Lecture Notes in Computer Science 10538

More information about this series at http://www.springer.com/series/7409

Zi Huang · Xiaokui Xiao · Xin Cao (Eds.)

Databases Theory and Applications

28th Australasian Database Conference, ADC 2017
Brisbane, QLD, Australia, September 25–28, 2017
Proceedings

≈ Springer

Editors
Zi Huang
University of Queensland
Brisbane, QLD
Australia

Xin Cao
University of New South Wales
Sydney, NSW
Australia

Xiaokui Xiao
Nanyang Technological University
Singapore
Singapore

Preface

It is our pleasure to present to you the proceedings of the 28th Australasian Database Conference (ADC 2017), which took place in Brisbane, Australia. The Australasian Database Conference is an annual international forum for sharing the latest research advancements and novel applications of database systems, data-driven applications, and data analytics between researchers and practitioners from around the globe, particularly Australia and New Zealand. The mission of ADC is to share novel research solutions to problems of today's information society that fulfill the needs of heterogeneous applications and environments and to identify new issues and directions for future research and development work. ADC seeks papers from academia and industry presenting research on all practical and theoretical aspects of advanced database theory and applications, as well as case studies and implementation experiences. All topics related to databases are of interest and within the scope of the conference. ADC gives researchers and practitioners a unique opportunity to share their perspectives with others interested in the various aspects of database systems.

As in previous years, the ADC 2017 Program Committee accepted those papers to be considered as being of ADC quality without setting any predefined quota. The conference received 32 submissions and accepted 22 papers, including 20 full research papers and two demo papers. Each paper was peer reviewed in full by at least three independent reviewers, and in some cases four referees produced independent reviews. A conscious decision was made to select the papers for which all reviews were positive and favorable. The Program Committee that selected the papers consists of 41 members from around the globe, including Australia, China, Finland, Japan, Korea, New Zealand, Singapore, Switzerland, the UK, and the USA, who were thorough and dedicated to the reviewing process.

We would like to thank all our colleagues who served on the Program Committee or acted as external reviewers. We would also like to thank all the authors who submitted their papers and all the attendees. This conference is held for you, and we hope that with these proceedings, you can have an overview of this vibrant research community and its activities. We encourage you to make submissions to the next ADC conference and contribute to this community.

August 2017

Zi Huang
Xiaokui Xiao
Xin Cao

General Chair's Welcome Message

Welcome to the proceedings of the 28th Australasian Database Conference (ADC 2017)! ADC is a leading Australia- and New Zealand-based international conference on research and applications of database systems, data-driven applications, and data analytics. In the past decade, ADC has been held in Sydney (2016), Melbourne (2015), Brisbane (2014), Adelaide (2013), Melbourne (2012), Perth (2011), Brisbane (2010), Wellington (2009), Wollongong (2008), and Ballarat (2007). This year, the ADC conference came to Brisbane.

In the past, the ADC conference series was held as a part of the Australasian Computer Science Week (ACSW). Starting from 2014, ADC conferences have departed from ACSW as the database research community in Australasia has grown significantly larger. Now the new ADC conference has an expanded research program and focuses on community building through a PhD School. ADC 2017 was the fourth of this new ADC conference series.

The conference this year had three eminent speakers to give keynote speeches: Divesh Srivastava from AT&T Labs-Research (USA), Masaru Kitsuregawa co-affiliated with the National Institute of Informatics and the University of Tokyo (Japan), and Mingsheng Ying from the University of Technology Sydney (Australia). In addition to 22 papers carefully selected by the Program Committee, we were also fortunate to have a distinguished lecture by Dr. Guoliang Li from Tsinghua University (China), and two invited talks presented by Gianluca Demartini from the University of Sheffield (UK) and Dacheng Tao from the University of Sydney (Australia). Furthermore, we had a PhD School program with great support from four invited speakers: Divesh Srivastava from AT&T Labs-Research (USA), Yu Zheng from Microsoft Research (China), Gianluca Demartini from the University of Sheffield (UK), and Rui Zhang from the University of Melbourne (Australia).

