EACL 2023

The 9th Workshop on Slavic Natural Language Processing 2023

Proceedings of the Workshop (SlavicNLP 2023)

May 6, 2023

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-959429-57-9

Introduction

This volume contains the papers presented at SlavNLP 2023: the 9^{th} Workshop on Natural Language Processing (NLP) for Slavic Languages. The workshop is organized by ACL SIGSLAV, the Special Interest Group of the Association for Computational Linguistics on NLP for Slavic Languages.

The SlavNLP / BSNLP workshops have been convening for over fifteen years, with a clear vision and purpose. On one hand, the languages from the Slavic group play an important role due to their widespread use and diverse cultural heritage. These languages are spoken by about one-third of all speakers of the official languages of the European Union, and by over 400 million speakers worldwide.

The current political and economic developments in Central and Eastern Europe—the foremost of which is the invasion of Ukraine by Russia—place the societies where Slavic languages are spoken at the center of events of global importance. Rapid technological advancement is urgently needed to help societies deal with massive flows of information—including counteracting the impact of disinformation, propaganda, etc.

On the other hand, despite the rapid growth of European consumer markets, research on theoretical and applied NLP in these languages still lags behind the "major" languages. In comparison to English, which has dominated the digital world since the advent of the Internet, many of these languages still lack resources, processing tools and applications—especially those with smaller communities of speakers.

The Slavic languages pose a wealth of fascinating scientific challenges. The linguistic phenomena specific to the Slavic languages—complex morphology and free word order—present non-trivial problems for the construction of NLP tools, and require rich morphological and syntactic resources.

The SlavNLP workshop brings together researchers in NLP for Slavic languages from academia and industry. We aim to stimulate research, foster the creation of tools and the dissemination of new results. The workshop serves as a forum for the exchange of ideas and experience and for discussing shared problems. One fascinating aspect of Slavic languages is their structural similarity, as well as an easily recognizable lexical and inflectional inventory spanning the entire group, which—despite the lack of mutual intelligibility—creates a special environment in which researchers can fully appreciate the shared problems and solutions.

In order to stimulate research and collaboration further, we have organized the fourth SIGSLAV Challenge: a Shared Task on multilingual named entity recognition (NER). Due to rich inflection, free word order, derivation, and other phenomena present in the Slavic languages, work on named entities is a challenging task.

Fostering research and development on the problems of named entities—detecting mentions of names, lemmatization (normalization), classification, and cross-lingual matching—is crucial for cross-lingual information access and for the wider use of NLP in Slavic languages. This edition of the challenge covers three languages: Czech, Polish, and Russian, building on the data from the second and the third editions of the shared task, which covered six languages: Bulgarian, Czech, Polish, Russian, Slovene, and Ukrainian. It covers five types of named entities: persons, locations, organizations, events, and products.

This year the workshop received 26 regular submissions, of which we selected 9 for oral presentation and 9 for poster presentation. Two additional presentations were based on ACL Findings papers, published by EACL separately. These papers cover a wide range of topics in NLP for various Slavic languages.

Seven teams registered to participate in the NER Challenge, of which three submitted results, and two submitted additional papers with descriptions of their NER systems. These papers are also included in this volume, and their work is discussed in the special session dedicated to the NER Challenge.

This workshop's presentation—the regular Workshop papers and the Shared Task Challenge—cover at least ten Slavic languages: Bosnian, Bulgarian, Croatian, Czech, Montenegrin, Polish, Russian, Serbian, Slovene, and Ukrainian.

This workshop continues the proud tradition established by the earlier BSNLP workshops, which were held in conjunction with the following venues:

- ACL 2007 Conference in Prague, Czech Republic.

- IIS 2009: Intelligent Information Systems, in Kraków, Poland.
- TSD 2011: 14th International Conference on Text, Speech and Dialogue in Plzeň, Czech Republic.
- ACL 2013 Conference in Sofia, Bulgaria.
- RANLP 2015 Conference in Hissar, Bulgaria.
- EACL 2017 Conference in Valencia, Spain.
- ACL 2019 Conference in Florence, Italy.
- EACL 2021 Conference in Kyiv, Ukraine.

We hope that this work will help stimulate further growth of our rich and exciting field.

The SlavNLP Organizers: Michał Marcińczuk, Preslav Nakov, Maciej Ogrodniczuk, Jakub Piskorski, Senja Pollak, Pavel Přibáň, Piotr Rybak, Josef Steinberger, Roman Yangarber

Organizing Committee

Workshop Organizer

Jakub Piskorski, Polish Academy of Sciences
Michał Marcińczuk, Wroclaw University of Science and Technology, Samurai Labs
Preslav Nakov, Mohamed bin Zayed University of Artificial Intelligence
Maciej Ogrodniczuk, Polish Academy of Sciences
Senja Pollak, Jožef Stefan Institute
Pavel Přibáň, University of West Bohemia
Piotr Rybak, Polish Academy of Sciences
Josef Steinberger, University of West Bohemia
Roman Yangarber, University of Helsinki

Program Committee

Program Committee

Željko Agić, Unity Technologies
Bogdan Babych, Heidelberg University
Radovan Garabík, Ľ. Štúr Institute of Linguistics, Slovak Academy of Sciences
Milos Jakubicek, Lexical Computing
Olha Kanishcheva, University of Jena
Anisia Katinskaia, University of Helsinki
Cvetana Krstev, University of Belgrade, Faculty of Philology
Vladislav Kubon, Charles University
Petya Osenova, Sofia University St. Kl. Ohridskiand IICT-BAS
Alexander Panchenko, Skolkovo Institue of Science and Technology
Maciej Piasecki, Wroclaw University of Science and Technology
Lidia Pivovarova, University of Helsinki
Pavel Přibáň, University of West Bohemia, Faculty of Applied Sciences
Marko Robnik-šikonja, University of Ljubljana, Faculty of Computer and Information Science
Alexandr Rosen, Charles University, Prague
Agata Savary, Paris-Saclay University
Serge Sharoff, University of Leeds
Josef Steinberger, University of West Bohemia
Marko Tadić, University of Zagreb, Faculty of Humanities and Social Sciences
Marcin Woliński, Institute of Computer Science, Polish Academy of Sciences

Table of Contents