

Special issue on best papers of VLDB 2013

Michael H. Böhlen · Christoph Koch

Published online: 10 September 2015
© Springer-Verlag Berlin Heidelberg 2015

VLDB is a premier annual international forum for data management, and it targets database researchers, users, vendors, and application developers. Each year the conference covers current issues in data management and database systems, which remain key technological cornerstones of emerging applications of the twenty-first century.

VLDB 2013 took place at the picturesque town of Riva del Garda in Italy. We received a total of 559 research paper submissions, and of these 127 submissions were accepted for presentation at the VLDB conference. A committee that included Vanja Josifovski, Mohamed Mokbel, Dan Olteanu, Ken Salem, Divesh Srivastava, and Jens Teubner selected the best papers that have been submitted to VLDB 2013. The authors of these papers were invited to submit an extended version of their papers to a special issue of the VLDB Journal. The result of this process is the following four papers:

- Active Learning in Keyword Search-Based Data Integration, Zhepeng Yan (University of Pennsylvania), Nan Zheng (University of Pennsylvania), Zachary G. Ives (University of Pennsylvania), Partha Pratim Talukdar (SERC Indian Institute of Science), Cong Yu (Google Research).
- ClouDiA: A Deployment Advisor for Public Clouds, Tao Zou (Cornell University), Ronan Le Bras (Cornell University), Marcos Vaz Salles (University of Copenhagen),

Alan Demers (Cornell University), Johannes Gehrke (Cornell University).

- Scheduled Approximation and Incremental Enhancement for Accuracy-aware Personalized PageRank, Fanwei Zhu (Zhejiang University City College), Yuan Fang (Institute for Infocomm Research, Singapore), Kevin C. Chang (University of Illinois at Urbana-Champaign), Jing Ying (Zhejiang University City College).
- VLL: A Lock Manager Redesign for Main Memory Database Systems, Kun Ren (Yale University), Alexander Thomson (Google), Daniel J. Abadi (Yale University).

The first paper, “Active Learning in Keyword Search-Based Data Integration,” proposes active learning to improve keyword-based data integration. The approach is to avoid a global schema and instead develop keyword search-based data integration where the system lazily discovers associations to join keyword matches and to return ranked results. The paper describes the Q-system that performs keyword search over databases based on feedback about the answers. The user plays an active role as (s)he is expected to understand the data domain and provide feedback about the quality of the answers. The Q-system generalizes such feedback to learn how to correctly integrate data. The main distinguishing feature of the approach is to consider not only the relevance of query results but complementing it with uncertainty and informativeness of feedback deriving from those results. The combinatorial problem under keyword search over structured data is a Steiner tree problem, and the feedback is on these structured results. The paper also considers the role of diversity and shows that diversity hurts more if searches become narrow. The paper proposes several techniques for estimating uncertainty and relevance of a query result, for ranking answers to a keyword query, for learning from user feedback, and for adjusting issues related to overlapping query results

M. H. Böhlen (✉)
Database Technology Group, Department of Informatics, University of
Zürich, Binzmühlestrasse 14, 8050 Zürich, Switzerland
e-mail: boehlen@ifi.uzh.ch
URL: <http://www.ifi.uzh.ch/dbtg/Staff/Boehlen.html>

C. Koch
School of Computer and Communication Sciences, Ecole polytechnique
fédérale de Lausanne, EPFL IC IIF DATA BC 260 (Bâtiment
BC) Station 14, 1015 Lausanne, Switzerland
URL: <http://people.epfl.ch/christoph.koch>

by means of a proper diversity management of the top-k query answers. The approach is validated over real data from several very different domains.

The second paper, “ClouDiA: A Deployment Advisor for Public Clouds,” proposes an advisor for mapping components of a distributed application to virtual machines in an IaaS environment like Amazon EC2. The authors observe that cloud providers allocate virtual machine instances non-contiguously to achieve high utilization, i.e., instances of a given application may end up in physically distant machines in the cloud. Clearly such an allocation strategy can lead to large differences in average latency between instances. ClouDiA addresses this problem by selecting application node deployments that minimize either the largest latency between application nodes or the longest critical path among all application nodes. Application components are assumed to communicate with each other according to a known pattern, and the advisor attempts to determine the mapping of the components that minimizes an objective function over the communication cost. ClouDiA employs a number of algorithmic techniques, including mixed-integer programming and constraint programming techniques, to efficiently search the space of possible mappings of application nodes to instances. Experiments with synthetic and real applications in Amazon EC2 show that mean latency is a robust metric to model communication cost in these applications.

The third paper, “Scheduled Approximation and Incremental Enhancement for Accuracy-aware Personalized PageRank,” describes an incremental and exponentially bounded approximation of personalized page rank. It is based on using the concept of inverse P-distance, but in an approximate and incremental way by considering the paths based on their length. The concept of hubs is introduced to make

the problem tractable. The authors use hubs to decompose the computation into a number of iterations based on the number of visited hubs along personalized page rank walks and use the Chapman–Kolmogorov equation to reuse computations across iterations. The authors explore multiple hub selection techniques to improve the performance of the algorithm. The paper is a good example of algorithmic work that can be practically useful and provides complexity and error bounds. The experiments using the new hub selection show significant improvement over previously reported work.

The fourth paper, “VLL: A Lock Manager Redesign for Main Memory Database Systems,” proposes a very lightweight locking (VLL) mechanism for database systems. The authors propose a global job queue, instead of a queue for each locked resource, to get a very lightweight locking for main-memory database systems. The proposed approach dramatically decreases locking overhead, while having only a modest impact on concurrency. Conflict queues are replaced by conflict counts, which are much more lightweight. The invocation of the selective conflict analysis is an elegant and useful technique to get a high transactional throughput. The ability to add range locking to VLL (resulting in VLLR), and the way it is done, is a novel idea as is the way the ranges are defined to which the counts apply. The performance experiments that are reported are extensive and convincing.

We thank the authors of the invited papers for investing substantial time and effort into extending the original conference version of the papers and have them reviewed again. We would also like to express our sincere thanks to the diligent reviewers who provided perspective, constructive advice, and assessment of the submissions. We hope you enjoy this special issue of the best papers of VLDB 2013.