved challenges and promising applications are summarized as a guidance to future research.

We would like to thank all the authors for their insightful contributions. We hope you enjoy reading these survey papers.

VLDB J. Special Issue on Survey Editors