

3rd Workshop on Deep Learning Approaches for Low-Resource Natural Language Processing (DeepLo 2022)

Online
14 July 2022

ISBN: 978-1-7138-5628-3

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. (2022)

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>Introducing QuBERT: A Large Monolingual Corpus and BERT Model for Southern Quechua</i> Rodolfo Zevallos, John Ortega, William Chen, Richard Castro, Núria Bel, Cesar Toshio, Renzo Venturas and Hilario Aradiel and Nelsi Melgarejo	1
<i>Improving Distantly Supervised Document-Level Relation Extraction Through Natural Language Inference</i> Clara Vania and Grace Lee and Andrea Pierleoni	14
<i>IDANI: Inference-time Domain Adaptation via Neuron-level Interventions</i> Omer Antverg and Eyal Ben-David and Yonatan Belinkov	21
<i>Generating unlabelled data for a tri-training approach in a low resourced NER task</i> Hugo Boulanger and Thomas Lavergne and Sophie Rosset	30
<i>ANTS: A Framework for Retrieval of Text Segments in Unstructured Documents</i> Brian Chivers, Mason P. Jiang, Wonhee Lee, Amy Ng and Natalya I. Rapstine and Alex Storer	38
<i>Cross-TOP: Zero-Shot Cross-Schema Task-Oriented Parsing</i> Melanie A. Rubino, Nicolas Guenon des mesnards, Uday Shah, Nanjiang Jiang and Weiqi Sun and Konstantine Arkoudas	48
<i>Help from the Neighbors: Estonian Dialect Normalization Using a Finnish Dialect Generator</i> Mika Härmäläinen and Khalid Alnajjar and Tuuli Tuisk	61
<i>Exploring diversity in back translation for low-resource machine translation</i> Laurie Burchell and Alexandra Birch and Kenneth Heafield	67
<i>Punctuation Restoration in Spanish Customer Support Transcripts using Transfer Learning</i> Xiliang Zhu, Shayna Gardiner, David Rossouw and Tere Roldán and Simon Corston-Oliver ..	80
<i>Pre-training Data Quality and Quantity for a Low-Resource Language: New Corpus and BERT Models for Maltese</i> Kurt Micallef, Albert Gatt, Marc Tanti and Lonneke van der Plas and Claudia Borg	90
<i>Building an Event Extractor with Only a Few Examples</i> Pengfei Yu, Zixuan Zhang, Clare Voss and Jonathan May and Heng Ji	102
<i>Task Transfer and Domain Adaptation for Zero-Shot Question Answering</i> Xiang Pan, Alex Sheng, David Shimshoni, Aditya Singhal and Sara Rosenthal and Avirup Sil	110
<i>Let the Model Decide its Curriculum for Multitask Learning</i> Neeraj Varshney and Swaroop Mishra and Chitta Baral	117
<i>AfriTeVA: Extending ?Small Data? Pretraining Approaches to Sequence-to-Sequence Models</i> Oduwayo Jude Ogundepo, Akintunde Oladipo, Mofetoluwa Adeyemi and Kelechi Ogueji and Jimmy Lin	126
<i>Few-shot Learning for Sumerian Named Entity Recognition</i> Guanghai Wang and Yudong Liu and James Hearne	136
<i>Deep Learning-Based Morphological Segmentation for Indigenous Languages: A Study Case on Innu-Aimun</i> Ngoc Tan Le, Antoine Cadotte, Mathieu Boivin and Fatiha Sadat and Jimena Terraza	146

<i>Clean or Annotate: How to Spend a Limited Data Collection Budget</i> Derek Chen and Zhou Yu and Samuel R. Bowman	152
<i>Unsupervised Knowledge Graph Generation Using Semantic Similarity Matching</i> Lixian Liu, Amin Omidvar, Zongyang Ma and Ameeta Agrawal and Aijun An	169
<i>FarFetched: Entity-centric Reasoning and Claim Validation for the Greek Language based on Textually Represented Environments</i> Dimitris Papadopoulos, Katerina Metropoulou and Nikolaos Papadakis and Nikolaos Matsatsinis	180
<i>Alternative non-BERT model choices for the textual classification in low-resource languages and environments</i> Syed Mustavi Maheen, Moshir Rahman Faisal and Md. Rafakat Rahman and Md. Shahriar Karim.....	192
<i>Generating Complement Data for Aspect Term Extraction with GPT-2</i> Amir Pouran Ben Veyseh, Franck Dernoncourt and Bonan Min and Thien Huu Nguyen	203
<i>How to Translate Your Samples and Choose Your Shots? Analyzing Translate-train & Few-shot Cross-lingual Transfer</i> Iman Jundi and Gabriella Lapesa	214
<i>Unified NMT models for the Indian subcontinent, transcending script-barriers</i> Gokul N.C.....	227