Guest Editorial Introduction to the Special Section on the 16th International Conference on Data Engineering

David B. Lomet and Gerhard Weikum

1 INTRODUCTION

The 16th International Conference on Data Engineering (ICDE 2000) was held in San Diego, California, from 29 February through 3 March 2000. As the first major database conference of the new millennium, it marked the threshold of a new database era with a proliferation of exciting data-intensive, network-centric applications, and increasing penetration of database technology into the software infrastructure of cyberspace. The conference technical program reflected this, covering topics ranging from data mining and knowledge discovery to XML, e-commerce, and mobile computing, while also capturing advances in traditional and still critical engine-technology.

The 287 submissions for the ICDE technical program required us to be very selective, but also resulted in an outstanding program. We accepted 41 research papers. In addition, 24 submissions were accepted for poster sessions. The technical program also included industrial sessions, panels, demos, and tutorials. The keynote addresses, by Jim Gray on “rules of thumb” and Dennis Tsichritzis on the “art of computer research” were thought-provoking highlights.

Thanks to the active encouragement and cooperation of Farokh Bastani, editor-in-chief of the IEEE Transactions on Data and Knowledge Engineering, we were chartered to bring the best of the ICDE 2000 technical contributions to this special section of TDKE, ensuring that these fine contributions are widely available in an archival journal. This special section contains the six contributions selected as the best papers of ICDE 2000. We solicited extended versions from the authors, to provide full, thorough, and archival quality presentations. The extended papers were then reviewed a second time and we are convinced of their outstanding technical strength and presentation quality.

This special section leads off with the paper selected for the ICDE 2000 Best Paper Award. This paper is “Automating Statistics Management for Query Optimizers” by Surajit Chaudhuri and Vivek Narasayya. It addresses the often neglected but pragmatically important area of automating system administration. The specific problem investigated is how to automatically select the columns of a relational database to maintain in histograms. An elegant solution is devised for choosing those columns that are essential to effective query optimization, and the solution is experimentally studied.

The second paper, “A Foundation for Conventional and Temporal Query Optimization Addressing Duplicates and Ordering” by Giedrius Slivinskas, Christian Jensen, and Richard Snodgrass, develops a general algebraic framework that includes powerful aggregation and sorting constructs that integrate conventional and temporal query processing. The framework is sufficiently flexible to enable it to accommodate different architectural approaches to the support of temporal functionality.

Our special section would be incomplete without papers from the increasingly important field of data mining. The first data mining paper, “DEMON: Mining and Monitoring Evolving Data” by Venkatesh Ganti, Johannes Gehrke, and Raghu Ramakrishnan, won the ICDE 2000 Best Student Paper Award. It addresses the problem of incremental mining where the underlying data gradually changes over time and it proposes “incremental” mining algorithms that exploit this gradual evolution for increased efficiency. The incremental approach is applied to a range of problems including discovery of association rules and clustering.

The other mining paper, “Finding Interesting Associations without Support Pruning” by Edith Cohen, Mayur Datar, Shinji Fujiwara, Aristides Gionis, Piotr Indyk, Rajeev Motwani, Jeffrey Ullman, and Cheng Yang, reconsiders association rules, e.g., on market baskets or web traffic logs. The new insight, motivated by these discovery applications, is that requiring a minimum support is counter-productive. But, removing this requirement interferes with pruning the search space. The paper develops a family of randomized algorithms that find the most interesting rules, regardless of support, with probability converging to one.

Our final two papers fall into the area of spatial and multimedia data management. The first one, “Multiple Similarity Queries: A Basic DBMS Operation for Mining in Metric Databases” by Bernhard Braunmueller, Martin Ester, Hans-Peter Kriegel, and Joerg Sander, aims at improving the performance of similarity queries, e.g., finding the k nearest neighbors of a given data object, on spatial and multimedia
databases. The paper introduces a generic scheme for combining multiple, independently sectioned similarity queries into a single index traversal and data scan and presents measurements documenting the resulting performance improvement.

