

Student Research Workshop 2021

Associated with the 13th International Conference on Recent Advances in Natural Language Processing (RANLP 2021)

Online
1 – 3 September 2021

ISBN: 978-1-7138-4096-1

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. (2022)

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>Humor Generation and Detection in Code-Mixed Hindi-English</i> Kaustubh Agarwal and Rhythm Narula	1
<i>Towards Code-Mixed Hinglish Dialogue Generation</i> Vibhav Agarwal, Pooja Rao and Dinesh Babu Jayagopi	7
<i>Hinglish to English Machine Translation using Multilingual Transformers</i> Vibhav Agarwal, Pooja Rao and Dinesh Babu Jayagopi	16
<i>SiPOS: A Benchmark Dataset for Sindhi Part-of-Speech Tagging</i> Wazir Ali, Zenglin Xu and Jay Kumar	22
<i>Sarcasm Detection and Building an English Language Corpus in Real Time</i> Oliver Cakebread-Andrews	31
<i>Correcting Texts Generated by Transformers using Discourse Features and Web Mining</i> Alexander Chernyavskiy, Dmitry Ilvovsky and Boris Galitsky	36
<i>Introducing linguistic transformation to improve translation memory retrieval. Results of a professional translators' survey for Spanish, French and Arabic</i> Souhila Djabri and Rocío Caro Quintana	44
<i>Using Transfer Learning to Automatically Mark L2 Writing Texts</i> Tim Elks	51
<i>Bilingual Terminology Extraction Using Neural Word Embeddings on Comparable Corpora</i> Darya Filippova, Burcu Can and Gloria Corpas Pastor	58
<i>Web-sentiment analysis of public comments (public reviews) for languages with limited resources such as the Kazakh language</i> Dinara Gimadi	65
<i>Disambiguating Grammatical Number and Gender With BERT</i> Annegret Janzso	69
<i>Towards a Language Model for Temporal Commonsense Reasoning</i> Mayuko Kimura, Lis Kanashiro Pereira and Ichiro Kobayashi	78
<i>Text Preprocessing and its Implications in a Digital Humanities Project</i> Maria Kunilovskaya and Alistair Plum	85
<i>Compiling a specialised corpus for translation research in the environmental domain</i> Anastasiia Laktionova	94
<i>Paragraph Similarity Matches for Generating Multiple-choice Test Items</i> Halyna Maslak and Ruslan Mitkov	99
<i>Neural Borrowing Detection with Monolingual Lexical Models</i> John Miller, Emanuel Pariasca and Cesar Beltran Castañon	109
<i>Does local pruning offer task-specific models to learn effectively ?</i> Abhishek Kumar Mishra and Mohna Chakraborty	118

<i>On Reducing Repetition in Abstractive Summarization</i>	
Pranav Nair and Anil Kumar Singh	126
<i>Improving Abstractive Summarization with Commonsense Knowledge</i>	
Pranav Nair and Anil Kumar Singh	135
<i>A Dataset for Research on Modelling Depression Severity in Online Forum Data</i>	
Isuri Anuradha Nanomi Arachchige, Vihangi Himaya Jayasuriya and Ruvan Weerasinghe	144
<i>Handling synset overgeneration: Sense Merging in BTB-WN</i>	
Ivaylo Radev and Zara Kancheva	154
<i>On the Evolution of Word Order</i>	
Idan Rejwan and Avi Caciularu	162
<i>EmoPars: A Collection of 30K Emotion-Annotated Persian Social Media Texts</i>	
Nazanin Sabri, Reyhane Akhavan and Behnam Bahrak	167
<i>A Review on Document Information Extraction Approaches</i>	
Kanishka Silva and Thushari Silva	174
<i>Towards New Generation Translation Memory Systems</i>	
Nikola Spasovski and Ruslan Mitkov	180
<i>Question answering in Natural Language: the Special Case of Temporal Expressions</i>	
Armand Stricker	184
<i>Building A Corporate Corpus For Threads Constitution</i>	
Lionel Tadonfouet Tadjou, Fabrice Bourge, Tiphaine Marie, Laurent Romary and Éric de la Clergerie	193
<i>Generating Answer Candidates for Quizzes and Answer-Aware Question Generators</i>	
Kristiyan Vachev, Momchil Hardalov, Georgi Karadzhov, Georgi Georgiev, Ivan Koychev and Preslav Nakov	203
<i>Toward Discourse-Aware Models for Multilingual Fake News Detection</i>	
Francielle Vargas, Fabrício Benevenuto and Thiago Pardo	210
<i>Automatic Transformation of Clinical Narratives into Structured Format</i>	
Sylvia Vassileva, Gergana Todorova, Kristina Ivanova, Boris Velichkov, Ivan Koychev, Galia Angelova and Svetla Boytcheva	219