Fifth Workshop on Computational Approaches to Linguistic Code-Switching 2021

Online
11 June 2021

ISBN: 978-1-7138-3018-4

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.

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

Political Discourse Analysis: A Case Study of Code Mixing and Code Switching in Political Speeches
Dama Sravani, Lalitha Kameswari and Radhika Mamidi 1

Challenges and Limitations with the Metrics Measuring the Complexity of Code-Mixed Text
Vivek Srivastava and Mayank Singh 6

Translate and Classify: Improving Sequence Level Classification for English-Hindi Code-Mixed Data
Devansh Gautam, Kshitij Gupta and Manish Shrivastava 15

Gated Convolutional Sequence to Sequence Based Learning for English-Hingilsh Code-Switched Machine Translation.
Suman Dowlagar and Radhika Mamidi 26

IITP-MT at CALCS2021: English to Hinglish Neural Machine Translation using Unsupervised Synthetic Code-Mixed Parallel Corpus
Ramakrishna Appicharla, Kamal Kumar Gupta, Asif Ekbal and Pushpak Bhattacharyya 31

Exploring Text-to-Text Transformers for English to Hinglish Machine Translation with Synthetic Code-Mixing
Ganesh Jawahar, El Moatez Billah Nagoudi, Muhammad Abdul-Mageed and Laks Lakshmanan, V.S. 36

CoMeT: Towards Code-Mixed Translation Using Parallel Monolingual Sentences
Devansh Gautam, Prashant Kodali, Kshitij Gupta, Anmol Goel, Manish Shrivastava and Ponnurangam Kumaraguru 47

Investigating Code-Mixed Modern Standard Arabic-Egyptian to English Machine Translation
El Moatez Billah Nagoudi, AbdelRahim Elmadany and Muhammad Abdul-Mageed 56

Much Gracias: Semi-supervised Code-switch Detection for Spanish-English: How far can we get?
Dana-Maria Iliescu, Rasmus Grand, Sara Qirko and Rob van der Goot 65

A Language-aware Approach to Code-switched Morphological Tagging
Şaziye Betül Özateş and Özlem Çetinoğlu 72

Can You Traducir This? Machine Translation for Code-Switched Input
Jitao Xu and François Yvon 84

On the logistical difficulties and findings of Jopara Sentiment Analysis
Marvin Agüero-Torales, David Vilares and Antonio López-Herrera 95

Unsupervised Self-Training for Sentiment Analysis of Code-Switched Data
Akshat Gupta, Sargam Menghani, Sai Krishna Rallabandi and Alan W Black 103

CodemixedNLP: An Extensible and Open NLP Toolkit for Code-Mixing
Sai Muralidhar Jayanthi, Kavya Nerella, Khyathi Raghavi Chandu and Alan W Black 113

Normalization and Back-Transliteration for Code-Switched Data
Dwija Parikh and Thamar Solorio 119

Abusive content detection in transliterated Bengali-English social media corpus
Salim Sazzed 125

Developing ASR for Indonesian-English Bilingual Language Teaching
Zara Maxwelll-Smith and Ben Foley 131

Transliteration for Low-Resource Code-Switching Texts: Building an Automatic Cyrillic-to-Latin Converter for Tatar
Chihiro Taguchi, Yusuke Sakai and Taro Watanabe 133

Code-Mixing on Sesame Street: Dawn of the Adversarial Polyglots
Samson Tan and Shafiq Joty 141

Are Multilingual Models Effective in Code-Switching?
Genta Indra Winata, Samuel Cahyawijaya, Zihan Liu, Zhaojiang Lin, Andrea Madotto and Pascale Fung 142