Guest Editors' Introduction to the Special Section on the 26th International Conference on Data Engineering

Shahram Ghandeharizadeh, Jayant R. Haritsa, *Senior Member*, *IEEE*, and Gerhard Weikum

◆

1 INTRODUCTION

THE 26th International Conference on Data Engineering, ICDE 2010, was held in Long Beach, California, during 1-6 March 2010. A program committee of 230 members evaluated the 523 research manuscripts submitted to the research track of ICDE, producing an outstanding technical program consisting of 69 full and 41 short research papers. These papers covered diverse topics ranging from data clouds to social networks and location-based services. The technical program also included industrial sessions, panels, demos, and tutorials. There were three thought-provoking keynote addresses, by Richard Winter and Pekka Kostamaa on Large Scale Data Warehousing: Trends and Observations, Jeffery Naughton on Lessons from the First 50 Years, Speculations for the Next 40, and Donald Kossmann on How New Is the Cloud?

With the active encouragement and cooperation of Professor Beng Chin Ooi (Editor-in-Chief of the *IEEE Transactions on Knowledge and Data Engineering*) and the steering committee of ICDE, we have brought together the best of ICDE 2010 technical contributions in this special section of *TKDE*. Leveraging the inputs of the conference best-paper award committee, we identified seven contributions as being outstanding in their technical strength and presentation quality, and solicited extended versions from their authors to produce manuscripts with materially enhanced technical value. These extended submissions underwent a second round of reviews to ensure compliance with *TKDE* publication standards.

This special section begins with "Efficient Top-k Approximate Subtree Matching in Small Memory" by Nikolaus Augsten, Denilson Barbosa, Michael Böhlen, and Themis Palpanas, which received the Best Paper award at the conference. This study presents an elegant and comprehensive solution to the classical problem of identifying the subtrees in a data tree with the smallest edit distances from a given query tree. A detailed performance evaluation demonstrates that the proposed solution scales, both theoretically and empirically, to large XML repositories.

The second paper is "Usher: Improving Data Quality with Dynamic Forms" by Kuang Chen, Harr Chen, Neil Conway, Joseph Hellerstein, and Tapan S. Parikh, which received the Best Student Paper award at the conference. It presents principled and machine-learning-inspired techniques to address the crucial but largely unexplored problem of assuring data quality right at its very root, when humans enter data via forms. An evaluation on real-world data sets indicates that data quality can be improved considerably, and relatively inexpensively, using these techniques.

The next paper, "Efficient and Accurate Discovery of Patterns in Sequence Data Sets" by Avrilia Floratou, Sandeep Tata, and Jignesh M. Patel, investigates efficient mining of approximate contiguous patterns for applications such as computational genomics. A new suffix-tree-based algorithm called FLAME is presented and shown to be complete in its answer set, fast and scalable in performance, and adaptable to different applications.

The fourth paper is "Frequent Item Computations on a Chip" by Jens Teubner, René Müller, and Gustavo Alonso. This study investigates a fundamental redesign of CPU-based algorithms to compute frequent item sets using field-programmable gate arrays (FPGA). It shows these designs to be beneficial by enhancing performance while reducing energy consumption. Moreover, it analyzes different FPGA features to quantify how a feature trades performance for scalability.

Location-based services are addressed in "Continuous Monitoring of Distance-Based Queries" by Muhammad Aamir Cheema, Ljiljana Brankovic, Xuemin Lin, Wenjie Zhang, and Wei Wang. This study focuses on distance-based range queries issued by devices that continuously change their location in a euclidean space. The concept of a "safe zone" is introduced for efficient processing of queries, and techniques to compute it efficiently are detailed. This approach is empirically found to be close to optimal and significantly faster than straightforward solution techniques.

