

2nd Workshop for NLP Open Source Software (NLP-OSS 2020)

Online
19 November 2020

ISBN: 978-1-7138-2000-0

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© (2020) 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

<i>A Framework to Assist Chat Operators of Mental Healthcare Services</i> Thiago Madeira, Heder Bernardino, Jairo Francisco De Souza, Henrique Gomide, Nathália Munck Machado, Bruno Marcos Pinheiro da Silva and Alexandre Vieira Pereira Pacelli	1
<i>ARBML: Democratizing Arabic Natural Language Processing Tools</i> zaid alyafeai and Maged Al-Shaibani	8
<i>CLEVR Parser: A Graph Parser Library for Geometric Learning on Language Grounded Image Scenes</i> Raeid Saqur and Ameet Deshpande	14
<i>End-to-end NLP Pipelines in Rust</i> Guillaume Becquin	20
<i>Fair Embedding Engine: A Library for Analyzing and Mitigating Gender Bias in Word Embeddings</i> Vaibhav Kumar, Tenzin Bhotia and Vaibhav Kumar	26
<i>Flexible retrieval with NMSLIB and FlexNeuART</i> Leonid Boytsov and Eric Nyberg	32
<i>fugashi, a Tool for Tokenizing Japanese in Python</i> Paul McCann	44
<i>Going Beyond T-SNE: Exposing whatlies in Text Embeddings</i> Vincent Warmerdam, Thomas Kober and Rachael Tatman	52
<i>Howl: A Deployed, Open-Source Wake Word Detection System</i> Raphael Tang, Jaejun Lee, Afsaneh Razi, Julia Cambre, Ian Bicking, Jofish Kaye and Jimmy Lin	61
<i>iNLTK: Natural Language Toolkit for Indic Languages</i> Gaurav Arora	66
<i>KLPT – Kurdish Language Processing Toolkit</i> Sina Ahmadi	72
<i>Open Korean Corpora: A Practical Report</i> Won Ik Cho, Sangwhan Moon and Youngsook Song	85
<i>Open-Source Morphology for Endangered Mordvinic Languages</i> Jack Rueter, Mika Hämäläinen and Niko Partanen	94
<i>Pimlico: A toolkit for corpus-processing pipelines and reproducible experiments</i> Mark Granroth-Wilding	101
<i>PySBD: Pragmatic Sentence Boundary Disambiguation</i> Nipun Sadvilkar and Mark Neumann	110
<i>iobes: Library for Span Level Processing</i> Brian Lester	115
<i>SacreROUGE: An Open-Source Library for Using and Developing Summarization Evaluation Metrics</i> Daniel Deutsch and Dan Roth	120

<i>TextAttack: Lessons learned in designing Python frameworks for NLP</i>	
John Morris, Jin Yong Yoo and Yanjun Qi	126
<i>TOMODAPI: A Topic Modeling API to Train, Use and Compare Topic Models</i>	
Pasquale Lisena, Ismail Harrando, Oussama Kandakji and Raphael Troncy	132
<i>User-centered & Robust NLP OSS: Lessons Learned from Developing & Maintaining RSMTool</i>	
Nitin Madnani and Anastassia Loukina	141
<i>WAFFLE: A Graph for WordNet Applied to FreeForm Linguistic Exploration</i>	
Berk Ekmekci and Blake Howald	147