Report on SEBD 2020: the 28th Italian Symposium on Advanced Database Systems

Maristella Agosti
University of Padua, Italy
maristella.agosti@unipd.it

Maurizio Atzori
University of Cagliari, Italy
atzori@unica.it

Paolo Ciaccia
University of Bologna, Italy
paolo.ciaccia@unibo.it

Letizia Tanca
Politecnico di Milano, Italy
letizia.tanca@polimi.it

Abstract

This paper reports on the *28th Italian Symposium on Advanced Database Systems* (SEBD 2020), held online as a virtual conference from the 21st to the 24th of June 2020. The topics that were addressed in this edition of the conference were organized in the sessions: ontologies and data integration, anomaly detection and dependencies, text analysis and search, deep learning, noSQL data, trajectories and diffusion, health and medicine, context and ranking, social and knowledge graphs, multimedia content analysis, security issues, and data mining.

1 Overview

Since 1993 SEBD (*the Italian Symposium on Advanced Database Systems*) represents the main annual forum for the Italian database community, where to meet, discuss and exchange experiences among all those, both in the academy and industry, who are interested in database systems and in all their broad range of advanced systems and applications[1]. SEBD focuses on the specific themes of databases, but since its first editions it has always been a conference open to kindred research and industrial themes, including information retrieval, query answering, knowledge discovery, data mining, information access and filtering, text classification, and recommender systems.

SEBD 2020 conference was planned to be held in Villasimius (in the southern part of Sardinia), Italy, at the Tanka Resort, by the crystal-clear sea of Simius beach. Unfortunately the SARS-CoV-2 emergency required exceptional containment measures for reducing the spread of the virus, and the decision of running the conference online, on the same days, had to be taken.

Thanks to the experience from EDBT 2020 virtual conference, reported by its organizers [Bonifati et al., 2020], we were able to make informed decisions on how to run the online organization of

[1] dblp page of SEBD https://dblp.org/db/conf/sebd

SEBD 2020. After a selection over a dozen platforms, Zoom was chosen because of its familiarity among the community and the ability of having face-showing meetings with up to 49 people on the same screen[2]. Given the peculiar social dimension of SEBD events, we did not use one-way communications of Zoom Webinars, opting instead for the Zoom Meeting environment, also for the three keynotes, tutorial and panel. The welcome drink on the first evening was run on a different online platform called Gather[3], a fun game-like environment on a sea-side conference map background, where participants can communicate via audio-video calls by just moving their avatars close to each other.

Each day of the conference was divided into two Zoom meeting sessions: one in the morning and one in the afternoon. Even coffee breaks were included in the conference program, so that attendees had the chance to speak freely over Zoom. We believe that these social moments and live presentations, despite being held remotely, prevented us from the risk of perceiving it as a prerecorded videocast and contributed to the success of the event. Thanks to the valuable job of 49 program committee members, the scientific program consisted of 41 papers accepted for presentation: 8 research papers, 32 discussion papers, 1 student paper. With only a few exceptions, almost all submissions were reviewed by three anonymous reviewers from the PC members, with a median score for accepted papers of 1.7 out of 3.0 (according to the EasyChair metric). The scientific program of the conference was enriched by three invited talks, whose contents are illustrated in Section 2, and a special session on "Open Science", whose contents are presented in detail in Section 3.

The proceedings of the conference [Agosti et al., 2020] are available online as Volume 2646 of the CEUR series of Workshop Proceedings[4].

Last but not least, unfortunately in the period of preparation for the conference and during the congress we lost two colleagues that we want to remember here: Florian Daniel and Antonio Picariello. On April 28th, 2020, Florian Daniel died prematurely, at 42, after a brief illness. Florian was an associate professor at Politecnico di Milano; a brilliant, active and passionate scholar, a beloved and appreciated teacher, with his irony and acumen he had won the esteem and recognition of colleagues and students. Antonio Picariello (aka Picus) passed away on June 23rd, 2020, while the SEBD 2020 conference was taking place: a real shock for all conference participants, many of them being close friends or collaborators of Antonio. Antonio Picariello was a full professor at the University of Naples "Federico II", he was one of the most passionate sustainers of SEBD and he was well-known for his contributions in the area of database and multimedia systems, and, more recently, of big data analytics.