The sixth paper is "Differential Privacy via Wavelet Transforms" by Xiaokui Xiao, Guozhang Wang, and Johannes Gehrke. An epsilon-differential privacy-preserving data publishing technique that provides accurate answers for count queries with range predicates is presented. The primary insight is to apply wavelet transforms on the data prior to adding noise. This study shows the

- *S. Ghandeharizadeh is with the Computer Science Department, University of Southern California, 941 W. 37th Place, USC College Park, Los Angeles, CA 90089-0781. E-mail: shahram@usc.edu.*
- *J.R. Haritsa is with SERC, Indian Institute of Science, Sir C V Raman Road, Bangalore 560012, India.*
  E-mail: haritsa@dsl.serc.iisc.ernet.in.
- *G. Weikum is with Max Planck Institute for Informatics, Campus E1.4, D-66123 Saarbruecken, Germany. E-mail: weikum@mpi-inf.mpg.de.*

For information on obtaining reprints of this article, please send e-mail to: tkde@computer.org.

effectiveness and efficiency of its proposed solution using both real and synthetic data.

Finally, a formal definition of Reverse Top-k queries is provided in "Monochromatic and Bichromatic Reverse Top-k Queries" by Akrivi Vlachou, Christos Doulkeridis, Yannis Kotidis, and Kjetil Nørvåg. Using this definition, queries are classified into two categories and customized query processing techniques are designed for each category. Experimental results are presented to demonstrate that these techniques reduce the required number of computations by orders of magnitude.

This special section was put together under tight deadlines, and we wish to express our heartfelt thanks to the authors and reviewers for their cooperation and responsiveness in this effort. Our deep appreciation also goes to the program committee, organizing committee, and participants of ICDE 2010 for making the conference eminently enjoyable and technically outstanding.

Shahram Ghandeharizadeh
Jayant R. Haritsa
Gerhard Weikum
Guest Editors

Jayant R. Haritsa received the BTech degree in electronics and communications engineering from the Indian Institute of Technology, Madras, and the MS and PhD degrees in computer science from the University of Wisconsin-Madison. He is a senior professor of database systems in the Supercomputer Education and Research Centre and the Department of Computer Science and Automation at the Indian Institute of Science, Bangalore. He is a Fellow of the Indian National Academy of Engineering (INAE), the National Academy of Sciences, India (NASI), and the Indian Academy of Sciences (IAS), and is also a distinguished scientist of the ACM and a senior member of the IEEE. He has been on the editorial boards of the *IEEE Transactions on Knowledge and Data Engineering*, *VLDB Journal*, *JDPD*, and *JRTS*. He has served as the program cochair for ICDE 2010 and DASFAA 2008, the program vice-chair for ICDM 2007 and ICDE 2005, the tutorials cochair for WWW 2010 and DASFAA 2005, the demo cochair for VLDB 2009, and the organizational cochair for ICDE 2003. He is a recipient of the Shanti Swarup Bhatnagar Prize, the Swarnajayanti Fellowship, and the Vikram Sarabhai Research Award. He recently received the Best Demonstration award at VLDB 2010 for the Picasso database query optimizer visualizer.

Shahram Ghandeharizadeh received the PhD degree in computer science from the University of Wisconsin-Madison, in 1990. Since then, he has been on the faculty at the University of Southern California. His research interests include design and implementation of novel architectures for high-performance data intensive applications, multimedia-based social networking systems, parallel database systems, and active databases. In 1992, he received the US National Science Foundation Young Investigator's award for his research on the physical design of parallel database systems. He was a recipient of the ACM Software System Award in 2008 for his contributions to the Gamma Parallel Database System. He has served on the organizing committees of numerous conferences and has served on the board of several professional organizations and academic institutions.

Gerhard Weikum received the diploma and doctoral degrees from the University of Darmstadt, Germany. He is a scientific director at the Max-Planck Institute for Informatics (MPII) in Saarbruecken, Germany, where he leads the department on databases and information systems. Earlier, he held positions at Saarland University in Saarbruecken, Germany, at ETH Zurich, Switzerland, at MCC in Austin, Texas, and he was a visiting senior researcher at Microsoft Research in Redmond, Washington. His research interests are in distributed data management, performance analysis and optimization, integrating database-systems and information-retrieval methods, and the automatic construction of large-scale knowledge bases by Web and text mining. He has coauthored a comprehensive textbook on transactional information systems. His work on automatic database tuning received the 2002 VLDB 10-Year Award. He is an ACM fellow, a fellow of the German Computer Society, and a member of the German Academy of Science and Engineering. He has served on various editorial boards, including the *Communications of the ACM*, and as program committee chair of conferences like ACM SIGMOD, IEEE Data Engineering, and CIDR. From 2003 through 2009, he was president of the VLDB Endowment.