

Special issue on best papers of VLDB 2011

Wolfgang Lehner · Sunita Sarawagi

Published online: 4 December 2012
© Springer-Verlag Berlin Heidelberg 2012

The VLDB conference is one of the most renowned venues for presenting advances in the research and practice of data management. The VLDB 2011 conference took place from August 29 to September 3 in Seattle (WA, USA) with more than 750 participants.

Out of 553 submissions to the research track, 100 papers got accepted for presentation at the VLDB 2011 conference. Of these, the following five papers were invited to deliver an extended version to be published in this special issue of VLDBJ.

- Efficiently Adapting Graphical Models for Selectivity Estimation.
by Kostas Tzoumas (TU Berlin), Amol Deshpande, Christian S. Jensen.
- RemusDB: Transparent High Availability for Database Systems.
by Umar Farooq Minhas (University of Waterloo), Shriram Rajagopalan, Brendan Cully, Ashraf Abounaga, Kenneth Salem, Andrew Warfield.
- OXPATh: A Language for Scalable Data Extraction, Automation, and Crawling on the Deep Web.
by Tim Furche (Oxford University), Georg Gottlob, Giovanni Grasso, Christian Schallhart, Andrew Sellers.
- Automating the Database Schema Evolution Process.
by Carlo A Curino (Microsoft), Hyun J Moon, Alin Deutsch, Carlo Zaniolo.

- Keyword Search on Form Results.
by Aditya Ramesh, S. Sudarshan (IIT Bombay), Purva Joshi, Manisha Naik Gaonkar

The selection covers a wide spectrum of database system-related topics, ranging from query optimization to information retrieval and model management. Also, the papers come from research groups all over the world: Redmond, Waterloo, Oxford, Berlin, and Bombay; a truly international mix! All contributions are significantly extended and improved with respect to the original conference version.

We hope you enjoy these “Best of VLDB 2011” papers—may the papers spark novel research ideas!!!

The first paper “EFFICIENTLY ADAPTING GRAPHICAL MODELS FOR SELECTIVITY ESTIMATION” of this special issue focuses on a classical problem within database query optimization: how to increase the accuracy of a cost model? Since traditional query optimization techniques are based on the independence assumption of individual columns, even small correlations in the database may result in significant estimation errors and yield a non-optimal plan for a given query. The paper takes a principled and practical approach to the problem by relying on joint probability distributions over pairs of columns within a database. The work also includes an implementation within the PostgreSQL DBMS. Experiments show a significant improvement of query plan quality (with respect to the query runtime) by a moderate increase in optimization time.

The second paper on “REMUSDB: TRANSPARENT HIGH AVAILABILITY FOR DATABASE SYSTEMS” presents a novel approach to achieve high availability (HA) for database systems using techniques from operating system. The presented solution is based on Remus, a virtualization layer with core functionality to offer high availability using efficient state replication of virtual machine images. With RemusDB,

W. Lehner (✉)
Technische Universität Dresden, Nöthnitzer Str. 46,
01187 Dresden, Germany
e-mail: wolfgang.lehner@tu-dresden.de

S. Sarawagi
Computer Science and Engineering, IIT Bombay, Kanwal Rekhi
Building, Powai, Mumbai 400076, India
e-mail: sunita@iitb.ac.in

the authors extended the Remus framework making it database aware by optimizing the replication scheme with respect to tasks only relevant to achieve HA on the database level. Experiments include benchmarks using two DBMSes showing only small overhead in combination with a fast failover. This paper triggers an interesting discussion on whether database services are supposed to be provided by the DBMS itself (the traditional way) or covered by an underlying infrastructure.

The contribution “OXPATH: A LANGUAGE FOR SCALABLE DATA EXTRACTION, AUTOMATION, AND CRAWLING ON THE DEEP WEB” provides a new language to cope with the growing variety of information sources on the Web. The paper tackles four different challenges in the context of web data extraction: interaction with different interfaces, capturing precision of extracted data, scaling with the number of web data sources, and embedding into existing technology stacks. The contribution outlines an extension of XPath to significantly improve the web extraction process. Insights into the implementation of these techniques and results of simulation and real benchmark experiments are given showing that OXPath outperforms existing approaches substantially.

The fourth paper in this special issue may be considered a representative of the wide research field of schema and model management. In the paper “AUTOMATING THE DATABASE SCHEMA EVOLUTION PROCESS”, the authors

focus on automating perspective of schema evolution when migrating a database and rewriting queries and updates. Their approach is based on PRISM/PRISM++ providing the user with a set of schema modification operators plus constraint mapping operators to reflect the migration of existing constraints to the target database model. In addition to the theoretical framework, the paper also outlines two novel tools to automatically collect and provide statistics and recommendation with respect to schema evolution tasks. Experiments with almost 2000 schema evolution steps and over a 100 years of evolution history show the feasibility of the overall approach taken.

The final paper on “KEYWORD SEARCH ON FORM RESULTS” investigates the problem of keyword search in enterprise applications considering explicit and implicit form parameters provided by users. The paper presents a technique to “invert” SQL queries based on the queries attached to the form. The paper gives a comprehensive insight into their implementation and reports on good performance in the presence of large databases and large number of forms.

We thank the authors of the invited papers for investing substantial time and effort into extending the original conference version of the papers. We would also like to express our sincere Thank You to the diligent reviewers who provided perspective, advice, and keen assessment of the submissions. We hope you enjoy this issue.