2 Invited Talks

The scientific program included three invited talks by the internationally renowned scientists Tova Milo, Gerhard Weikum and Carlo Zaniolo.

[2]Zoom Video Communications, Inc. Zoom Meeting online platform https://zoom.us/
[3]Gather Presence Inc. Gather Town online platform https://gather.town/
[4]SEBD 2020 proceedings: http://ceur-ws.org/Vol-2646/

Getting Rid of Data – *Tova Milo*, Tel Aviv University, Israel

Rationale: We are experiencing an amazing data-centered revolution. Incredible amounts of data are collected, integrated and analyzed, leading to key breakthroughs in science and society. This well of knowledge, however, is at a great risk if we do not dispense with some of the data flood. First, the amount of generated data grows exponentially and already at 2025 is expected to be more than five times the available storage. Second, even disregarding storage constraints, uncontrolled data retention risks privacy and security, as recognized, e.g., by the recent EU Data Protection reform. Data disposal policies must be developed to benefit and protect organizations and individuals. Retaining the knowledge hidden in the data while respecting storage, processing and regulatory constraints is a great challenge. The difficulty stems from the distinct, intricate requirements entailed by each type of constraint, the scale and velocity of data and the constantly evolving needs. While multiple data sketching, summarization and deletion techniques were developed to address specific aspects of the problem, we are still very far from a comprehensive solution. Every organization has to battle the same tough challenges, with ad hoc solutions that are application specific and rarely sharable.
Contents: The talk addressed the logical, algorithmic, and methodological foundations required for the systematic disposal of large-scale data, for constraints enforcement and for the development of applications over the retained information. An overview of relevant related work was presented, together with the presentation of new research challenges and potential reuse of existing techniques, as well as the research performed in this direction in the Tel Aviv Databases group.

15 Years of Knowledge Graphs: Lessons, Challenges, Opportunities – *Gerhard Weikum*, Max Planck Institute for Informatics, Germany

Rationale: Machines with comprehensive knowledge of the world's entities and their relationships is a long-standing vision and challenge of AI. Over the last 15 years, huge knowledge bases, also known as knowledge graphs, were automatically constructed from web data and text sources, and became a key asset for search, analytics, recommendations and data integration. This digital knowledge can be harnessed to semantically interpret textual phrases in news, social media and web tables, contributing to question answering, natural language processing and data analytics.
Contents: The talk reviewed these advances and discussed lessons learned. Moreover, it also identified open challenges and new research opportunities.

Usability, Performance and Scalability for Expressive Data Languages via Cardinality-Based Aggregates – *Carlo Zaniolo*, University of California, Los Angeles, USA

Rationale: In theory, the combined use of recursion and negation allows the declarative expression of very powerful algorithms in languages such as Datalog or SQL. However, the difficulty of writing such programs and proving that they satisfy formal non-monotonic semantics makes such an approach totally impractical for software developers.
Contents: The talk presented a new general solution to the problem of developing advanced applications in logic-based languages. The adopted approach is based on the combined use of

recursion and aggregates endowed with a unified definition of their semantics based on cardinality. In fact, it showed that this entails the expression on a wide range of algorithms used in ML, data-mining and graph applications, which can be very appealing for practitioners because of (i) the compactness of their code, (ii) a simplified proof or verification for their stable-model semantics, (iii) their efficient implementations via a max-optimized semi-naive fixpoint algorithm, and (iv) their superior scalability via Stale-Synchronous Parallelism. To demonstrate and further enhance the usability of the proposed framework, including the several efficient and scalable applications that were developed, the speaker provided a Logical Algorithm Library (Llib) and a Logical Data-Frame System (LFrame). By integrating access to Llib library with other Apache Spark libraries, and supporting the interoperability of BigDatalog, RaSQL, and Datalog-ML systems with Scala, Java and Python, LFrame turns Datalog into a powerful and attractive tool for advanced application development in the Spark ecosystem.

