

First Workshop on Corpus Generation and Corpus Augmentation for Machine Translation

Held at AMTA 2022

Orlando, Florida, USA
12-16 September 2022

ISBN: 978-1-7138-6759-3

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© (2022) 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. (2023)

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 English-Russian Data Augmentation for Neural Machine Translation
Nikita Teslenko Grygoryev, Mercedes Garcia Martinez, Francisco Casacuberta Nolla,
Amando Estela Pastor, Manuel Herranz

- 11 Efficient Machine Translation Corpus Generation
Kamer Ali Yuksel, Ahmet Gunduz, Shreyas Sharma, Hassan Sawaf

- 18 Building and Analysis of Tamil Lyric Corpus with Semantic Representation
Karthika Ranganathan, Geetha T V

- 28 Ukrainian-To-English Folktale Corpus: Parallel Corpora Creation and Augmentation
for Machine Translation in Low-Resource Languages
Olena Burda-Lassen