

# **2nd Workshop on Data Science with Human-in-the-loop: Language Advances (DaSH-LA 2021)**

Online  
11 June 2021

ISBN: 978-1-7138-3021-4

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2021) by the Association for Computational Linguistics  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006

Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>Leveraging Wikipedia Navigational Templates for Curating Domain-Specific Fuzzy Conceptual Bases</i> Kрати Saxena, Tushita Singh, Ashwini Patil, Sagar Sunkle and Vinay Kulkarni .....	1
<i>It is better to Verify: Semi-Supervised Learning with a human in the loop for large-scale NLU models</i> Verena Weber, Enrico Piovano and Melanie Bradford .....	8
<i>ViziTex: Interactive Visual Sense-Making of Text Corpora</i> Natraj Raman, Sameena Shah, Tucker Balch and Manuela Veloso .....	16
<i>A Visualization Approach for Rapid Labeling of Clinical Notes for Smoking Status Extraction</i> Saman Enayati, Ziyu Yang, Benjamin Lu and Slobodan Vucetic .....	24
<i>Semi-supervised Interactive Intent Labeling</i> Saurav Sahay, Eda Okur, Nagib Hakim and Lama Nachman .....	31
<i>Human-In-The-Loop Entity Linking for Low Resource Domains</i> Jan-Christoph Klie, Richard Eckart de Castilho and Iryna Gurevych .....	41
<i>Bridging Multi-disciplinary Collaboration Challenges in ML Development via Domain Knowledge Elicitation</i> Soya Park .....	44
<i>Active learning and negative evidence for language identification</i> Thomas Lippincott and Ben Van Durme .....	47
<i>Towards integrated, interactive, and extensible text data analytics with Leam</i> Peter Griggs, Cagatay Demiralp and Sajjadur Rahman .....	52
<i>Data Cleaning Tools for Token Classification Tasks</i> Karthik Muthuraman, Frederick Reiss, Hong Xu, Bryan Cutler and Zachary Eichenberger .....	59
<i>Building Low-Resource NER Models Using Non-Speaker Annotations</i> Tatiana Tsygankova, Francesca Marini, Stephen Mayhew and Dan Roth .....	62
<i>Evaluating and Explaining Natural Language Generation with GenX</i> Kayla Duskin, Shivam Sharma, Ji Young Yun, Emily Saldanha and Dustin Arendt .....	70
<i>CrossCheck: Rapid, Reproducible, and Interpretable Model Evaluation</i> Dustin Arendt, Zhuanyi Shaw, Prasha Shrestha, Ellyn Ayton, Maria Glenski and Svitlana Volkova	79
<i>TopGuNN: Fast NLP Training Data Augmentation using Large Corpora</i> Rebecca Iglesias-Flores, Megha Mishra, Ajay Patel, Akanksha Malhotra, Reno Kriz, Martha Palmer and Chris Callison-Burch .....	86
<i>Everyday Living Artificial Intelligence Hub</i> Raymond Finzel, Esha Singh, Martin Michalowski, Maria Gini and Serguei Pakhomov .....	102
<i>A Computational Model for Interactive Transcription</i> William Lane, Mat Bettinson and Steven Bird .....	105