

6th Workshop on Balto-Slavic Natural Language Processing (BSNLP 2017)

Valencia, Spain
4 April 2017

ISBN: 978-1-5108-3862-8

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© (2017) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2017)

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>Toward Pan-Slavic NLP: Some Experiments with Language Adaptation</i> Serge Sharoff	1
<i>Clustering of Russian Adjective-Noun Constructions using Word Embeddings</i> Andrey Kutuzov, Elizaveta Kuzmenko and Lidia Pivovarova	3
<i>A Preliminary Study of Croatian Lexical Substitution</i> Domagoj Alagić and Jan Šnajder	14
<i>Projecting Multiword Expression Resources on a Polish Treebank</i> Agata Savary and Jakub Waszczuk	20
<i>Lexicon Induction for Spoken Rusyn – Challenges and Results</i> Achim Rabus and Yves Scherrer	27
<i>The Universal Dependencies Treebank for Slovenian</i> Kaja Dobrovoljc, Tomaž Erjavec and Simon Krek	33
<i>Universal Dependencies for Serbian in Comparison with Croatian and Other Slavic Languages</i> Tanja Samardžić, Mirjana Starović, Željko Agić and Nikola Ljubešić	39
<i>Spelling Correction for Morphologically Rich Language: a Case Study of Russian</i> Alexey Sorokin	45
<i>Debunking Sentiment Lexicons: A Case of Domain-Specific Sentiment Classification for Croatian</i> Paula Gombar, Zoran Medić, Domagoj Alagić and Jan Šnajder	54
<i>Adapting a State-of-the-Art Tagger for South Slavic Languages to Non-Standard Text</i> Nikola Ljubešić, Tomaž Erjavec and Darja Fišer	60
<i>Comparison of Short-Text Sentiment Analysis Methods for Croatian</i> Leon Rotim and Jan Šnajder	69
<i>The First Cross-Lingual Challenge on Recognition, Normalization, and Matching of Named Entities in Slavic Languages</i> Jakub Piskorski, Lidia Pivovarova, Jan Šnajder, Josef Steinberger and Roman Yangarber	76
<i>Liner2 — a Generic Framework for Named Entity Recognition</i> Michał Marcińczuk, Jan Kocoń and Marcin Oleksy	86
<i>Language-Independent Named Entity Analysis Using Parallel Projection and Rule-Based Disambiguation</i> James Mayfield, Paul McNamee and Cash Costello	92
<i>Comparison of String Similarity Measures for Obscenity Filtering</i> Ekaterina Chernyak	97
<i>Stylometric Analysis of Parliamentary Speeches: Gender Dimension</i> Justina Mandravickaite and Tomas Krilavičius	102
<i>Towards Never Ending Language Learning for Morphologically Rich Languages</i> Kseniya Buraya, Lidia Pivovarova, Sergey Budkov and Andrey Filchenkov	108

Gender Profiling for Slovene Twitter communication: the Influence of Gender Marking, Content and Style
Ben Verhoeven, Iza Škrjanec and Senja Pollak 119