DaSH-LA

The 2nd Workshop on Data Science with Human-in-the-loop: Language Advances

Proceedings of the Workshop

June 11, 2021

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-954085-39-8

Message from the Workshop Co-chairs

The 2nd Workshop on Data Science with Human-in-the-loop (DaSH) builds on the success of the inaugural workshop that took place at KDD in 2020. The current workshop (DaSH-LA) is co-located with NAACL-HLT 2021, and its focus is on human-in-the-loop aspects in computational linguistics and natural language processing (NLP).

The aim of the DaSH-LA workshop is to stimulate research on human-computer interaction challenges in data science within the broad areas related to language, including but not limited to information extraction, text classification, machine translation, dialog systems, question answering, language generation, information retrieval, digital humanity, and more. We expect the overall series of the DaSH workshops to help develop and grow a strong community of researchers who are interested in this topic and to yield future collaborations and scientific exchanges across the relevant areas of computational linguistics, data mining, machine learning, data and knowledge management, human-machine interaction, and user interfaces.

The participants of the DaSH-LA workshop include researchers and practitioners interested in understanding how to optimize human-computer cooperation and how to minimize human effort along various NLP pipelines in a wide range of tasks and real-life applications. The full-day program includes two keynote talks (by Dan Weld and Joyce Chai), three regular sessions with 14 accepted papers, a special session with highlights from two recent papers with human-in-the-loop focus, as well as a panel of experts (including Danqi Chen, Joel Tetreault and the two keynote speakers).

We would like to thank all people who in one way or another helped with the workshop. We are thankful to the members of the program committee who did an excellent job in reviewing the submitted papers under strict time constraints, and also to the steering committee for their helpful suggestions. Last but not least we would like to thank all authors, speakers and participants at the workshop.

Eduard Dragut, Yunyao Li, Lucian Popa, and Slobodan Vucetic
June 2021

Workshop Co-chairs

Eduard Dragut, Temple University
Lucian Popa, IBM Research
Slobodan Vucetic, Temple University
Yunyao Li, IBM Research

Program Committee

Mohit Bansal, University of North Carolina (UNC) Chapel Hill
Cornelia Caragea, University of Illinois at Chicago
Marina Danilevsky, IBM Research
Aritra Dasgupta, New Jersey Institute of Technology
Nemanja Djuric, Aurora Innovation
Kenneth Forbus, Northwestern University
Anna Lisa Gentile, IBM Research
Iryna Gurevych, Technical University of Darmstadt
Lifu Huang, Virginia Tech
Dongyeop Kang, Carnegie Mellon University
Jonathan K. Kummerfeld, University of Michigan
Bing Liu, University of Illinois at Chicago
Jeff Pan, University of Edinburgh
Soujanya Poria, Singapore University of Technology and Design
Daniel Preotiuc-Pietro, Bloomberg
Kun Qian, Amazon
Xiang Ren, University of Southern California
Shashank Srivastava, University of North Carolina (UNC) Chapel Hill
Gabriel Stanovsky, Hebrew University of Jerusalem
Benjamin Van Durme, Johns Hopkins University
Dakuo Wang, IBM Research
Rui Zhang, Penn State University

Steering Committee

AnHai Doan, University of Wisconsin
ChengXiang Zhai, University of Illinois at Urbana-Champaign
Dan Weld, University of Washington
Marti A. Hearst, University of California, Berkeley
Sunita Sarawagi, IIT Bombay

Table of Contents

Conference Program

Keynote Talk 1: Daniel Weld

Regular Session 1: Support for Text Analytics with Human in the Loop

Leveraging Wikipedia Navigational Templates for Curating Domain-Specific Fuzzy Conceptual Bases
Krati Saxena, Tushita Singh, Ashwini Patil, Sagar Sunkle and Vinay Kulkarni

It is better to Verify: Semi-Supervised Learning with a human in the loop for large-scale NLU models
Verena Weber, Enrico Piovano and Melanie Bradford

ViziTex: Interactive Visual Sense-Making of Text Corpora
Natraj Raman, Sameena Shah, Tucker Balch and Manuela Veloso

A Visualization Approach for Rapid Labeling of Clinical Notes for Smoking Status Extraction
Saman Enayati, Ziyu Yang, Benjamin Lu and Slobodan Vucetic

Semi-supervised Interactive Intent Labeling
Saurav Sahay, Eda Okur, Nagib Hakim and Lama Nachman

Highlights: Human in the Loop Papers from Recent Conferences

Human-In-The-LoopEntity Linking for Low Resource Domains
Jan-Christoph Klie, Richard Eckart de Castilho and Iryna Gurevych

Bridging Multi-disciplinary Collaboration Challenges in ML Development via Domain Knowledge Elicitation
Soya Park

Regular Session 2: Human in the Loop for NLP Tasks

Keynote Talk 2: Joyce Chai

Regular Session 3: Human in the Loop Tools

Panel