Database Theory Column*
Report on PODS 2018

Marcelo Arenas

PODS 2018 PC Chair

Pontificia Universidad Católica & IMFD, Chile

marenas@ing.puc.cl

November 20, 2018

The 37th edition of the ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS) took place from June 11 to June 13, 2018, in Houston, Texas, USA. As in previous years, the symposium was held jointly with the ACM SIGMOD International Conference on Management of Data (SIGMOD). PODS focuses on theoretical aspects of data management systems and techniques, and the co-location with SIGMOD stimulates interaction between theory-oriented and system-oriented research.

The proceedings of PODS 2018 are published by ACM Press, and can be found on the ACM Digital Library (`https://dl.acm.org/citation.cfm?id=3196959`). The conference program included a keynote talk by Michael Benedikt (University of Oxford), two invited tutorials by Arvind Narayanan (Princeton University) and Rajeev Raman (University of Leicester), and 29 contributed papers selected by the Program Committee from 83 submissions.

Michael Benedikt's keynote talk, titled "How can reasoners simplify database querying (and why haven't they done it yet)?", can be watched at `https://www.youtube.com/watch?v=jvYahnh9PRg`. This keynote was a reflection on the state of tools and algorithms for applying reasoning to simplify database querying, from within and outside database theory. Within database theory, a large amount of research has been devoted to reasoning problems, such as query containment, equivalence, simplification and rewriting. Thus, a first motivation for this tutorial was to reflect upon what has been done, highlighting some accomplishments as well as limitations. Outside the database area, there has been tremendous progress in the development of reasoning systems, which has included not only advances in theory, but also in system building and the development of a culture of experimental evaluation. Thus, the second motivation for this tutorial was to reflect upon how much the development of these reasoning systems could benefit the work on reasoning for query simplification.

In the tutorial by Arvind Narayanan, titled "Blockchains: past, present, and future", Arvind first gave a detailed account of how blockchain technology is a careful assembly of pieces that have been studied before in the literature, such as linked timestamping, consensus,

---

and proof of work. Then he presented some abstract models of blockchains, which helped the audience understand the similarities and differences between proposed blockchain designs. Finally, Arvind described several research directions, focusing on those of interest to the database community. In particular, it was mentioned in the tutorial that the blockchain is usually fully replicated by every node participating in the protocol, leading to massive inefficiency and severely limiting transaction throughput. Thus, an interesting question to the database community is how to improve the way the blockchain is stored without weakening security.

The tutorial by Rajeev Raman, title "In-memory representations of databases via succinct data structures", can be watched at `https://www.youtube.com/watch?v=_3q7T_5JaTI`. In this tutorial, Rajeev gave an overview of the growing field of succinct data structures. Such structures allow to compress the data so that they can be stored in main memory, and with the additional and fundamental feature of supporting operations on the data with little or no slow-down compared to their conventional counterparts. In this respect, succinct data structures can be applied to a variety of problems in the database area, some of which were explored in this tutorial.

Among the contributed papers, the Program Committee selected "Entity matching with active monotone classification" by Yufei Tao (Chinese University of Hong Kong) for the PODS 2018 Best Paper Award. Given two sets X and Y of entities, the entity matching problem is the task of deciding, given $x \in X$ and $y \in Y$, whether x and y represent the same entity. This is a fundamental problem, which is usually solved using approaches that have the following classification task at their cores. The input is a set P of points in $\mathbb{R}^d$, each of which carries a binary label 0 or 1. A classifier is a function from $\mathbb{R}^d$ to $\{0, 1\}$, and the objetive is to find a classifier that correctly gives the labels of a large number of point in P. In this paper, the author represents this classification task as an instance of active learning where the goal is to learn a monotone classifier. In this formulation, the labels of all points in P are initially hidden, and a learning algorithm can invoke an oracle to know the label of a point $p \in P$. Such an oracle models the help of an external source, which as the last resort can be a human expert called upon to decide the label of a point $p \in P$ (or, more generally, to decide whether x and y represent the same entity). As such, the goal of the paper is to construct a learning algorithm with a good balance between accuracy and cost, where the cost is measured as the number of times the oracle is called. More specifically, the author describes algorithms with non-trivial guarantees on the cost and accuracy simultaneously, and he also proves lower bounds that establish the asymptotic optimality of the solutions proposed in the paper. The talk on this article by Yufei is available at `https://www.youtube.com/watch?v=AdvKLYkZ2es`.

Similarly to preceding years, the contributed papers covered a wide range of topics. These articles were clustered in sessions on: graphs and hypergraphs techniques on databases; similarity search and clustering; information extraction and efficient enumeration of answers; consistent query answering, certain answers and repairs; query evaluation and containment; learning and streaming; algorithms, privacy and workflows. PODS 2018 further solicited contributions on new formal frameworks and the experimental validation of theoretical approaches. Hence, some of the papers in the conference were selected in these categories.

The PODS 2018 conference also presented the ACM PODS Alberto O. Mendelzon Test-of-Time Award. This award is given each year to one or a small number of papers pub-

lished in the PODS proceedings ten years prior that had the most impact over the intervening decade. This year's committee, consisting of Maurizio Lenzerini (Università di Roma La Sapienza), Wim Martens (University of Bayreuth) and Nicole Schweikardt (Humboldt-University Berlin), selected the paper "The chase revisited" by Alin Deutsch (UC San Diego), Alan Nash, and Jeffrey Remmel. The chase procedure is a famous technique in databases, which has proved very useful in providing solutions to several problems related to reasoning on data. In this respect, the paper makes several fundamental contributions to our understanding of this procedure. In particular, it revisits its standard version, studying its properties and applicability to classical database problems, and it settles the open problem of decidability of termination of the standard chase procedure. The talk on this article given by Alin can be seen at `https://www.youtube.com/watch?v=gzFIjlVGEKk&t=126s`.

In addition, PODS 2018 also included two Gems of PODS talks, which were meant to publicize to a wider audience approaches and results originating from the database theory community that have been highly influential. These two talks were selected by a committee consisting of Tova Milo (Tel Aviv University), Dan Olteanu (University of Oxford) and myself. The first selected talk was titled "Reflections on schema mappings, data exchange, and metadata management" and was presented by Phokion Kolaitis (UC Santa Cruz and IBM Almaden Research Center). The second selected talk was titled "Worst-case optimal join algorithms: techniques, results, and open problems" and was presented by Hung Ngo (RelationalAI). These talks can be watched at `https://www.youtube.com/watch?v=H6Ld8BCEIEw` and `https://www.youtube.com/watch?v=FmDsIVPCAbI`.

I would like to thank all invited speakers, authors, PC members, organizers and attendees for making PODS 2018 a wonderful and interesting conference. To conclude, the next PODS will take place in Amsterdam, The Netherlands, from July 1 to 3, 2019. As usual, PODS will be co-located with SIGMOD. Dan Suciu (University of Washington) will be the PODS General Chair and Christoph Koch (Swiss Federal Institute of Technology, Lausanne) will be the PC Chair. More information can be found at `http://sigmod2019.org/`.