

8th Workshop on Balto-Slavic Natural Language Processing (BSNLP 2021)

Online
20 April 2021

ISBN: 978-1-7138-2958-4

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

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>HerBERT: Efficiently Pretrained Transformer-based Language Model for Polish</i> Robert Mroczkowski, Piotr Rybak, Alina Wróblewska and Ireneusz Gawlik	1
<i>Russian Paraphrasers: Paraphrase with Transformers</i> Alena Fenogenova	11
<i>Abusive Language Recognition in Russian</i> Kamil Saitov and Leon Derczynski	20
<i>Detecting Inappropriate Messages on Sensitive Topics that Could Harm a Company’s Reputation</i> Nikolay Babakov, Varvara Logacheva, Olga Kozlova, Nikita Semenov and Alexander Panchenko	26
<i>BERTić - The Transformer Language Model for Bosnian, Croatian, Montenegrin and Serbian</i> Nikola Ljubešić and Davor Lauc	37
<i>RuSentEval: Linguistic Source, Encoder Force!</i> Vladislav Mikhailov, Ekaterina Taktasheva, Elina Sigdel and Ekaterina Artemova	43
<i>Exploratory Analysis of News Sentiment Using Subgroup Discovery</i> Anita Valmarska, Luis Cabrera-Diego, Elvys Linhares Pontes and Senja Pollak	66
<i>Creating an Aligned Russian Text Simplification Dataset from Language Learner Data</i> Anna Dmitrieva and Jörg Tiedemann	73
<i>Multilingual Named Entity Recognition and Matching Using BERT and Dedupe for Slavic Languages</i> Marko Prelevikj and Slavko Zitnik	80
<i>Priberam Labs at the 3rd Shared Task on SlavNER</i> Pedro Ferreira, Ruben Cardoso and Afonso Mendes	86
<i>Multilingual Slavic Named Entity Recognition</i> Rinalds Viksna and Inguna Skadina	93
<i>Using a Frustratingly Easy Domain and Tagset Adaptation for Creating Slavic Named Entity Recognition Systems</i> Luis Adrián Cabrera-Diego, Jose G. Moreno and Antoine Doucet	98
<i>Benchmarking Pre-trained Language Models for Multilingual NER: TraSpaS at the BSNLP2021 Shared Task</i> Marek Suppa and Ondrej Jariabka	105
<i>Named Entity Recognition and Linking Augmented with Large-Scale Structured Data</i> Paweł Rychlikowski, Adrian Lancucki, Adam Kaczmarek, Bartłomiej Najdecki, Adam Wawrzyński and Wojciech Janowski	115
<i>Slav-NER: the 3rd Cross-lingual Challenge on Recognition, Normalization, Classification, and Linking of Named Entities across Slavic Languages</i> Jakub Piskorski, Bogdan Babych, Zara Kancheva, Olga Kanishcheva, Maria Lebedeva, Michał Marcińczuk, Preslav Nakov, Petya Osenova, Lidia Pivovarova, Senja Pollak, Pavel Přibáň, Ivaylo Radev, Marko Robnik-Sikonja, Vasyl Starko, Josef Steinberger and Roman Yangarber	122