Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2021)

Student Research Workshop

Online
6 – 11 June 2021

ISBN: 978-1-7138-3012-2

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

Sampling and Filtering of Neural Machine Translation Distillation Data
Vilém Zouhar 1

IceSum: An Icelandic Text Summarization Corpus
Jón Daðason, Hrafn Loftsson, Salome Sigurðardóttir and Þorsteinn Björnsson 9

Negation typology and general representation models for cross-lingual zero-shot negation scope resolution in Russian, French, and Spanish.
Anastassia Shaitarova and Fabio Rinaldi 15

Representations of Meaning in Neural Networks for NLP: a Thesis Proposal
Tomáš Musil 24

Towards Layered Events and Schema Representations in Long Documents
Hans Ole Hatzel and Chris Biemann 32

Parallel Text Alignment and Monolingual Parallel Corpus Creation from Philosophical Texts for Text Simplification
Stefan Paun 40

Syntax-Based Attention Masking for Neural Machine Translation
Colin McDonald and David Chiang 47

Multi-Modal Image Captioning for the Visually Impaired
Hiba Ahsan, Daivat Bhatt, Kaivan Shah and Nikita Bhalla 53

Open-Domain Question Answering with Pre-Constructed Question Spaces
Jinfeng Xiao, Lidan Wang, Franck Dernoncourt, Trung Bui, Tong Sun and Jiawei Han 61

A Sliding-Window Approach to Automatic Creation of Meeting Minutes
Jia Jin Koay, Alexander Roustai, Xiaojin Dai and Fei Liu 68

Exploration and Discovery of the COVID-19 Literature through Semantic Visualization
Jingxuan Tu, Marc Verhagen, Brent Cochran and James Pustejovsky 76

Shuffled-token Detection for Refining Pre-trained RoBERTa
Subhadarshi Panda, Anjali Agrawal, Jeewon Ha and Benjamin Bloch 88

Morphology-Aware Meta-Embeddings for Tamil
Arjun Sai Krishnan and Seyoon Ragavan 94

Seed Word Selection for Weakly-Supervised Text Classification with Unsupervised Error Estimation
Yiping Jin, Akshay Bhatia and Dittaya Wanvarie 112

Multi-Task Learning of Generation and Classification for Emotion-Aware Dialogue Response Generation
Tatsuya Ide and Daisuke Kawahara 119

Comparison of Grammatical Error Correction Using Back-Translation Models
Aomi Koyama, Kengo Hotate, Masahiro Kaneko and Mamoru Komachi 126

Parallel sentences mining with transfer learning in an unsupervised setting
Yu Sun, Shaolin Zhu, Feng Yifan and Chenggang Mi 136

Sentence Concatenation Approach to Data Augmentation for Neural Machine Translation
Seiichiro Kondo, Kengo Hotate, Tosho Hirasawa, Masahiro Kaneko and Mamoru Komachi . . . 143

Emotion Classification in a Resource Constrained Language Using Transformer-based Approach
Avishek Das, Omar Sharif, Mohammed Moshiul Hoque and Iqbal H. Sarker 150

Hie-BART: Document Summarization with Hierarchical BART
Kazuki Akiyama, Akihiro Tamura and Takashi Ninomiya 159