We wish to take this opportunity to thank all speakers, authors, and organizers. I would especially like to thank our Organizing Committee members, the Program Committee co-chairs Helen Huang and Xiaokui Xiao, for their dedication to ensuring a high-quality program, proceedings chair Xin Cao, for his efforts in delivering the conference proceedings, local organization co-chairs Hongzhi Yin and Sen Wang, for their efforts in covering every detail of the conference logistics, publicity and Web chair Jun Zhou for his efforts in maintaining the conference website, tutorial and distinguished lecture chair Sebastion Link for his efforts in selecting and inviting the tutorial/lecture speakers, panel chair Athman Bouguettaya for his efforts in choosing the topic of an inspiring panel discussion, and PhD School coordinator Junhao Gan for designing an exciting program for the PhD school. I would also like to thank the University of Queensland for the support that it gave to the conference. Without them, this year's ADC would not have been a success.

Brisbane is a multicultural city, and ADC 2017 was held on the St. Lucia campus of the University of Queensland. We trust that all ADC 2017 participants had a wonderful experience with the conference, the campus, and the city.

Yufei Tao

Organization

General Chair

Yufei Tao   University of Queensland, Australia

Program Committee Co-chairs

Zi Huang   University of Queensland, Australia
Xiaokui Xiao   Nanyang Technological University, Singapore

Local Organization Chairs

Hongzhi Yin   University of Queensland, Australia
Sen Wang   Griffith University, Australia

Panel Chair

Athman Bouguettaya   University of Sydney, Australia

Publication Chair

Xin Cao   University of New South Wales, Australia

Publicity and Web Chair

Jun Zhou   Griffith University, Australia

Tutorial and Distinguished Lecture Chair

Sebastion Link   University of Auckland, New Zealand

PhD School Coordinator

Junhao Gan   University of Queensland, Australia

Steering Committee

Rao Kotagiri   University of Melbourne, Australia
Timos Sellis   RMIT University, Australia
Gill Dobbie   University of Auckland, New Zealand
Alan Fekete   University of Sydney, Australia
Xuemin Lin   University of New South Wales, Australia

| Yanchun Zhang | Victoria University, Australia |
| Xiaofang Zhou | University of Queensland, Australia |

Program Committee

Zhifeng Bao	RMIT University, Australia
Runyao Duan	University of Technology Sydney, Australia
Junbin Gao	University of Sydney, Australia
Janusz Getta	University of Wollongong, Australia
Gabriel Ghinita	University of Massachusetts, Boston, USA
Yusuke Gotoh	Okayama University, Japan
Yu Gu	Northeastern University, China
Wook-Shin Han	POSTECH, Korea
Michael E. Houle	National Institute of Informatics, Japan
Guangyan Huang	Deakin University, Australia
Arijit Khan	Nanyang Technological University, Singapore
Guoliang Li	Tsinghua University, China
Feifei Li	University of Utah, USA
Jianxin Li	University of Western Australia, Australia
Xue Li	University of Qeensland, Australia
Zhixu Li	Soochow University, China
Xiang Lian	Kent State University, UK
Jiaheng Lu	University of Helsinki, Finland
Muhammad Naeem	University of Auckland, New Zealand
Lu Qin	University of Technology Sydney, Australia
Goce Ristanoski	Data61, Australia
Shazia Sadiq	University of Queensland, Australia
Timos Sellis	Swinburne University of Technology, Australia
Haichuan Shang	National Institute of Information and Communications, Japan
Quan Z. Sheng	University of Adelaide, Australia
Jingkuan Song	University of Queensland, Australia
Bela Stantic	Griffith University, Australia
Farhan Tauheed	Oracle Labs Zurich, Switzerland
Anwaar Ulhaq	Victoria University, Australia
Junhu Wang	Griffith University, Australia
Yan Wang	Macquarie University, Australia
Yinghui Wu	University of California Santa Barbara, USA
Chuan Xiao	Nagoya University, Japan
Yang Yang	University of Electronic Science and Technology of China
Weiren Yu	Aston University, UK
Ying Zhang	University of Technology Sydney, Australia
Wenjie Zhang	University of New South Wales, Australia
Kai Zheng	Soochow University, China
Yi Zhou	University of Western Sydney, Australia