3 Special Session on Open Science

The session was organised into two moments: an introductory tutorial on Open Science and a discussion with a panel of experts. The slides that were presented during the tutorial and the panel can be downloaded from the Web site of the conference[5].

Tutorial

The speakers were *Paolo Manghi*, CNR-ISTI, CTO of the European OpenAIRE e-infrastructure, and *Emma Lazzeri*, CNR-ISTI, OpenAIRE National Open Access Desk.

The desired outcomes were to address the motivations and the main features of Open Science, linking it to the research integrity and reproducibility of science, with a focus on the challenges in the ICT and database systems domains. The following contents were dealt with in particular:

- How does science work today?
  - Scientific process and scientific communication
  - Scientific indicators
  - Market interests and business models in the current research evaluation system
- Open Science
  - Open Science: science set free
  - Open Access: advantages, cultural barriers, European mandates and obligations
  - Open Data: why and how to manage research data
  - FAIRness of research products
- Reproducibility of science and state of the art in Computer Science.

[5]SEBD 2020 tutorial and panel slides: https://sebd2020.unica.it/openscience

Panel

After the tutorial, a panel of experts explored some specific issues, in particular: *Paolo Manghi* dealt with the themes of OpenScience discovery and monitoring, and the OpenAIRE graph; *Shalini Kurapati*, of the Polytechnic of Turin, dealt with the topic if Data Stewardship because she is an expert in research data management and data management plan, having been a former data steward at TU-Delft and currently working on the same issues at the Polytechnic of Turin; *Stefano Bianco*, a researcher at the National Institute of Nuclear Physics, as a member of the Expert Group of Plan-S, presented the plan launched at the end of 2018 by Coalition-S, a group of research funding bodies that aim to obtain full Open Access at all publications produced with public funds since 2021; finally, *Emma Lazzeri*, which coordinates the National Open Access Desk of OpenAIRE in Italy, presented the main projects funded by the European Community on Open Access.

4 Best Paper-Title Award

A "Best Paper-Title Award" was introduced in this edition of SEBD for the first time because the authors *Paolo Ciaccia*, *Davide Martinenghi* and *Riccardo Torlone* managed to deal with a demanding scientific topic under the certainly playful and original title

> Where Porceddu is better than Pasta

which recalls two typical ingredients of Italian cuisine, which are pork, which is called "Porceddu" in the Sardinian language (and SEBD 2020 took place in Sardinia), and "Pasta", a generic name to identify any recipe based on pasta in any Italian region. Many participants appeared really amused during the delivery of this new and original prize.

Acknowledgements

We gratefully acknowledge CC ICT-SUD for its sponsorship and support in the organization. We wish to thank the authors who submitted papers and the conference participants. We are also grateful to the members of the program committee and the external referees for their thorough work, the invited speaker and panelists, and all the other people involved in the organization of this unusual edition of SEBD.

References

Maristella Agosti, Maurizio Atzori, Paolo Ciaccia, and Letizia Tanca, editors. *Proceedings of the 28th Italian Symposium on Advanced Database Systems, Villasimius, Sud Sardegna, Italy (virtual due to Covid-19 pandemic), June 21-24, 2020*, volume 2646 of *CEUR Workshop Proceedings*, 2020. CEUR-WS.org.

Angela Bonifati, Giovanna Guerrini, Carsten Lutz, Wim Martens, Lara Mazilu, Norman Paton, Marcos Antonio Vaz Salles, Marc H. Scholl, and Yongluan Zhou. Holding a Conference Online and Live, Due to COVID-19. *Communications of the ACM*, April 21, 2020.