Finally, the paper “On the ‘Dimensionality Curse’ and the ‘Self-Similarity Blessing’” by Flip Korn, Bernd-Uwe Pagel, and Christos Faloutsos tackles the problem of execution cost estimation for similarity queries and the underlying analysis of data properties. The paper develops accurate formulas for certain realistic data distributions that characterize the distances to neighbors in high dimensional data spaces and permit accurate I/O cost estimations. Convincing experimental evidence for the adequacy of this analysis is presented.

We would like to thank both the authors and our otherwise unrecognized referees for their hard work, cooperation, and responsiveness in the timely production of this special section. Finally, we want to express our appreciation to the members of the ICDE steering committee, organizing committee, and program committee, and to the participants of ICDE 2000 for making the conference both technically outstanding and socially enjoyable.

David B. Lomet
Gerhard Weikum

David B. Lomet received a PhD degree in computer science from the University of Pennsylvania. He has been a senior researcher and manager of the Database Group at Microsoft Research, Redmond, Washington since 1995. Prior to joining Microsoft, he was a senior consulting engineer at Digital Equipment Corporation’s Cambridge Research Lab (CRL), having worked previously in Digital’s database product group. Earlier in his career, he was a research staff member at the IBM Thomas J. Watson Research Center in Yorktown Heights, New York, and, subsequently, a professor of information technology at the Wang Institute of Graduate Studies. Dr. Lomet spent a sabbatical at the University of Newcastle-upon-Tyne working on software reliability with Brian Randell. He is best known for his work in database systems and transactions. He is one of the inventors of the transaction concept. His earlier work included research and product engineering in machine architecture, programming languages and compilers, and distributed systems. His work in database systems has focused on access methods, concurrency control, and recovery. He is an author of more than 60 conference or journal papers and holds 19 patents. He has twice been a SIGMOD “best paper” award author.

Dr. Lomet has served on major conference program committees, including SIGMOD, PODS, VLDB, ICDE, and EDBT. He was program committee cochair for ICDE 2000 and PC chair of the FODO ’93 conference. He has participated in several “research directions” workshops sponsored by both the US National Science Foundation and the ACM and wrote summaries for two of them. Dr. Lomet has been editor-in-chief of the Data Engineering Bulletin since 1992 and is an editor of the ACM Transactions on Database Systems, the VLDB Journal, and the Journal of Distributed and Parallel Databases. He is an IEEE Golden Core Member and has received an IEEE Outstanding Contribution Award.

Gerhard Weikum received the diploma degree (Dipl.-Inform.) and the doctoral degree (Dr.-Ing.), both in computer science, from the University of Darmstadt, Germany, in 1982 and 1986, respectively. Since April 1994, Dr. Weikum has been a full professor in the Department of Computer Science of the University of the Saarland at Saarbruecken, Germany, where he is leading a research group on database systems and information systems. During his sabbatical in the summer of 1997, he was a visiting senior researcher in the Database Group at Microsoft Research in Redmond, Washington. Earlier, he held positions at ETH Zurich, Switzerland, and with Microelectronics and Computer Technology Corporation in Austin, Texas. Dr. Weikum’s research interests include parallel and distributed information systems, multimedia data management, transaction processing and workflow management, and database optimization and performance evaluation. Dr. Weikum serves on the editorial boards of the ACM Transactions on Database Systems, the VLDB Journal, the Distributed and Parallel Databases Journal, the ACM SIGMOD Digital Review, and the Data Engineering Bulletin, and has served on numerous program committees of international conferences. He was a program committee cochair of the Fourth ACM International Conference on Parallel and Distributed Information Systems in 1996 and the 16th IEEE CS International Conference on Data Engineering in 2000. He is also a member of the board of trustees of the VLDB Endowment, the steering committee of one of the premier international database conferences. Dr. Weikum has published more than 70 refereed papers in journals and conferences.