

Workshop on Data Science with Human-in-the-loop: Language Advances (DaSH 2022)

Abu Dhabi, United Arab Emirates
8 December 2022

ISBN: 978-1-7138-9107-9

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© (2022) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

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

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
Web: www.proceedings.com

Table of Contents

<i>MEGAnno: Exploratory Labeling for NLP in Computational Notebooks</i>	
Dan Zhang, Hannah Kim, Rafael Li Chen, Eser Kandogan and Estevam Hruschka	1
<i>Cross-lingual Short-text Entity Linking: Generating Features for Neuro-Symbolic Methods</i>	
Qiuhaio Lu, Sairam Gurajada, Prithviraj Sen, Lucian Popa, Dejing Dou and Thien Nguyen	8
<i>Crowdsourcing Preposition Sense Disambiguation with High Precision via a Priming Task</i>	
Shira Wein and Nathan Schneider	15
<i>DoSA : A System to Accelerate Annotations on Business Documents with Human-in-the-Loop</i>	
Neelesh Shukla, Msp Raja, Raghu Katikeri and Amit Vaid	23
<i>Execution-based Evaluation for Data Science Code Generation Models</i>	
Junjie Huang, Chenglong Wang, Jipeng Zhang, Cong Yan, Haotian Cui, Jeevana Priya Inala, Colin Clement and Nan Duan	28
<i>A Gamified Approach to Frame Semantic Role Labeling</i>	
Emily Amspoker and Miriam R L Petruck	37
<i>A Comparative Analysis between Human-in-the-loop Systems and Large Language Models for Pattern Extraction Tasks</i>	
Maeda Hanafi, Yannis Katsis, Ishan Jindal and Lucian Popa	43
<i>Guiding Generative Language Models for Data Augmentation in Few-Shot Text Classification</i>	
Aleksandra Edwards, Asahi Ushio, Jose Camacho-collados, Helene Ribaupierre and Alun Preece	51
<i>Partially Humanizing Weak Supervision: Towards a Better Low Resource Pipeline for Spoken Language Understanding</i>	
Ayush Kumar, Rishabh Tripathi and Jithendra Vepa	64
<i>Improving Human Annotation Effectiveness for Fact Collection by Identifying the Most Relevant Answers</i>	
Pranav Kamath, Yiwen Sun, Thomas Semere, Adam Green, Scott Manley, Xiaoguang Qi, Kun Qian and Yunyao Li	74
<i>AVA-TMP: A Human-in-the-Loop Multi-layer Dynamic Topic Modeling Pipeline</i>	
Viseth Sean, Padideh Danaee, Yang Yang and Hakan Kardes	81
<i>Improving Named Entity Recognition in Telephone Conversations via Effective Active Learning with Human in the Loop</i>	
Md Tahmid Rahman Laskar, Cheng Chen, Xue-yong Fu and Shashi Bhushan Tn	88
<i>Interactively Uncovering Latent Arguments in Social Media Platforms: A Case Study on the Covid-19 Vaccine Debate</i>	
Maria Leonor Pacheco, Tunazzina Islam, Lyle Ungar, Ming Yin and Dan Goldwasser	94
<i>User or Labor: An Interaction Framework for Human-Machine Relationships in NLP</i>	
Ruyuan Wan, Naome Etori, Karla Badillo-urquiola and Dongyeop Kang	112