Lei Zou	Peking University, China
Feida Zhu	Singapore Management University, Singapore

External Reviewers

Sen Wang	Griffith University, Australia
Henry Nguyen	Griffith University, Australia
Saiful Islam	Griffith University, Australia
Binbin Gu	Griffith University, Australia
Jian Chen	Griffith University, Australia
Dong Wen	University of Technology Sydney, Australia
Xubo Wang	University of New South Wales, Australia
Seyed-Mohssen Ghafari	Macquarie University, Australia
Wei Emma Zhang	Macquarie University, Australia
Feng Zhu	Macquarie University, Australia

ADC Keynotes

Quantum Graph Reachability Problem

Mingsheng Ying

University of Technology Sydney, Australia

Abstract. Graph reachability is a fundamental problem in database theory and many other areas of computer science. In this talk, we consider quantum graph reachability problem, which originally arose in verification and analysis of quantum programs and model-checking quantum systems, but may also interest database community. We will discuss the following issues: 1. How we can naturally define a graph structure in the state Hilbert space of a quantum system from its (discrete-time) dynamics? 2. Why the approaches to classical graph reachability problem do not work for quantum reachability problem? 3. Strongly connected component decomposition theorem for quantum graphs. At the end of the talk, a series of open problems will be pointed out, including possible applications to database search in future quantum computers.

Short Biography. Mingsheng Ying was Cheung Kong Chair Professor at Tsinghua University and Director of the Scientific Committee, the National Key Laboratory of Intelligent Technology and Systems, China. Since 2008, he is Distinguished Professor and Research Director of the Centre for Quantum Software and Information, University of Technology Sydney, Australia. He is also Deputy Director for Research of the Institute of Software, Chinese Academy of Sciences.

Ying's research interests include quantum computation and quantum information, programming theory, and logical foundations of artificial intelligence. In particular, he developed Hoare logic for quantum programs and proved its (relative) completeness (TOPLAS'11). He defined the notion of invariants for quantum programs (POPL'17). He initiated the research line of model checking quantum Markov chains (CONCUR'12–14, TOCL'14). Ying is the author of Foundations of Quantum Programming (Morgan Kaufmann 2016).

Big Data Integration

Divesh Srivastava[1] and Masaru Kitsuregawa[2]

[1] AT&T Labs Research, USA
[2] National Institute of Informatics, University of Tokyo, Japan

Abstract. The Big Data era is upon us: data are being generated, collected and analyzed at an unprecedented scale, and data-driven decision making is sweeping through all aspects of society. Since the value of data explodes when it can be linked and fused with other data, addressing the big data integration (BDI) challenge is critical to realizing the promise of Big Data. BDI differs from traditional data integration in many dimensions: (i) the number of data sources, even for a single domain, has grown to be in the tens of thousands, (ii) many of the data sources are very dynamic, as a huge amount of newly collected data are continuously made available, (iii) the data sources are extremely heterogeneous in their structure, with considerable variety even for substantially similar entities, and (iv) the data sources are of widely differing qualities, with significant differences in the coverage, accuracy and timeliness of data provided. This talk presents techniques to address these novel challenges faced by big data integration, and identifies a range of open problems for the community.

