

8th Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial '2021)

Online
20 April 2021

ISBN: 978-1-7138-2971-3

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© (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

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>Findings of the VarDial Evaluation Campaign 2021</i> Bharathi Raja Chakravarthi, Gaman Mihaela, Radu Tudor Ionescu, Heidi Jauhiainen, Tommi Jauhiainen, Krister Lindén, Nikola Ljubešić, Niko Partanen, Ruba Priyadharshini, Christoph Purschke, Eswari Rajagopal, Yves Scherrer and Marcos Zampieri	1
<i>Hierarchical Transformer for Multilingual Machine Translation</i> Albina Khusainova, Adil Khan, Adín Ramírez Rivera and Vitaly Romanov	12
<i>Regression Analysis of Lexical and Morpho-Syntactic Properties of Kiezdeutsch</i> Diego Frassinelli, Gabriella Lapesa, Reem Alatrash, Dominik Schlechtweg and Sabine Schulte im Walde	21
<i>Representations of Language Varieties Are Reliable Given Corpus Similarity Measures</i> Jonathan Dunn	28
<i>Whit's the Richt Pairt o Speech: PoS tagging for Scots</i> Harm Lameris and Sara Stymne	39
<i>Efficient Unsupervised NMT for Related Languages with Cross-Lingual Language Models and Fidelity Objectives</i> Rami Aly, Andrew Caines and Paula Buttery	49
<i>Fine-tuning Distributional Semantic Models for Closely-Related Languages</i> Kushagra Bhatia, Divyanshu Aggarwal and Ashwini Vaidya	60
<i>Discriminating Between Similar Nordic Languages</i> René Haas and Leon Derczynski	67
<i>Naive Bayes-based Experiments in Romanian Dialect Identification</i> Tommi Jauhiainen, Heidi Jauhiainen and Krister Lindén	76
<i>UnibucKernel: Geolocating Swiss German Jodels Using Ensemble Learning</i> Gaman Mihaela, Sebastian Cojocariu and Radu Tudor Ionescu	84
<i>Optimizing a Supervised Classifier for a Difficult Language Identification Problem</i> Yves Bestgen	96
<i>Comparing the Performance of CNNs and Shallow Models for Language Identification</i> Andrea Ceolin	102
<i>Dialect Identification through Adversarial Learning and Knowledge Distillation on Romanian BERT</i> George-Eduard Zaharia, Andrei-Marius Avram, Dumitru-Clementin Cercel and Traian Rebedea	113
<i>Comparing Approaches to Dravidian Language Identification</i> Tommi Jauhiainen, Tharindu Ranasinghe and Marcos Zampieri	120
<i>N-gram and Neural Models for Uralic Language Identification: NRC at VarDial 2021</i> Gabriel Bernier-Colborne, Serge Leger and Cyril Goutte	128
<i>Social Media Variety Geolocation with geoBERT</i> Yves Scherrer and Nikola Ljubešić	135