4th Workshop on Technologies for MT of Low Resource Languages (LoResMT 2021)

Held at MT Summit 2021

Online
16 – 20 August 2021

ISBN: 978-1-7138-3922-4

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© (2021) by the Association for Computational Linguistics
All rights reserved.
Copyright for individual papers remains with the authors and are licensed under a Creative Commons 4.0 license, CC-BY-ND. (https://creativecommons.org/licenses/by-nd/4.0/)

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

Contents

1 Dealing with the Paradox of Quality Estimation

Sugyeong Eo, Chanjun Park, Hyeonseok Moon, Jaehyung Seo and Heuiseok Lim

11 Small-Scale Cross-Language Authorship Attribution on Social Media Comments

Benjamin Murauer and Gunther Specht

20 Morphologically-Guided Segmentation For Translation of Agglutinative Low-Resource Languages

William Chen and Brett Fazio

32 Active Learning for Massively Parallel Translation of Constrained Text into Low Resource Languages

Zhong Zhou and Alex Waibel

44 Love Thy Neighbor: Combining Two Neighboring Low-Resource Languages for Translation

John E. Ortega, Richard Alexander Castro Mamani and Jaime Rafael Montoya Samame

52 Structural Biases for Improving Transformers on Translation into Morphologically Rich Languages

Paul Soulos, Sudha Rao, Caitlin Smith, Eric Rosen, Asli Celikyilmaz , R. Thomas McCoy, Yichen Jiang, Coleman Haley, Roland Fernandez, Hamid Palangi, Jianfeng Gao and Paul Smolensky

68 A Comparison of Different NMT Approaches to Low-Resource Dutch-Albanian Machine Translation

Arbnor Rama and Eva Vanmassenhove

78 Manipuri-English Machine Translation using Comparable Corpus

Lenin Laitonjam and Sanasam Ranbir Singh

89 EnKhCorp1.0: An English–Khasi Corpus

Sahinur Rahman Laskar, Abdullah Faiz Ur Rahman Khilji Darsh Kaushik, Partha Pakray and Sivaji Bandyopadhyay

96 Zero-Shot Neural Machine Translation with Self-Learning Cycle

Surafel M. Lakew, Matteo Negri and Marco Turchi

114 Findings of the LoResMT 2021 Shared Task on COVID and Sign Language for Low-resource Languages

Atul Kr. Ojha, Chao-Hong Liu, Katharina Kann, John Ortega, Sheetal Shatam and Theodorus Fransen

124 A3-108 Machine Translation System for LoResMT Shared Task @MT Summit 2021 Conference

Saumitra Yadav and Manish Shrivastava

129 The UCF Systems for the LoResMT 2021 Machine Translation Shared Task

William Chen and Brett Fazio

134 Attentive fine-tuning of Transformers for Translation of low-resourced languages @LoResMT 2021

Karthik Puranik, Adeep Hande, Ruba Priyadharshini, Thenmozi Durairaj, Anbukkarasi Sampath, Kingston Pal Thamburaj and Bharathi Raja Chakravarthi

144 Machine Translation in the Covid domain: an English-Irish case study for LoResMT 2021

Seamus Lankford, Haithem Afli and Andy Way

151 English-Marathi Neural Machine Translation for LoResMT 2021

Vandan Mujadia and Dipti Misra Sharma

158 Evaluating the Performance of Back-translation for Low Resource English-Marathi Language Pair: CFILT-IITBombay @ LoResMT 2021

Aditya Jain, Shivam Mhaskar and Pushpak Bhattacharyya