Short Biography. Divesh Srivastava is the head of Database Research at AT&T Labs-Research. He is a Fellow of the Association for Computing Machinery (ACM) and the managing editor of the Proceedings of the VLDB Endowment (PVLDB). He has served as a trustee of the VLDB Endowment, as an associate editor of the ACM Transactions on Database Systems (TODS), as an associate Editor-in-Chief of the IEEE Transactions on Knowledge and Data Engineering (TKDE), and as a general or program committee co-chair of many conferences. He has presented keynote talks at several international conferences, and his research interests and publications span a variety of topics in data management. He received his Ph.D. from the University of Wisconsin, Madison, USA, and his Bachelor of Technology from the Indian Institute of Technology, Bombay, India.

Short Biography. Masaru Kitsuregawa received his Information Engineering Ph.D. degree from the University of Tokyo in 1983. Since then he joined the Institute of Industrial Science, the University of Tokyo, and is currently a professor. He is also a professor at Earth Observation Data Integration & Fusion Research Initiative of the University of Tokyo since 2010. He also serves Director General of National Institute of Informatics since 2013. Dr. Kitsuregawa's research interests include Database Engineering, and he had been principal researcher of Funding Program for World-Leading Innovative R&D on Science and Technology, MEXT Grant-in-Aids

Program for "Info-Plosion", and METI's Information Grand Voyage Project. He had served President of Information Processing Society of Japan from 2013 to 2015. He served as a committee member for a number of international conferences, including ICDE Steering Committee Chair. He is an IEEE Fellow, ACM Fellow, IEICE Fellow and IPSJ Fellow, and he won ACM SIGMOD E.F.Codd Contributions Award, Medal with Purple Ribbon, 21st Century Invention Award, and C&C Prize.

ADC Invited Talks

Human Computation
for Entity-Centric Information Access

Gianluca Demartini

The University of Sheffield, UK

Abstract. Human Computation is a novel approach used to obtain manual data processing at scale by means of crowdsourcing. In this talk we will start introducing the dynamics of crowdsourcing platforms and provide examples of their use to build hybrid human-machine information systems. We will then present ZenCrowd: a hybrid system for entity linking and data integration problems over linked data showing how the use of human intelligence at scale in combination with machine-based algorithms outperforms traditional systems. In this context, we will then discuss efficiency and effectiveness challenges of micro-task crowdsourcing platforms including spam, quality control, and job scheduling in crowdsourcing.

Short Biography. Dr. Gianluca Demartini is a Senior Lecturer in Data Science at the University of Sheffield, Information School. His research is currently supported by the UK Engineering and Physical Sciences Research Council (EPSRC) and by the EU H2020 framework program. His main research interests are Information Retrieval, Semantic Web, and Human Computation. He received the Best Paper Award at the European Conference on Information Retrieval (ECIR) in 2016 and the Best Demo Award at the International Semantic Web Conference (ISWC) in 2011. He has published more than 70 peer-reviewed scientific publications including papers at major venues such as WWW, ACM SIGIR, VLDBJ, ISWC, and ACM CHI. He has given several invited talks, tutorials, and keynotes at a number of academic conferences (e.g., ISWC, ICWSM, WebScience, and the RuSSIR Summer School), companies (e.g., Facebook), and Dagstuhl seminars. He is an ACM Distinguished Speaker since 2015. He serves as area editor for the Journal of Web Semantics, as Student Coordinator for ISWC 2017, and as Senior Program Committee member for the AAAI Conference on Human Computation and Crowdsourcing (HCOMP), the International Conference on Web Engineering (ICWE), and the ACM International Conference on Information and Knowledge Management (CIKM). He is Program Committee member for several conferences including WWW, SIGIR, KDD, IJCAI, ISWC, and ICWSM. He was co-chair for the Human Computation and Crowdsourcing Track at ESWC 2015. He co-organized the Entity Ranking Track at the Initiative for the Evaluation of XML Retrieval in 2008 and 2009. Before joining the University of Sheffield, he was post-doctoral researcher at the eXascale Infolab at the University of Fribourg in Switzerland, visiting researcher at UC Berkeley, junior researcher at the L3S Research Center in Germany, and intern at Yahoo! Research in Spain. In 2011, he obtained a Ph.D. in Computer Science at the Leibniz University of Hanover focusing on Semantic Search.

