

Special Section on the International Conference on Data Engineering 2015

Wolfgang Lehner, Johannes Gehrke, and Kyuseok Shim

THIS special section covers extended versions of top-ranked papers of the 31st International Conference on Data Engineering that was held in Seoul, Korea, on April 13–17, 2015. The ICDE 2015 conference received 645 submissions within the research track. With an acceptance rate of 13 percent—which reflects the goal of ICDE to keep the highest quality standards—84 research papers were accepted. Within the extremely well structured reviewing process, every paper was at least reviewed by three members of the program committee (PC). The PC consisted of 142 members; this year, special emphasis was laid on a balanced structure of the committee by appointing experienced senior researchers as well as young researchers starting out in a career in academia. Every PC member’s review workload was about eight and nine papers. In addition, every paper was assigned to a topic area. ICDE 2015 was structured into 15 topic areas handled by individual area chairs. With topics ranging from Big Data system architectures via modern hardware support to temporal, spatial mobile data management, ICDE 2015 provided a rich bouquet of specialized areas managed by internationally renowned experts of the field. In addition to research papers, the conference also offered an industrial track with 15 papers accepted out of 33 submissions, an application track with four papers (11 submissions) as well as a demo track with 73 submissions and 29 accepted demos resulting in an acceptance rate of 39.7 percent.

Out of the 84 papers presented at the conference, five outstanding papers were selected and the authors were offered the opportunity to provide an extended version to be published in this special section of the *IEEE Transactions on Knowledge and Data Engineering*. After an additional—and again—rigorous review process, all papers are now substantially extended in comparison to their conference version. Those five papers cover a wide range of topics and very nicely reflect the breadth of research work in the context of data management.

The paper “Understand Short Texts by Harvesting and Analyzing Semantic Knowledge” by Wen Hua, Zhongyuan

- W. Lehner is with the Department of Computer Science, Technische Universität Dresden, Dresden D-01062, Germany.
E-mail: wolfgang.lehner@tu-dresden.de.
- J. Gehrke is with Microsoft, Destiny, NY 13204.
E-mail: johannes@cs.cornell.edu.
- K. Shim is with the Department of Electrical and Computer Engineering, Seoul National University, Seoul 151-742, Korea.
E-mail: kshim@snu.ac.kr.

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TKDE.2016.2639978

Wang, Haixun Wang, Kai Zheng, and Xiaofang Zhou addresses the challenge of how to understand short texts, which is crucial to many applications, because they do not always observe the syntax of a written language or do not contain sufficient statistical signals to support many state-of-the-art approaches for text mining such as topic modeling. The authors present a prototype system for short text understanding which exploits semantic knowledge provided by a well-known knowledgebase automatically harvested from a web corpus including a comprehensive performance evaluation on real-life data.

The paper “Network Motif Discovery: A GPU Approach” by Wenqing Lin, Xiaokui Xiao, Xing Xie, and Xiao-Li Li tackles the challenge of identifying network motifs as important applications in numerous domains. This paper presents a novel study on network motif discovery using GPUs in order to parallelize a large number of subgraph matching tasks in computing subgraph frequencies from random graphs. The paper explores the design space of GPU-based subgraph matching algorithms, with careful analysis of several crucial factors (such as branch divergences and memory coalescing) that affect the performance of GPU programs.

The paper “Bump Hunting in the Dark: Local Discrepancy Maximization on Graphs” by Aristides Gionis, Michael Mathioudakis, and Antti Ukkonen studies the problem of discrepancy maximization on graphs to identify a connected subgraph that contains many more nodes from a set of nodes of an underlying graph than other nodes. Considering two different access models, the authors prove that the basic problem of discrepancy maximization on graphs is NP-hard and empirically evaluate the performance of four heuristics for solving it.

The paper “Enumeration of Maximal Cliques from an Uncertain Graph” by Arko Mukherjee, Pan Xu, and Srikanth Tirthapura considers the enumeration of dense substructures (maximal cliques) from an uncertain graph. The paper presents matching upper and lower bounds on the number of α -maximal cliques possible within an uncertain graph by providing an algorithm to enumerate α -maximal cliques whose worst-case runtime is near-optimal. An experimental evaluation shows the practical utility of the algorithm.

