
Advances in Database Technology — EDBT 2021

24th International Conference
on Extending Database Technology
Nicosia, Cyprus, March 23–26, 2021
Proceedings

Editors

Yannis Velegarakis
Demetris Zeinalipour
Panos K. Chrysanthis
Francesco Guerra



Advances in Database Technology – EDBT 2021
Proceedings of the 24th International Conference
on Extending Database Technology
Nicosia, Cyprus, March 23–26, 2021

Series ISSN: 2367-2005

Editors

Yannis Velegrakis, University of Trento, Italy and Utrecht University, Netherlands
Demetris Zeinalipour, University of Cyprus, Cyprus
Panos K. Chrysanthis, University of Cyprus, Cyprus and University of Pittsburgh, USA
Francesco Guerra, University of Modena and Reggio Emilia, Italy



OpenProceedings.org
University of Konstanz
University Library
78457 Konstanz, Germany

COPYRIGHT NOTICE: Copyright © 2021 by the authors of the individual papers.

Distribution of all material contained in this volume is permitted under the terms of the Creative Commons license CC-by-nc-nd 4.0

OpenProceedings ISBN: 978-3-89318-084-4

DOI of this front matter: 10.5441/002/edbt.2021.01

Foreword by the PC Chair

The International Conference on Extending Database Technology (EDBT) is an established forum for researchers and practitioners alike to disseminate knowledge and research results related to data management. This year, the 24th edition of EDBT, scheduled to take place between March 23rd and March 26, 2021 in Nicosia, Cyprus, has instead been held entirely online, due to the circumstances and the travel restrictions imposed by the 2020/2021 pandemic. It has been jointly organized with the International Conference on Database Theory (ICDT).

The organizing committee solicited contributions in many different areas, including, but not limited to, Data Preparation, Data Privacy, Database Engines, Distributed Data Systems, Graph Management, Reasoning over Data, Machine Learning and AI in Databases, Novel Database Architectures, Semi-structured Data Management, and User Interfaces for Data. Novel in this year's edition is the solicitation of contributions in the area of Applied Database Systems for Data Science. The goal is to give the opportunity to researchers from different areas dealing with interesting data management challenges in the context of Data Science, to disseminate them to the data management community, and at the same time to give database researchers working on interesting data science scenarios to publish their works.

The main Research track, the Industrial & Application track, as well as the Demonstration track have remained as in previous years, while the Short paper track has been slightly extended to accommodate papers of 6 pages, to allow more technical content.

With some rare exceptions, research papers were reviewed by 4 reviewers. Discussions and decisions were taken under the coordination of a senior PC member, with expertise in the respective area of the paper under review. The reviewing phase involved two cycles: resulting in a number of papers being revised and having their quality improved significantly.

Consistent with the tradition, the program included two keynotes, one on Data Profiling by Felix Naumann (Hasso Plattner Institute, Germany), and one on Knowledge Management by Katja Hose (Aalborg University, Denmark). These keynotes were complemented by the two additional keynotes of the co-organized ICDT. Last but not least, the program had the traditional tutorial track, featuring four tutorials on topics related to knowledge graphs, blockchains, text-to-SQL, and time series management. The overall program was accompanied by 6 Workshops.

The Program committee reviewed 108 full research papers, of which 27 were accepted. For short papers, 34 papers out of 113 were selected. The Industry and Application track received 22 submissions of which it selected 11 for publication, while the demo track received 24 works of which 15 were accepted.

Given that this year the conference took place online, it was a good opportunity to exploit at the maximum the opportunities that digital technologies can offer. Furthermore, it was our intention to make sure that the paper contributions are disseminated to a broader audience, especially outside the conference participants. For this reason, the authors were asked to provide a pre-recorded presentation of 10 min that will remain in the proceedings, a 30 sec pitch video that advertises the results of the contribution and a graphics ad.

In order to recognize significant contributions and give credits to the authors, the technical program committee awarded the Best Paper award to one of the research papers and the Best Demonstration award to one of the demos. Furthermore, following the EDBT tradition, it awarded the Test-of-Time award to a paper from the EDBT 2011 proceedings that has been deemed to have the greatest impact among those of that year. A novelty in this year's EDBT edition is the additional recognition of the Best Short Paper.

The realization of EDBT 2021 is a result of a collaborative effort of the different chairs and program committee members. Congratulations are in place to the senior PC members for guiding the discussions in such a professional and timely manner, ensuring the selection of the best quality papers. Special thanks for the excellent collaboration and quality of work go to the Industrial and Application Chair Eric Simon, the Demonstration Chair Sihem Amer-Yahia, the tutorial chairs Stefan Manegold and Wang-Chiew Tan, and the Workshop Chair Evaggelia Pitoura. A great deal of credits go to the Proceedings Chair Francesco Guerra for all the extra work he has put in organizing the material and making sure that all the proceedings information is in place. My appreciation goes also to the members of the paper award committees for the effort they put in evaluating the candidate papers and providing the decisions in a timely manner. I find it impossible to describe the passion, professionalism, consistency and collaborative attitude the general chairs Demetris Zeinalipour and Panos K. Chrysanthis have demonstrated. It was great working with them. I would also like to thank all the program committee members since due to them the high quality program became possible, the award committees, as well as Angela Bonifati and Marc H. Scholl from the EDBT Board for their numerous advices and support. Last but not least, I would like to thank all the authors for the works they submitted to the conference, the keynote speakers, the tutorial presenters, and the demonstration presenters.

Yannis Velegrakis, EDBT 2021 PC Chair

Message from the General Chairs

The 24th edition of the International Conference on Extending Database Technology (EDBT) was held between the 23rd and 26th of March 2021, despite the COVID-19 worldwide pandemic. Continuing its longstanding tradition as a premier data management research forum taking place in a European location, EDBT 2021, jointly organized with the International Conference on Database Theory (ICDT), was held virtually in Nicosia, the capital of the beautiful island of Cyprus.

At the onset, our aim has been to offer as close to an in-person, face-to-face experience as possible, maximizing the dissemination of knowledge and sharing of research results among the EDBT/ICDT participants at no health risk to them, and together discover the Cypriot culture and hospitality. EDBT/ICDT 2021 was initially planned to become the first hybrid EDBT/ICDT conference, combining physical presentations with an extended online audience, although this plan later evolved into an online format due to the ongoing situation with COVID-19. To this end, we have been committed to offering the best possible online experience to attendees by capitalizing and expanding on the success of earlier conferences.

