

M. Tamer Özsu · Donald Kossmann · Renée J. Miller

Special issue: Best papers of VLDB 2004

Published online: 30 October 2005
© Springer-Verlag 2005

This special issue includes four papers from the 30th International Conference on Very Large Data Bases, held in Toronto, Canada, 31 August—3 September 3 2004. The conference program included 81 research papers selected from 463 submissions. Eight papers from these 81 were invited by the Program Chairs for inclusion in this special issue. Substantially expanded and enhanced versions of these papers were subjected to a new round of reviewing and, as a result, the four that are included in this issue were selected.

The four papers cover a wide range of topics: sensor networks, data warehouses, data quality, and scientific databases.

The first paper “AWESOME—A Data Warehouse-based System for Adaptive Website Recommendations” by A. Thor, N. Golovin, and E. Rahm, deals with recommendations that a Web system could make to users regarding additional sites they can visit. Recommendation systems are important in large Web systems where navigating within the site may be difficult. Examples of these are electronic commerce sites such as Amazon that offer a large number of products accessed by a large number of users with differing profiles. Recommendation systems guide each user, based on a number of factors, to those parts of the Web site that may be relevant to their needs or requests. The paper proposes a recommendation system that dynamically chooses

certain recommendation algorithms from a library of recommendation algorithms that it maintains.

The second paper, “An Annotation Management System for Relational Databases” by D. Bhagwat, L. Chiticariu, W.-C. Tan, and G. Vijayvargiya addresses a data quality issue in relational databases, the management of annotations. Their system permits annotations on data to be automatically propagated as the data is transformed by queries. Annotations are an effective tool for understanding the provenance and flow of data, permitting scientistis or data curators to verify the origin or quality of the data. The paper develops techniques for managing annotations in a relational database system and proposes an extension to SQL for propagating annotations.

The third paper in this special issue, “Algebraic Manipulation of Scientific Datasets” by Bill Howe and David Maier focus on the long-standing problem of supporting the data management requirements of scientific applications. It characterizes scientific datasets by their “topological structure” (which the authors call grid) over which they are defined—time series is defiend over a one-dimensional grid while the solution to a partial differential equation using finite-element method might be defined over a three-dimensional grid. The paper inviestigates the algebraic processing strategies for large scientific datasets that possibly have an irregular grid structure. It defines an algebra of grid-fields for manipulating arbitrary gridded datasets, proposes optimization techniques for this algebra, and describes an implementation.

One of the fastest growing applications areas of database technology are stream data, in particular as they are generated by sensor networks. The final paper, “Model-Driven Data Acquisition in Sensor Networks” by A. Deshpande, C. Guestrin, S.R. Madden, J.M. Hellerstein, and W. Hong addresses a particular issue within this context, namely that of properly modeling sensor data as a database where the database does not (contrary to traditional systems) capture the entire information. In these environments, two problems are prevalent in modeling data: misrepresentation of data since it is impossible to gather all the relevant data, and inefficient approximate querying where the results to the queries

M. T. Özsu (✉)
(Technical Program Chair)
School of Computer Science, University of Waterloo,
200 University Avenue West, Waterloo, ON N2L 3G1, Canada
E-mail: tozsu@uwaterloo.ca

D. Kossmann
(Core Database Technology Program Chair)
ETH Zurich, Institut f. Informationssysteme, ETH-Zentrum, Univer-
sitätstrasse 6, CH-8092 Zürich, Switzerland
E-mail: donald.kossmann@inf.ethz.ch

R. J. Miller
(Infrastructure for Information Systems Program Chair)
Department of Computer Science, University of Toronto, Toronto,
Ontario M5S 2E4, Canada
E-mail: miller@cs.toronto.edu

are expected to be approximate, but the execution techniques that are typically employed assume completeness. The paper proposes techniques to compensate for both of these deficiencies by incorporating statistical models of real-world processes.

Since this is the final task of the Program Chairs, we would like to extend our thanks to everyone who worked hard to make VLDB 2004 a success. In particular, we would like to thank the 137 colleagues who served on the

three Program Committees that were formed along themes (core database, infrastructure for information systems, and industrial and applications). They had to carry significant load within a short period of time, and some of them had to do extra work for further reviewing the papers that were selected for consideration for this special issue.

M. Tamer Özsu · Donald Kossmann · Renée J. Miller