

2nd Ukrainian Natural Language Processing Workshop (UNLP 2023)

Dubrovnik, Croatia
5 May 2023

ISBN: 978-1-7138-7388-4

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© (2023) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2023)

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 UberText 2.0: A Corpus of Modern Ukrainian at Scale</i> Dmytro Chaplynskyi	1
<i>Contextual Embeddings for Ukrainian: A Large Language Model Approach to Word Sense Disambiguation</i> Yurii Laba, Volodymyr Mudryi, Dmytro Chaplynskyi, Mariana Romanyshyn and Oles Dobosevych	11
<i>Learning Word Embeddings for Ukrainian: A Comparative Study of FastText Hyperparameters</i> Nataliia Romanyshyn, Dmytro Chaplynskyi and Kyrylo Zakharov	20
<i>GPT-2 Metadata Pretraining Towards Instruction Finetuning for Ukrainian</i> Volodymyr Kyrylov and Dmytro Chaplynskyi	32
<i>The Evolution of Pro-Kremlin Propaganda From a Machine Learning and Linguistics Perspective</i> Veronika Solopova, Christoph Benz Müller and Tim Landgraf	40
<i>Abstractive Summarization for the Ukrainian Language: Multi-Task Learning with Hromadske.ua News Dataset</i> Svitlana Galeshchuk	49
<i>Extension Multi30K: Multimodal Dataset for Integrated Vision and Language Research in Ukrainian</i> Nataliia Saichyshyna, Daniil Maksymenko, Oleksii Turuta, Andriy Yerokhin, Andrii Babii and Olena Turuta	54
<i>Silver Data for Coreference Resolution in Ukrainian: Translation, Alignment, and Projection</i> Pavlo Kuchmiichuk	62
<i>Exploring Word Sense Distribution in Ukrainian with a Semantic Vector Space Model</i> Nataliia Cheilytko and Ruprecht von Waldenfels	73
<i>The Parliamentary Code-Switching Corpus: Bilingualism in the Ukrainian Parliament in the 1990s-2020s</i> Olha Kanishcheva, Tetiana Kovalova, Maria Shvedova and Ruprecht von Waldenfels	79
<i>Creating a POS Gold Standard Corpus of Modern Ukrainian</i> Vasyl Starko and Andriy Rysin	91
<i>UA-GEC: Grammatical Error Correction and Fluency Corpus for the Ukrainian Language</i> Oleksiy Syvokon, Olena Nahorna, Pavlo Kuchmiichuk and Nastasiia Osidach	96
<i>Comparative Study of Models Trained on Synthetic Data for Ukrainian Grammatical Error Correction</i> Maksym Bondarenko, Artem Yushko, Andrii Shportko and Andrii Fedorych	103
<i>A Low-Resource Approach to the Grammatical Error Correction of Ukrainian</i> Frank Palma Gomez, Alla Rozovskaya and Dan Roth	114
<i>RedPenNet for Grammatical Error Correction: Outputs to Tokens, Attentions to Spans</i> Bohdan Didenko and Andrii Sameliuk	121
<i>The UNLP 2023 Shared Task on Grammatical Error Correction for Ukrainian</i> Oleksiy Syvokon and Mariana Romanyshyn	132