Particularly, we decided on a number of novelties compared to previous conferences. Firstly, given that an optimal online experience cannot be achieved with pre-recorded presentations, but only with live presentations and direct interactions between the speakers and the audience, we decided in coordination with the Program Chairs on the online synchronous delivery of all events, including all research presentations, tutorials and live demonstrations. The next challenge was to select the appropriate conference management platform. Our personal experience as participants in several online conferences during the past year, and our evaluation of a number of popular web socializing platforms, led us to the decision that a custom conference attendance experience can only be delivered by an in-house platform that relies on a reliable teleconferencing channel, an intuitive interface, and a low learning curve.

To this end, the Data Management Systems Laboratory at the University of Cyprus, under the leadership of Demetris Zeinalipour developed and hosted a novel web-based platform named VGATE (Virtual Gate) to online conferences. VGATE allowed the Organizing Committee to collaborate over a Web-based Sheet interface (like Google Sheets) to plan and manage the organization. VGATE provided numerous helpful building-blocks in the online organization, namely, zoom license management and user status integration, integration of proceedings and multimedia content, industry booths, live sessions, social rooms, and for the first time an online helpdesk supported by Easy-Conferences Ltd. Special thanks to Paschalis Mpeis (University of Cyprus, Cyprus), Soteris Constantinou (University of Cyprus, Cyprus) and Constantinos Costa (University of Pittsburgh, USA) for their support and code contributions to the project under an extremely tight schedule. Special thanks also to Zoom Video Communications, Inc., for facilitating and sponsoring EDBT/ICDT 2021.

Novel this year is the participation of EDBT/ICDT to the newly founded Diversity and Inclusion (D&I) initiative of the Data Management community. EDBT/ICDT (alongside SIGMOD, VLDB, SoCC, and ICDE) celebrates the diversity in our community and welcomes everyone regardless of age, sex, gender identity, race, ethnicity, socioeconomic background, country of origin, religion, sexual orientation, physical ability, education, work experience, etc. To introduce this initiative, Panos K. Chrysanthis, the EDBT/ICDT D&I Chair, together with Sihem Amer-Yahia (CNRS, University Grenoble Alpes, France), a D&I Core Member, organized a panel as part of the Reception on DAY1 of the conference.

With D&I and broad participation in mind, the conference program was structured in a way that is convenient to at least half of the world at any given time. Specifically, the program was split into morning sessions (CET), which were convenient for EU-ASIA participants, on DAY1 and DAY3, and the afternoon sessions (CET), which were convenient for EU-America participants on DAY2 and DAY4. The Workshops were scheduled on DAY1 in the morning or afternoon based on the geographical location of their presenters. Social events were split along the same lines, presenting the host country, Cyprus, through music and videos on culture, geography, food, leisure, and other enjoyable aspects of Cyprus. Virtual corridor and hallway discussions were supported by a number of unmoderated and dynamic social rooms on VGATE at all times.

This year we also introduced some additional novelties, namely: (i) We introduced a special ceremony during Reception on DAY2, titled “*Pandemic Greetings from the Pioneers and the Next Data Management Challenge*”, with greetings by Philip A. Bernstein (Microsoft Research, WA, USA), Laura M. Haas (University of Massachusetts – Amherst, MA, USA), Yannis Ioannidis (University of Athens, Greece) and Jeffrey D. Ullman (Stanford, CA, USA); (ii) we introduced the concept of a sponsored industry talk during Dinner on DAY3, with title “*Behind the Scenes of Snowflake’s new Search Optimization Service*”, by Ismail Oukid (Snowflake, Germany) and Stefan Richter (Snowflake, Germany); (iii) we introduced a dedicated “*Demos in Action: Meet the Authors!*” session during Dinner on DAY3, which allowed Demo presenters to individually showcase their demos to participants in their private presentation space through

VGATE. Finally, we also continued to support the Climate Change discussion with a session on DAY4, led by Antoine Amarilli (Télécom Paris, France), with guest Benjamin Pierce (University of Pennsylvania, USA).

In all above endeavors, we had the support and encouragement of our incredible Organization Committee and the EDBT Executive Board, in particular, the President of EDBT Executive Board Angela Bonifati (Lyon 1 University, France), as well as the Chair of the ICDT Council Wim Martens (University of Bayreuth, Germany).

We are grateful to the entire Technical Program Committees, which, under the excellent leadership of the EDBT Program Chair Yannis Velegrakis (University of Trento, Italy and Utrecht University, Netherlands) and the ICDT Program Chair Ke Yi (Hong Kong University of Science and Technology, Hong Kong), brought forward an exciting technical program with a rich production of technical content (proceedings, videos, pitches, ads, etc.). Our gratitude also extends personally to the EDBT Demonstration Chair Sihem Amer-Yahia (CNRS, University Grenoble Alpes, France), the EDBT Workshop Chair Evaggelia Pitoura (University of Ioannina, Greece), the Climate Chair Antoine Amarilli (Télécom Paris, France), the EDBT Applied Database Systems for Data Science Vice-Chair Paul Groth (University of Amsterdam, Netherlands), the EDBT Industrial/Application Chair Eric Simon (SAP, France), the Tutorial Chairs Stefan Manegold (CWI, Netherlands) and Wang-Chiew Tan (Megagon Labs, USA), for the excellent collaboration, and exchange of ideas and insightful discussions. We would also like to highlight EDBT Demonstration Chair Sihem Amer-Yahia's success in putting together EDBT's first all-women Demonstration Track Committee. We would also like to thank the organizers of the six, co-located workshops DOLAP, BigVis, BMDA, DARLI-AP, SIMPLIFY and PIE+Q for enriching the scope of the technical program.

Several people contributed to the successful organization of the EDBT/ICDT 2021 conference. Special thanks to the following valued collaborators for their passion in organizing a memorable event: the Sponsorship Chair Divyakant Agrawal (University of California-Santa Barbara, USA), the Publicity Chair Herodotos Herodotou (Cyprus University of Technology, Cyprus), the Finance Chair George Pallis (University of Cyprus, Cyprus), the EDBT Proceedings Chair Francesco Guerra (University of Modena and Reggio Emilia, Italy), the ICDT Proceedings Chair Zhewei Wei (Renmin University, China), the Workshops Proceedings Chair Constantinos Costa (University of Pittsburgh, USA) and the Website support by Nicolas Kantzilaris (Easy Conferences, Cyprus).