The Progress of AI

Dacheng Tao

University of Sydney, Australia

Abstract. Since the concept of Turing machine has been first proposed in 1936, the capability of machines to perform intelligent tasks went on growing exponentially. Artificial Intelligence (AI), as an essential accelerator, pursues the target of making machines as intelligent as human beings. It has already reformed how we live, work, learning, discover and communicate. In this talk, I will review our recent progress on AI by introducing some representative advancements from algorithms to applications, and illustrate the stairs for its realization from perceiving to learning, reasoning and behaving. To push AI from the narrow to the general, many challenges lie ahead. I will bring some examples out into the open, and shed lights on our future target. Today, we teach machines how to be intelligent as ourselves. Tomorrow, they will be our partners to get into our daily life.

Short Biography. Dacheng Tao is Professor of Computer Science and ARC Future Fellow in the School of Information Technologies and the Faculty of Engineering and Information Technologies at The University of Sydney. He was Professor of Computer Science and Director of the Centre for Artificial Intelligence in the University of Technology Sydney. He mainly applies statistics and mathematics to Artificial Intelligence and Data Science. His research interests spread across computer vision, data science, image processing, machine learning, and video surveillance. His research results have expounded in one monograph and 500+ publications at top journals and conferences, such as IEEE T-PAMI, T-NNLS, T-IP, JMLR, IJCV, IJCAI, AAAI, NIPS, ICML, CVPR, ICCV, ECCV, ICDM; and ACM SIGKDD, with several best paper awards, such as the best theory/algorithm paper runner up award in IEEE ICDM'07, the best student paper award in IEEE ICDM'13, and the 2014 ICDM 10-year highest-impact paper award. He received the 2015 Australian Scopus-Eureka Prize, the 2015 ACS Gold Disruptor Award and the 2015 UTS Vice-Chancellor's Medal for Exceptional Research. He is a Fellow of the IEEE, OSA, IAPR and SPIE.

Hybrid Human-Machine Big Data Integration (Distinguished Lecture)

Guoliang Li

Tsinghua University, China

Abstract. Data integration cannot be completely addressed by automated processes. We proposed a hybrid human-machine method that harnesses human ability to address this problem. The framework first uses machine algorithms to identify possible matching pairs and then utilizes the crowd to compute actual matching pairs from these candidate pairs. In this talk, I will introduce our two systems on hybrid human-machine big data integration. (1) DIMA: A distributed in-memory system on big-data integration that can use SQL to integrate heterogenous data. DIAM can be used to identify candidate matching pairs in big data integration. (2) CDB: A crowd-powered database system that provides declarative programming interfaces and allows users to utilize an SQL-like language for posing crowdsourced queries. CDB can be used to refine the candidate pairs in big data integration.

Short Biography. Guoliang Li is an Associate Professor of Department of Computer Science, Tsinghua University, Beijing, China. His research interests include crowd-sourced data management, large-scale data cleaning and integration, and big spatio-temporal data analytics. He has regularly served as the PC members of many premier conferences, such as SIGMOD, VLDB, KDD, ICDE, WWW, IJCAI, and AAAI. He was a PC co-chair of WAIM 2014, WebDB 2014, NDBC 2016, and an area chair of CIKM 2016-2017. He is an associated editor of some premier journals, such as TKDE, Big Data Research, and FCS. He has published more than 80 papers in premier conferences and journals, such as SIGMOD, VLDB, ICDE, SIGKDD, SIGIR, ACM TODS, VLDB Journal, and TKDE. His papers have been cited more than 3600 times. He received IEEE TCDE Early Career Award, and best paper awards/ nominations at DASFAA 2014 and APWeb 2014.

Contents

AI for Big Data

Potpourri

Demo Papers