The paper “Efficient Metric Indexing for Similarity Search and Similarity Joins” by Lu Chen, Yunjun Gao, Xinhan Li, Christian S. Jensen, and Gang Chen proposes a versatile and efficient disk-based index for metric data, the space-filling curve, and pivot-based B+-tree called SPB-tree in order to better support spatial queries including similarity search and similarity joins. The paper covers the

presentation of the algorithm as well as corresponding cost models and compares the data structure within an extensive set of experiments using both real and synthetic data.

CONCLUSION

Finally, we would like to thank all of the authors of this special section for providing an extended version of their conference paper. We also thank the reviewers for fulfilling the additional task of comparing and reviewing the submissions. Looking back at ICDE 2015, we would also like to thank all members of the program committee and especially all area chairs for their effort in making the research program of ICDE 2015 a great success. We are convinced that this special section provides a very nice collection of high-quality research papers in the field of data engineering.

Wolfgang Lehner
Johannes Gehrke
Kyuseok Shim

Guest Editors and ICDE 2015 Program Committee Chairs



Wolfgang Lehner is a full professor and head of the Database Technology Group, Technische Universität Dresden (TU Dresden), Germany. He was a PostDoc with IBM Almaden, San Jose, CA, as well as a visiting scientist with Microsoft Research in Redmond, WA, with GfK Nuremberg, with SAP Walldorf, and with SAP Palo Alto as well as with the University of Waterloo. He is a member of the VLDB Endowment as well as the spokesman of the Review Board on Computer Science of DFG (German Research Foundation).

Since 2014, he has also been a member of Academia Europaea (The Academy of Europe). His research interests range from DB system design for exploiting the opportunities of main-memory centric database architectures with an emphasis on modern hardware capabilities (including NVRAM support, HW/SW CoDesign) to supporting data-intensive applications and processes in large distributed information systems. Apart from basic and mostly theoretic research questions, he puts a strong emphasis on practical research work in combination with industrial cooperations.



Johannes Gehrke is a distinguished engineer with Microsoft working on data science and the distributed systems infrastructure in Office 365. He is part of the senior leadership team of Office 365. Until 2015, he was the Tisch University Professor in the Department of Computer Science at Cornell University doing research on database systems, distributed systems, data science, and data privacy. He has received a US National Science Foundation Career Award, an Arthur P. Sloan Fellowship, an IBM Faculty Award, the Cornell College of Engineering James and Mary Tien Excellence in Teaching Award, the Cornell University Provost's Award for Distinguished Scholarship, a Humboldt Research Award from the Alexander von Humboldt Foundation, the 2011 IEEE Computer Society Technical Achievement Award, and the 2011 Blavatnik Award for Young Scientists from the New York Academy of Sciences. He co-authored the undergraduate textbook *Database Management Systems* (McGrawHill, 2002, currently in its third edition), used at universities all over the world. He was co-chair of KDD 2004, VLDB 2007, ICDE 2012, SOCC 2014, and ICDE 2015.



Kyuseok Shim received the BS degree in electrical engineering from Seoul National University, in 1986, and the MS and PhD degrees in computer science from the University of Maryland, College Park, in 1988 and 1993, respectively. He is currently a professor in the Electrical and Computer Engineering Department, Seoul National University, Korea. Before that, he was an assistant professor in the Computer Science Department, KAIST, and a member of technical staff for the Serendip Data Mining Project at Bell Laboratories.

He was also a member of the Quest Data Mining Project, IBM Almaden Research Center, and has visited Microsoft Research in Redmond, WA, several times as a visiting scientist. He was named an ACM Fellow for his contributions to scalable data mining and query processing research in 2013. He has been working in the area of databases focusing on data mining, search engines, recommendation systems, MapReduce algorithms, privacy preservation, query processing, and query optimization. His writings were published in a number of professional conferences and journals including ACM, VLDB, and IEEE publications. He served as a program committee member for SIGKDD, SIGMOD, ICDE, ICDM, ICDT, EDBT, PAKDD, VLDB, and WWW conferences. He also served as a program committee co-chair for PAKDD 2003, WWW 2014, ICDE 2015, and APWeb 2016. He was previously on the editorial boards of VLDB as well as the *IEEE Transactions on Knowledge and Data Engineering* and is currently a member of the VLDB Endowment Board of Trustees.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.