Our sincere gratitude to our platinum sponsor, Snowflake, and our bronze sponsors, Oracle and Zoom, as well as the contact persons behind the support, namely Martin Hentschel (Snowflake), Ann Brisson (Oracle) and Alberto Colautti (Zoom).

Organizing EDBT 2021 at the University of Cyprus was Prof. George Samaras†' passion. After his unexpected passing two years ago, his friends and colleagues at the University of Cyprus volunteered to make his passion a reality. We are grateful to the EDBT Executive Board for giving us this opportunity and trusting us to organize EDBT 2021 alongside ICDT 2021 in Cyprus, as proposed by George in 2018. We dedicate the EDBT/ICDT 2021 in honor of the memory of Prof. George Samaras† (1959–2018).

We hope you enjoyed EDBT/ICDT 2021's exciting technical program and your virtual visit to Cyprus!

Demetris Zeinalipour, University of Cyprus, Cyprus

Panos K. Chrysanthis, University of Cyprus, Cyprus and University of Pittsburgh, USA

EDBT/ICDT 2021 General Co-Chairs

Program Committee Members

Research Program Committee Chair

Yannis Velegarakis, U Trento, Italy & Utrecht U, The Netherlands

Senior Program Committee Members

Periklis Andritsos, U Toronto, Canada
Nikos Bikakis, Athena Res. Ctr., Greece
Elena Ferrari, U Insubria, Italy
Avigdor Gal, Technion – Israel IT, Israel
Lukasz Golaz, U Waterloo, Canada
Paul Groth, U Amsterdam, Netherlands
Sergio Greco, U Calabria, Italy

Christian S. Jensen, Aalborg U, Denmark
Laks V.S. Lakshmanan, U British Columbia, Canada
Qiong Luo, HKUST, China
Raymond Ng, U British Columbia, Canada
Tamer Özsu, U Waterloo, Canada
Pinar Tozun, IT U Copenhagen, Denmark

Program Committee Members

Karl Aberer, EPFL, Switzerland
Ashraf Aboulnaga, Qatar Comp. Res. Inst., HBKU, Qatar
Bernd Amann, Sorbonne U, France
Nicolas Ancaux, INRIA, France
Walid G. Aref, Purdue U, USA
Akhil Arora, EPFL, Switzerland
Manos Athanassoulis, Boston U, USA
Nikolaus Augsten, U Salzburg, Austria
Elena Baralis, Politecnico di Torino, Italy
Denilson Barbosa, U Alberta, Canada
Ilaria Bartolini, U Bologna, Italy
Senjuti Basu Roy, New Jersey IT, USA
Luigi Bellomarini, Banca d'Italia, Italy
Michael Benedikt, U Oxford, UK
Alexander Boehm, SAP SE, Germany
Luc Bouganim, INRIA, France
Andrea Cali, Birkbeck U London, UK
Marco Calautti, U Trento, Italy
Bogdan Cautis, U Paris-Sud, France
Mel Chekol, Utrecht U, Netherlands
Vassilis Christophides, ENSEA, ETIS, France
Dario Colazzo, U Paris Dauphine – PSL, France
Bin Cui, Peking U, China
Alfredo Cuzzocrea, ICAR-CNR & U Calabria, Italy
Sabrina De Capitani di Vimercati, U Milano, Italy
Antonios Deligiannakis, TU Crete, Greece
Çağatay Demiralp, Sigma Computing, USA
Stefania Dumbrava, ENSIIE, France
George Fakas, Uppsala U, Sweden
Ju Fan, Renmin U China, China
Donatella Firmani, Roma Tre U, Italy
George Fletcher, Eindhoven UT, Netherlands
Daniele Foroni, Huawei, Germany
Johann-Christoph Freytag, HU Berlin, Germany
Avigdor Gal, Technion – Israel IT, Israel
Johann Gamper, Free U Bozen-Bolzano, Italy

Yunjun Gao, Zhejiang U, China
Rainer Gemulla, U Mannheim, Germany
Boris Glavic, Illinois IT, USA
Paolo Guagliardo, U Edinburgh, UK
Michael Gubanov, Florida State U, USA
Xi He, U Waterloo, Canada
Melanie Herschel, U Stuttgart, Germany
Jan Hidders, U London, UK
Katja Hose, Aalborg U, Denmark
Vagelis Hristidis, U California – Riverside, USA
Haoyu Huang, Google, USA
Xin Huang, Hong Kong Baptist U, Hong Kong SAR
Ekaterini Ioanou, Tilburg U, Netherlands
Zsolt István, ITU Copenhagen, Denmark
Panos Kalnis, King Abdullah UST, Saudi Arabia
Vana Kalogeraki, Athens U Eco. & Busin., Greece
Verena Kantere, National TU Athens, Greece
Panagiotis Karras, Aarhus U, Denmark
Asterios Katsifodimos, TU Delft, Netherlands
Anastasios Kementsietsidis, Google Research, USA
Haridimos Kondylakis, FORTH-ICS, Greece
Nick Koudas, U Toronto, Canada
Georgia Koutrika, Athena Research Center, Greece
Alexandros Labrinidis, U Pittsburgh, USA
Ulf Leser, HU Berlin, Germany
Guoliang Li, Tsinghua U, China
Han Li, Amazon, USA
Xiang Lian, Kent State U, USA
Chunbin Lin, Amazon AWS, USA
Matteo Lissandrini, Aalborg U, Denmark
Eric Lo, Chinese U Hong Kong, Hong Kong SAR
Ping Lu, Beihang U, China
Nikos Mamoulis, U Ioannina, Greece
Ioana Manolescu, INRIA & Inst. Poly. Paris, France
Sebastian Michel, TU Kaiserslautern, Germany
Paolo Missier, Newcastle U, UK

