

Preface of the Workshop Organizers

The Second International Workshop on “Challenges and Experiences from Data Integration to Knowledge Graphs” (DI2KG 2020) hosted different contributions on data integration problems from noisy and heterogeneous sources.

These problems have been studied by different research communities, and we claim the need for a more cross-disciplinary approach. For this reason, we proposed to participants a broad spectrum of topics, including every possible step of an ideal pipeline that starts from data extraction from the Web to the construction of a complete and correct knowledge graph.

We also hosted the DI2KG challenge: we provided a dataset of product specifications extracted from the Web, suggested a set of fundamental integration tasks with their own manually-checked ground truth, and published a weekly updated leaderboard to help researchers evaluate their solution against competitors.

Submitted papers underwent a zero-blind review process. The challenge finalists had the opportunity to publish a paper describing their solution, which went through a light review process. We also accepted papers from finalists of the SIGMOD 2020 Programming Contest¹, with a similar goal but focused also on efficiency of solutions.

The workshop was held online due to the ongoing pandemic, and about 40 people from different research communities and industries participated.² We had three invited speakers (William W. Cohen from Google, Renée Miller from Northeastern University, and Mike Tung from Diffbot), the presentation of 6 papers, and a student panel with 3 speakers (Bahar Ghadiri Bashardoost from University of Toronto, Riccardo Cappuzzo from EURECOM and Daniel Obraczka from ScaDS) moderated by Donatella Firmani.

All of the participants underlined the importance of benchmarks, and panelists as well as organizers were happy with the discussion and ideas that emerged from the discussion. Overall we are confident that new editions of DI2KG can stimulate new research on holistic data integration methods, and help building a community that can contribute to the evolution of the DI2KG challenge benchmark.

We would like to thank the program committee members for their work in the review process: Edward Porter, Domenico Beneventano, Christian Bizer, Yash Govind, Laure Berti-Equille, Daniel Obraczka, Martin Theobald, Juan F. Sequeda, Colin Lockard, Nicholas Monath, Denilson Barbosa, Mourad Ouzzani, and H. V. Jagadish. Finally, we thank the authors of submitted papers, the challenge participants, invited speakers, panelists and workshop attendees who contributed to the success of the event.

October 2020

Federico Piai, Donatella Firmani, Divesh Srivastava, Xin Luna Dong, Andrea de Angelis, Maurizio Mazzei, Paolo Merialdo.

¹<http://www.inf.uniroma3.it/db/sigmod2020contest/index.html>

²The detailed workshop program is available on-line at <http://di2kg.dia.uniroma3.it/2020>.