Fifth Workshop on Arabic Natural Language Processing (WANLP 2020)

Held online due to COVID-19

Barcelona, Spain
12 December 2020

ISBN: 978-1-7138-2843-3

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© (2020) 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. (https://creativecommons.org/licenses/by/4.0/)

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

German-Arabic Speech-to-Speech Translation for Psychiatric Diagnosis
Juan Hussain, Mohammed Mediani, Moritz Behr, M. Amin Cheragui, Sebastian Stüker and Alexander Waibel 1

Hate Speech Detection in Saudi Twittersphere: A Deep Learning Approach
Raghad Alshaalan and Hend Al-Khalifa 12

Is it Great or Terrible? Preserving Sentiment in Neural Machine Translation of Arabic Reviews
Hadeel Saadany and Constantin Orasan 24

Deep Diacritization: Efficient Hierarchical Recurrence for Improved Arabic Diacritization
Badr AlKhamissi, Muhammad ElNokrashy and Mohamed Gabr 38

A Semi-Supervised BERT Approach for Arabic Named Entity Recognition
Chadi Helwe, Ghassan Dib, Mohsen Shamas and Shady Elbassuoni 49

Empathy-driven Arabic Conversational Chatbot
Tarek Naous, Christian Hokayem and Hazem Hajj 58

Machine Generation and Detection of Arabic Manipulated and Fake News
El Moatez Billah Nagoudi, AbdelRahim Elmadany, Muhammad Abdul-Mageed and Tariq Alhindi 69

Transliteration of Judeo-Arabic Texts into Arabic Script Using Recurrent Neural Networks
Ori Terner, Kfir Bar and Nachum Dershowitz 85

NADI 2020: The First Nuanced Arabic Dialect Identification Shared Task
Muhammad Abdul-Mageed, Chiyu Zhang, Houda Bouamor and Nizar Habash 97

Multi-dialect Arabic BERT for Country-level Dialect Identification
Bashar Talafha, Mohammad Ali, Muhy Eddin Za'ter, Haitham Seelawi, Ibraheem Tuffaha, Mostafa Samir, Wael Farhan and Hussein Al-Natsheh 111

On the Importance of Tokenization in Arabic Embedding Models
Mohamed Alkaoud and Mairaj Syed 119

Tracing Traditions: Automatic Extraction of Isnads from Classical Arabic Texts
Ryan Muther and David Smith 130

Embed More Ignore Less (EMIL): Exploiting Enriched Representations for Arabic NLP
ahmed younes and Julie Weeds 139

MANorm: A Normalization Dictionary for Moroccan Arabic Dialect Written in Latin Script
Randa Zarnoufi, Hamid Jaafar, Walid Bachri and Mounia Abik 155

A Unified Model for Arabizi Detection and Transliteration using Sequence-to-Sequence Models
Ali Shazal, Aiza Usman and Nizar Habash 167

Multi-Task Sequence Prediction For Tunisian Arabizi Multi-Level Annotation
elisa gugliotta, Marco Dinarelli and Olivier Kraif 178

AraWEAT: Multidimensional Analysis of Biases in Arabic Word Embeddings
Anne Lauscher, Rafik Takieddin, Simone Paolo Ponzetto and Goran Glavaš 192

Parallel resources for Tunisian Arabic Dialect Translation
Saméh Kchaou, Rahma Boujelbane and Lamia Hadrich-Belguith 200

Bert Transformer model for Detecting Arabic GPT2 Auto-Generated Tweets
Fouzi Harrag, Maria Dabbah, Kareem Darwish and Ahmed Abdelali 207

Contextual Embeddings for Arabic-English Code-Switched Data
Caroline Sabty, Mohamed Islam and Slim Abdennadher 215

Improving Arabic Text Categorization Using Transformer Training Diversification
Shammur Absar Chowdhury, Ahmed Abdelali, Kareem Darwish, Jung Soon-Gyo, Joni Salminen and Bernard J. Jansen 226

Team Alexa at NADI Shared Task
mutaz younes, Nour Al-khdour and Mohammad AL-Smadi 237

Comparison between Voting Classifier and Deep Learning methods for Arabic Dialect Identification
GHOUL dhaou and Gaël Lejeune 243

Simple vs Oversampling-based Classification Methods for Fine Grained Arabic Dialect Identification in Twitter
Mohamed Lichouri and Mourad Abbas 250

Semi-supervised Fine-grained Approach for Arabic dialect detection task
Nitin Nikamanth Appiah Balaji and Bharathi B 257

Arabic Dialect Identification Using BERT-Based Domain Adaptation
Ahmad Beltagy, Abdelrahman Abouelenin and Omar ElSherief 262

Weighted combination of BERT and N-GRAM features for Nuanced Arabic Dialect Identification
Abdellah El Mekki, Ahmed Alami, Hamza Alami, Ahmed Khoumsi and Ismail Berrada 268

Arabic dialect identification: An Arabic-BERT model with data augmentation and ensembling strategy
Kamel Gaanoun and Imade Benelallam 275

Faheem at NADI shared task: Identifying the dialect of Arabic tweet
Nouf AlShenaifi and Aqil Azmi 282

Identifying Nuanced Dialect for Arabic Tweets with Deep Learning and Reverse Translation Corpus Extension System
Rawan Tahssin, Youssef Kishk and Marwan Torki 288

The QMUL/HRBDT contribution to the NADI Arabic Dialect Identification Shared Task
Abdulrahman Aloraini, Massimo Poesio and Ayman Alhelbawy 295

Arabic Dialects Identification for All Arabic countries
Ahmed aliwy, Hawraa Taher and Zena AboAltaheen 302

Arabic Dialect Identification Using BERT Fine-Tuning
Moataz Mansour, Moustafa Tohamy, Zeyad Ezzat and Marwan Torki 308

LTG-ST at NADI Shared Task 1: Arabic Dialect Identification using a Stacking Classifier
Samia Touileb 313