Mohamed Mokbel, U Minnesota – Twin Cities, USA
Mirella M. Moro, U Federal Minas Gerais, Brazil
Davide Mottin, Aarhus U, Denmark
Hieu Nguyen, eBay, USA
Behrooz Omidvar-Tehrani, LIG, France
Mourad Ouzzani, Qatar Comp. Res. Inst., HBKU, Qatar
George Papadakis, U Athens, Greece
Olga Papaemmanouil, Brandeis U, USA
Odysseas Papapetrou, TU Eindhoven, Netherlands
Paolo Papotti, Eurecom, France
Alok Pareek, Striim, USA
Torben Bach Pedersen, Aalborg U, Denmark
Eric Peukert, Leipzig U, Germany
Dimitris Plexousakis, ICS-FORTH, Greece
Laura Po, U Modena & Reggio Emilia, Italy
Arnau Prat, Sparsity Technologies, Spain
 Nicoleta Preda, U Versailles, France
Abdulhakim Qahtan, Utrecht U, Netherlands
Louiqa Raschid, U Maryland, USA
Mohammad Sadoghi, U California, Davis, USA
Carlo Sartiani, U Basilicata, Italy
Kai-Uwe Sattler, TU Ilmenau, Germany
Sebastian Schelter, U Amsterdam, Netherlands

Industrial Track Program Committee

Tyler Akidau, Snowflake, USA
Minhea Andrei, SAP, France
Angela Bonifati, U Lille, France
Jianjun Chen, Bytedance US Lab, USA
Thomas Fanghaenel, Salesforce, USA
Avrilia Floratou, Microsoft, USA
Prasanta Ghosh, Microsoft, USA
Yash Govind, Informatica LLC, USA
Laura Haas, U Mass. Amherst, USA
Zack Ives, U Pennsylvania, USA
Martin Kersten, MonetDB Solutions, Netherlands
Hariharan Lakshmanan, Oracle, USA
Dustin Lange, Amazon Research, USA
Jyoti Leeka, Microsoft, USA
Stefan Mandl, EXASOL AG, Germany
Anisoara Nica, SAP Labs Waterloo, Canada
Berthold Reinwald, IBM Research-Almaden, USA
Alejandro Salinger, SAP SE, Germany
Dennis Shasha, New York U, USA
Mohamed Soliman, Datometry, USA
Nesime Tatbul, Intel Labs and MIT, USA
Wei Wang, Nat. U Singapore, Singapore
Xuezhi Wang, Google, USA
Yongsik Yoon, Snowflake, USA
Kai Zeng, Alibaba Group, China

Steffi Scherzinger, U Passau, Germany
Petra Selmer, Neo4j, UK
Juan Sequeda, data.world, USA
Lidan Shou, Zhejiang U, China
Giovanni Simonini, U Modena & Reggio Emilia, Italy
Hala Skaf-Molli, U Nantes, France
Kostas Stefanidis, Tampere U, Finland
Gabor Szarnyas, CWI, Netherlands
Letizia Tanca, Politecnico di Milano, Italy
Ernest Teniente, U Politècnica Catalunya, Spain
Arash Termehchy, Oregon State U, USA
Chao Tian, Alibaba, China
Riccardo Torlone, Roma Tre U, Italy
Farouk Toumani, Clermont Auvergne U, CNRS, France
Katerina Tzompanaki, CY Cergy Paris U, France
Vasilis Vassalos, Athens U Eco. & Busin., Greece
Panos Vassiliadis, U Ioannina, Greece
Yannis Velegrakis, Utrecht U, Netherlands
Jianguo Wang, Purdue U, USA
Wendy Hui Wang, Stevens IT, USA
Wolfram Wingerath, Baqend, Germany
Nikolay Yakovets, TU Eindhoven, Netherlands
Meihui Zhang, Beijing IT, China

Demonstration Track Program Committee

Anastasia Ailamaki, EPFL, Switzerland
Elena Baralis, Polytecnico di Torino, Italy
Senjuti Basu Roy, NJIT, USA
Angela Bonifati, Lyon U, France
Renata Borovica-Gajic, U Melbourne, Australia
Malu Castellanos, TERADATA, USA
Sarah Cohen Boulakia, U Paris Sud, France
Maria Luisa Damiani, U Milan, Italy
Anna Fariha, UMass Amherst, USA
Irina Fundulaki, FORTH, GRnet, Greece
Katja Hose, Aalborg U, Denmark
Vana Kalogeraki, Athens U Eco. & Busin., Greece
Zoi Kaoudi, TU Berlin, Germany
Georgia Koutrika, ATHENA, Greece
Ioana Manolescu, INRIA, France
Renée J. Miller, Northeastern U, USA
Silvia Nittel, U Maine, USA
Fatma Özcan, Google, USA
Nesime Tatbul, Intel Labs & MIT, USA
Pinar Tözün, ITU, Denmark
Esther Pacitti, U Montpellier (INRIA & CNRS), France
Danica Porobic, Oracle, Switzerland
Agma Traina, ICMC-USP, Brazil
Zografoula Vagena, U Paris, France and RelationalAI, USA
Xiaolan Wang, Megagon Labs, USA
Karine Zeitouni, U Versailles, France

Additional Reviewers

Yaniv Gur, IBM Research-Almaden, USA
Ahmed Al-Baghdadi, Kent State U, USA
Niranjan Rai, Kent State U, USA
Weilong Ren, Kent State U, USA
Tahereh Arabghalizi, U Pittsburgh, USA
Evangelos Karageorgos, U Pittsburgh, USA
Xiaoting Li, U Pittsburgh, USA
Anthony Sicilia, U Pittsburgh, USA
Prithu Banerjee, U BC, Vancouver, Canada
Saket Gurukar, Ohio State U, Ohio, USA
Garima Gaur, IIT Kanpur, India
Sainyam Galhotra, U Mass., Amherst, USA

Nikos Giatrakos, TU Crete, Greece
Angjela Davitkova, TU Kaiserslautern, Germany
Leonardo Gazzarri, U Stuttgart, Germany
Flavio Giobergia, Politecnico di Torino, Italy
Eliana Pastor, Politecnico di Torino, Italy
Uta Störl, Darmstadt U of Applied Sciences, Germany
Wolfgang Mauerer, TU of Appl. Sc. Regensburg, Germany
Vasilis Efthymiou FORTH ICS Greece
Nikos Myrtakis U Crete, Greece
Benjamin Wollmer, Baqend and U Hamburg, Germany
Andrea Rossi, U Roma Tre, Italy
Tobias Lindaaker, Neo4j, Sweden

Conference Organization

General Chairs

Demetris Zeinalipour, University of Cyprus, Cyprus

Panos K. Chrysanthis, University of Cyprus, Cyprus and University of Pittsburgh, USA

EDBT Program Chair

Yannis Velegarakis, University of Trento, Italy and Utrecht University, Netherlands

ICDT Program Chair

Ke Yi, Hong Kong University of Science and Technology, Hong Kong

EDBT Industrial/Application Chair

Eric Simon, SAP, France

EDBT Applied Database Systems for Data Science Vice-Chair

Paul Groth, University of Amsterdam, Netherlands

EDBT Demonstrations Chair

Sihem Amer-Yahia, CNRS, Univ. Grenoble Alpes, France

Tutorial Chairs

Stefan Manegold, CWI, Netherlands

Wang-Chiew Tan, Megagon Labs, USA

Workshops Chair

Evaggelia Pitoura, University of Ioannina, Greece

Climate Chair

Antoine Amarilli, Télécom Paris, France

EDBT Proceedings Chair

Francesco Guerra, University of Modena and Reggio Emilia, Italy

ICDT Proceedings Chair

Zhewei Wei, Renmin University, China

Workshops Proceedings Chair

Constantinos Costa, University of Pittsburgh, USA

Diversity and Inclusion Chair

Panos K. Chrysanthis, University of Pittsburgh, USA

Sponsorship Chair

Divyakant Agrawal, UC Santa Barbara, USA

Publicity Chair

Herodotos Herodotou, Cyprus University of Technology, Cyprus

Finance Chair

George Pallis, University of Cyprus, Cyprus

Website Chair

Kyriakos Georgiades, EasyConferences, Cyprus

Test-of-Time Award

Established in 2014, the Test-of-Time Award of the Extended Database Technology (EDBT) Conference recognizes papers presented at the EDBT Conferences that have had the most impact in terms of research, methodology, conceptual contribution, or transfer to practice over the past ten years. The 2021 Test-of-Time Award committee looked and evaluated the impact of the papers in the EDBT 2011 proceedings, and selected

SeMiTri: a framework for semantic annotation of heterogeneous trajectories

by Zhixian Yan, Dipanjan Chakraborty, Christine Parent, Stefano Spaccapietra, and Karl Aberer
published in the EDBT 2011 Proceedings, pp. 259–270, DOI: 10.1145/1951365.1951398,

because it is one of the earliest papers to propose a general method for enriching moving object trajectories with semantics useful for supporting location-based services, which have been and still are in high demand across several application sectors. Since its publication, SeMiTri has generated significant interest, and follow-up work on semantic processing of mobile data and trajectories.

The EDBT 2021 Test of Time Award committee was formed by Barbara Catania, University of Genova, Italy, Gautam Das, University of Texas at Arlington, USA, Beng Chin OOI, National University of Singapore, Singapore, Themis Palpanas, University of Paris, France, and Yufei Tao, Chinese University of Hong Kong, China.

The EDBT Test-of-Time award for 2021 will be presented during the EDBT/ICDT 2021 Conference in Nicosia, Cyprus, as part of the Awards session on March 24, 2021.

Best Paper Award

The Best Paper Award Committee has looked at the papers accepted in the conference and selected one that was distinguishing itself in terms of research quality, presentation, technical challenges, and novelty. The selected paper is

DomainNet: Homograph Detection for Data Lake Disambiguation

by Aristotelis Leventidis, Laura Di Rocco, Wolfgang Gatterbauer,
Renée J. Miller and Mirek Riedewald.

DOI: 10.5441/002/edbt.2021.03

The paper presents DomainNet, a system that disambiguates values from heterogeneous datasets by creating a network representing co-occurring values and computing their graph centrality. The system is unsupervised, its accuracy outperforms the state-of-the-art, and it is accompanied by an open benchmark. The paper is of high significance: the problem is important, the proposed solution is effective, and the benchmark facilitates further research.

Abstract: Modern data lakes are deeply heterogeneous in the vocabulary that is used to describe data. We study a problem of disambiguation in data lakes: how can we determine if a data value occurring more than once in the lake has different meanings and is therefore a homograph? While word and entity disambiguation have been well studied in computational linguistics, data management and data science, we show that data lakes provide a new opportunity for disambiguation of data values since they represent a massive network of interconnected values. We investigate to what extent this network can be used to disambiguate values. DomainNet uses network-centrality measures on a bipartite graph whose nodes represent values and attributes to determine, without supervision, if a value is a homograph. A thorough experimental evaluation demonstrates that state-of-the-art techniques in domain discovery cannot be re-purposed to compete with our method. Specifically, using a domain discovery method to identify homographs has a precision and a recall of 38% versus 69% with our method on a synthetic benchmark. By applying a network-centrality measure to our graph representation, DomainNet achieves a good separation between homographs and data values with a unique meaning. On a real data lake our top- 200 precision is 89%.

The EDBT 2021 Best Paper Award committee was formed by Avigdor Gal, Technion Israel Institute of Technology, Israel, Lucasz Golab, University of Waterloo, Canada, Christian Jensen, Aalborg University, Denmark, and Qiong Luo, HKUST, China.

The EDBT Best Paper Award for 2021 will be presented during the EDBT/ICDT 2021 Conference in Nicosia, Cyprus, on March 24, 2021.

Best Short Paper Award

The Best Short Paper Award Committee has looked at the papers accepted in the conference and selected one that was distinguishing itself in terms of research quality, presentation, technical challenges, novelty, and potential impact to the broader research community and industry. The selected paper is

Answer Graph: Factorization Matters in Large Graphs

by Zahid Abul-Basher, Nikolay Yakovets, Parke Godfrey, Stanley Clark, and Mark Chignell.

DOI: 10.5441/002/edbt.2021.56

The paper proposes a two-step method for answering SPARQL conjunctive queries. The first step constructs an answer graph consisting only of node-edge-node triples that are answers to the query (a factorized answer set). In the second step, this answer graph is used to compute the actual embeddings for the query. The authors present detailed implementations for generating answer graphs and computing the final embedding, particularly in the case of cyclic conjunctive queries. Some first results from an experimental evaluation are also provided.

The EDBT 2021 Best Short Paper Award committee was formed by Manos Athanasoulis, Boston University, USA, Johann Gamper, Free University of Bozen-Bolzano, Italy, Ioana Manolescu, INRIA, France, Letizia Tanca, University of Milan, Italy, and Yannis Kotidis, Athens University of Economics and Business, Greece

The EDBT Best Short Paper Award for 2021 will be presented during the EDBT/ICDT 2021 Conference in Nicosia, Cyprus.

Best Demonstration Award

The Best Demonstration Award Committee has reviewed the video recordings of the 15 demos accepted at EDBT/ICDT and selected the most engaging demonstration that serves as an example of a structured and well illustrated demonstration and showcases an end-to-end system on a state-of-the-art topic that promotes data management beyond its boundaries. The selected demo is

Conversational OLAP in Action

by Matteo Francia, Enrico Gallinucci, and Matteo Golfarelli
(DISI – University of Bologna)
DOI: 10.5441/002/edbt.2021.74

For demonstrating COOL, a tool supporting natural language COntersational OLap sessions. COOL interprets and translates a natural language dialogue into an OLAP session that starts with a GPSJ query. The demonstration is engaging and showcases the usability of COOL and its capabilities in assisting query formulation and ambiguity resolution.

The EDBT 2021 Best Demonstration Award committee was formed by Sihem Amer-Yahia, CNRS Univ. Grenoble Alpes, France, Elena Baralis, Politecnico di Torino, Italy, Maria Luisa Damiani, University of Milan, Italy, Anna Fariha, UMass Amherst, USA, Irini Fundulaki, FORTH, GRnet, Greece, Zoi Kaoudi, TU Berlin, Germany, Georgia Koutrika, ATHENA, Greece, Esther Pacitti, University of Montpellier (Inria&CNRS), France, and Agma Traina, ICMC-USP, Brazil.

The EDBT Best Demonstration Award for 2021 will be presented during the EDBT/ICDT 2021 Conference in Nicosia, Cyprus, on March 24, 2021.

Table of Contents

Foreword by the PC Chair	i
Message from the General Chairs	ii
Program Committee Members	iv
Conference Organization	vii
Test-of-Time Award	viii
Best Paper Award	ix
Best Short Paper Award	x
Best Demonstration Award	xi
Table of Contents	xii
Research Papers	
Exchanging Data under Policy Views <i>Angela Bonifati, Ugo Comignani, Efthymia Tsamoura</i>	1
DomainNet: Homograph Detection for Data Lake Disambiguation <i>Aristotelis Leventidis, Laura Di Rocco, Wolfgang Gatterbauer, Renée J. Miller, Mirek Riedewald</i>	13
GPU-INSCY: A GPU-Parallel Algorithm and Tree Structure for Efficient Density-based Subspace Clustering <i>Jakob Rødsgaard Jørgensen, Katrine Scheel, Ira Assent</i>	25
JIT happens: Transactional Graph Processing in Persistent Memory meets Just-In-Time Compilation <i>Muhammad Attahir Jibril, Alexander Baumstark, Philipp Götze, Kai-Uwe Sattler</i>	37
Fixing Wikipedia Interlinks Using Revision History Patterns <i>Tova Milo, Slava Novgorodov, Kathy Razmadze</i>	49
Automating Data Quality Validation for Dynamic Data Ingestion <i>Sergey Redyuk, Zoi Kaoudi, Volker Markl, Sebastian Schelter</i>	61
Provenance-Based Algorithms for Rich Queries over Graph Databases <i>Yann Ramusat, Silviu Maniu, Pierre Senellart</i>	73
Sequence detection in event log files <i>Ioannis Mavroudpoulos, Theodoros Toliopoulos, Christos Bellas, Andreas Kosmatopoulos, Anastastios Gounaris</i>	85
A Comparative Evaluation of Anomaly Explanation Algorithms <i>Nikolaos Myrtakis, Vassilis Christophides, Eric Simon</i>	97
Scaling Density-Based Clustering to Large Collections of Sets <i>Daniel Kocher, Nikolaus Augsten, Willi Mann</i>	109
Assess Queries for Interactive Analysis of Data Cubes <i>Matteo Francia, Matteo Golfarelli, Patrick Marcel, Stefano Rizzi, Panos Vassiliadis</i>	121
SolveDB+: SQL-Based Prescriptive Analytics <i>Laurynas Siksnys, Torben Bach Pedersen, Thomas Dyhre Nielsen, Davide Frazzetto</i>	133
Multi-Objective Influence Maximization <i>Shay Gershtein, Tova Milo, Brit Youngmann</i>	145
Subjectivity Aware Conversational Search Services <i>Yacine Gaci, Jorge Ramirez, Boualem Benatallah, Fabio Casati, Khalid Benabdeslem</i>	157

GeoBlocks: A Query-Cache Accelerated Data Structure for Spatial Aggregation over Polygons <i>Christian Winter, Andreas Kipf, Christoph Anneser, Eleni Tzirita Zacharitou, Thomas Neumann, Alfons Kemper</i>	169
Indoor Spatial Queries: Modeling, Indexing, and Processing <i>Tiantian Liu, Huan Li, Hua Lu, Muhammad Aamir Cheema, Lidan Shou</i>	181
Structure Detection in Verbose CSV Files <i>Lan Jiang, Gerardo Vitagliano, Felix Naumann</i>	193
FRESQUE: A Scalable Ingestion Framework for Secure Range Query Processing on Clouds <i>Hoang Tran Van, Tristan Allard, Laurent d’Orazio, Amr El Abbadi</i>	205
Cache on Track (CoT): Decentralized Elastic Caches for Cloud Environments <i>Victor Zakhary, Lawrence Lim, Divy Agrawal, Amr El Abbadi</i>	217
Knowledge Graph Management on the Edge <i>Weiqin Xu, Olivier Curé, Philippe Calvez</i>	229
PolyFit: Polynomial-based Indexing Approach for Fast Approximate Range Aggregate Queries <i>Zhe Li, Tsz Nam Chan, Man Lung Yiu, Christian Jensen</i>	241
Shift-Table: A Low-latency Learned Index for Range Queries using Model Correction <i>Ali Hadian, Thomas Heinis</i>	253
An Efficient and Secure Location-based Alert Protocol using Searchable Encryption and Huffman Codes <i>Sina Shaham, Gabriel Ghinita, Cyrus Shahabi</i>	265
Concealer: SGX-based Secure, Volume Hiding, and Verifiable Processing of Spatial Time-Series Datasets <i>Peeyush Gupta, Sharad Mehrotra, Shantanu Sharma, Nalini Venkatasubramanian, Guoxi Wang</i>	277
Evaluation of Hardening Techniques for Privacy-Preserving Record Linkage <i>Martin Franke, Ziad Sehili, Florens Rohde, Erhard Rahm</i>	289
Proof-of-Execution: Reaching Consensus through Fault-Tolerant Speculation <i>Suyash Gupta, Jelle Hellings, Sajjad Rahnama, Mohammad Sadoghi</i>	301
Scalable Linear Algebra Programming for Big Data Analysis <i>Leonidas Fegaras</i>	313
Short Papers	
Automated Machine Learning for Entity Matching Tasks <i>Matteo Paganelli, Francesco Del Buono, Marco Pevarello, Francesco Guerra, Maurizio Vincini</i>	325
COCOA: COrrelation COefficient-Aware Data Augmentation <i>Mahdi Esmailoghli, Jorge Arnulfo Quiane Ruiz, Ziawasch Abedjan</i>	331
Efficient Exploratory Clustering Analyses with Qualitative Approximations <i>Manuel Fritz, Dennis Tschechlov, Holger Schwarz</i>	337
AutoML4Clust: Efficient AutoML for Clustering Analyses <i>Dennis Tschechlov, Manuel Fritz, Holger Schwarz</i>	343
Feature-driven Time Series Clustering <i>Donato Tiano, Angela Bonifati, Raymond Ng</i>	349
Indexed Log File: Towards Main Memory Database Instant Recovery <i>Arlino Magalhães, Angelo Brayner, José Maria Monteiro, Gustavo Moraes</i>	355
HorsePower: Accelerating Database Queries for Advanced Data Analytics <i>Hanfeng Chen, Joseph D’silva, Laurie Hendren, Bettina Kemme</i>	361

Robust and Memory-Efficient Database Fragment Allocation for Large and Uncertain Database Workloads <i>Rainer Schlosser, Stefan Halfpap</i>	367
TD-AC: Efficient Data Partitioning based Truth Discovery <i>Mouhamadou Lamine BA, Osias Noël Nicodème Finagnon Tossou</i>	373
Towards Automated Concept-based Decision Tree Explanations for CNNs <i>Radwa El Shawi, Youssef Sherif, Sherif Sakr</i>	379
Efficient Maintenance of Distance Labelling for Incremental Updates in Large Dynamic Graphs <i>Muhammad Farhan, Qing Wang</i>	385
KISS - A fast kNN-based Importance Score for Subspaces <i>Anna Beer, Ekaterina Allerborn, Valentin Hartmann, Thomas Seidl</i>	391
SceneRec: Scene-Based Graph Neural Networks for Recommender Systems <i>Gang Wang, Ziyi Guo, Xiang Li, Dawei Yin, Shuai Ma</i>	397
Efficient Contact Similarity Query over Uncertain Trajectories <i>Xichen Zhang, Suprio Ray, Farzaneh Shoeleh, Rongxing Lu</i>	403
AdCom: Adaptive Combiner for Streaming Aggregations <i>Felipe Gutierrez, Kaustubh Beedkar, Abel Souza, Volker Markl</i>	409
Querying Top-k Dominant Traffic Flows on Large Urban Road Networks <i>Stella Maropaki, Paolo Sottovia, Stefano Bortoli</i>	415
On Supporting Scalable Active Learning-based Interactive Data Exploration with Uncertainty Estimation Index <i>Xiaoyu Ge, Panos Chrysanthis</i>	421
Efficient Discovery of Approximate Order Dependencies <i>Reza Karegar, Parke Godfrey, Lukasz Golab, Mehdi Kargar, Divesh Srivastava, Jaroslav Szlichta</i>	427
Towards Scalable Data Discovery <i>Javier Flores, Sergi Nadal, Oscar Romero</i>	433
Adaptive Multi-Model Reinforcement Learning for Online Database Tuning <i>Yaniv Gur, Dongsheng Yang, Frederik Stalschus, Berthold Reinwald</i>	439
Optimising Fairness Through Parametrised Data Sampling <i>Vladimiro González-Zelaya, Julián Salas, Dennis Prangle, Paolo Missier</i>	445
Using Landmarks for Explaining Entity Matching Models <i>Andrea Baraldi, Francesco Del Buono, Matteo Paganelli, Francesco Guerra</i>	451
Human-Interpretable Rules for Anomaly Detection in Time-Series <i>Ines Ben Kraiem, Faiza Ghozzi, André Péninou, Geoffrey Roman-Jimenez, Olivier Teste</i>	457
DBMS Performance Troubleshooting in Cloud Computing Using Transaction Clustering <i>Arunprasad Marathe</i>	463
Revisiting Multidimensional Adaptive Indexing [Experiment & Analysis] <i>Anders Hammershøj Jensen, Frederik Lauridsen, Fatemeh Zardbani, Stratos Idreos, Panagiotis Karras</i>	469
Twin Subsequence Search in Time Series <i>Georgios Chatzigeorgakidis, Dimitrios Skoutas, Kostas Patroumpas, Themis Palpanas, Spiros Athanasiou, Spiros Skiadopoulos</i>	475
Progressive Mergesort: Merging Batches of Appends into Progressive Indexes <i>Pedro Holanda, Stefan Manegold</i>	481
Multiple-Source Context-Free Path Querying in Terms of Linear Algebra <i>Arseniy Terekhov, Vlada Pogozhelskaya, Vadim Abzalov, Timur Zinnatulin, Semyon Grigorev</i>	487

Answer Graph: Factorization Matters in Large Graphs <i>Zahid Abul-Basher, Nikolay Yakovets, Parke Godfrey, Stanley Clark, Mark Chignell</i>	493
Schema Inference for Property Graphs <i>Hanâ Lbath, Angela Bonifati, Russ Harmer</i>	499
Optimizing SPARQL Queries using Shape Statistics <i>Kashif Rabbani, Matteo Lissandrini, Katja Hose</i>	505
Preserving Diversity in Anonymized Data <i>Mostafa Milani, Yu Huang, Fei Chiang</i>	511
Automatic Tuning of Read-Time Tolerances for Optimized On-Demand Data-Streaming from Sensor Nodes <i>Julius Hülsmann, Chiao-Yun Li, Jonas Traub, Volker Markl</i>	517
SOJA: A Memory-efficient Small-large Outer Join for MPI <i>Liang Liang, Guang Yang, Thomas Heinis, David Taniar</i>	523
Industrial Papers	
JENGA - A Framework to Study the Impact of Data Errors on the Predictions of Machine Learning Models <i>Sebastian Schelter, Tammo Rukat, Felix Biessmann</i>	529
Decongestant: A Breath of Fresh Air for MongoDB Through Freshness-aware Reads <i>Chenhao Huang, Michael Cahill, Alan Fekete, Uwe Roehm</i>	535
DLC: A New Compaction Scheme for LSM-tree with High Stability and Low Latency <i>Peiquan Jin, Jianchuang Li, Hai Long</i>	547
Financial Data Exchange with Statistical Confidentiality: A Reasoning-based Approach <i>Luigi Bellomarini, Livia Blasi, Rosario Laurendi, Emanuel Sallinger</i>	558
Generating Realistic Test Datasets for Duplicate Detection at Scale Using Historical Voter Data <i>Fabian Panse, André Düjon, Wolfram Wingerath, Benjamin Wollmer</i>	570
Path Indexing in the Cypher Query Pipeline <i>Jochem Kuijpers, George Fletcher, Tobias Lindaaker, Nikolay Yakovets</i>	582
A Deep Learning Architecture for Audience Interest Prediction of News Topic on Social Media <i>Ciprian-Octavian Truică, Elena-Simona APOSTOL, Teodor Ștefu, Panos Karras</i>	588
AutoDBaaS: Autonomous Database as a Service for managing relational database services <i>Mayank Tiwary, Pritish Mishra, Shashank Mohan Jain, Kshira Sahoo</i>	600
Scalable Spatio-temporal Indexing and Querying over a Document-oriented NoSQL Store <i>Nikolaos Koutroumanis, Christos Doulkeridis</i>	611
Production Experiences from Computation Reuse at Microsoft <i>Alekh Jindal, Shi Qiao, Hiren Patel, Abhishek Roy, Jyoti Leeka, Brandon Haynes</i>	623
WILSON: A Divide and Conquer Approach for Fast and Effective News Timeline Summarization <i>Yiming Liao, Shuguang Wang, Dongwon Lee</i>	635
Demos	
Conversational OLAP in Action <i>Matteo Francia, Enrico Gallinucci, Matteo Golfarelli</i>	646
Smart City Data Analysis via Visualization of Correlated Attribute Patterns <i>Yuya Sasaki, Keizo Hori, Daiki Nishihara, Ohashi Sora, Yusuke Wakuta, Kei Harada, Makoto Onizuka, Yuki Arase, Shinji Shimojo, Kenji Doi, Hongdi He, Zhong-ren Peng</i>	650

SciNeM: A Scalable Data Science Tool for Heterogeneous Network Mining <i>Serafeim Chatzopoulos, Thanasis Vergoulis, Panagiotis Deligiannis, Dimitrios Skoutas, Theodore Dalamagas, Christos Tryfonopoulos</i>	654
IMCF: The IoT Meta-Control Firewall for Smart Buildings <i>Soteris Constantinou, Antonis Vasileiou, Andreas Konstantinidis, Panos Chrysanthis, Demetrios Zeinalipour-Yazti</i>	658
BBoxDB Streams: Distributed Processing of Real-World Streams of Position Data <i>Jan Kristof Nidzwetzki, Ralf Hartmut Güting</i>	662
Correlation graph analytics for stock time series data <i>Tong Liu, Paolo Coletti, Anton Dignös, Johann Gamper, Maurizio Murgia</i>	666
Conquering a Panda's weaker self - Fighting laziness with laziness <i>Stefan Hagedorn, Steffen Kläbe, Kai-Uwe Sattler</i>	670
DocDesign 2.0: Automated Database Design for Document Stores with Multi-criteria Optimization <i>Moditha Hewasinghage, Sergi Nadal, Alberto Abelló</i>	674
Visualizing and Exploring Big Datasets based on Semantic Community Detection <i>Maria Krommyda, Konstantinos Tsitseklis, Verena Kantere, Vasileios Karyotis, Symeon Papavassiliou</i>	678
Exploration and Analysis of Temporal Property Graphs <i>Christopher Rost, Kevin Gomez, Philip Fritzsche, Andreas Thor, Erhard Rahm</i>	682
Coronis: Towards Integrated and Open COVID-19 Data <i>Giorgos Santipantakis, George Vouros, Christos Doulkeridis</i>	686
Effective and Scalable Data Discovery with NextiaJD <i>Javier Flores, Sergi Nadal, Oscar Romero</i>	690
A Tool for JSON Schema Witness Generation <i>Lyes Attouche, Mohamed-Amine Baazizi, Dario Colazzo, Francesco Falleni, Giorgio Ghelli, Cristiano Landi, Carlo Sartiani, Stefanie Scherzinger</i>	694
covRew: a Python Toolkit for Pre-Processing Pipeline Rewriting Ensuring Coverage Constraint Satisfaction <i>Chiara Accinelli, Barbara Catania, Giovanna Guerrini, Simone Minisi</i>	698
EasyBDI: Near Real-Time Data Analytics over Heterogeneous Data Sources <i>Bruno Silva, Jose Moreira, Rogério Luís Costa</i>	702
Tutorials	
Tutorial on the Internals of Permissioned Blockchains and on How to Build Applications with Hyperledger Fabric <i>Zsolt István</i>	706
Deep Learning Approaches for Text-to-SQL Systems <i>George Katsogiannis-Meimarakis, Georgia Koutrika</i>	710
Big Sequence Management: Scaling up and Out <i>Karima Echihabi, Kostas Zoumpatianos, Themis Palpanas</i>	714