Sixth Workshop on Arabic Natural Language Processing (WANLP 2021)

Held online due to COVID-19

Kiev, Ukraine
19 April 2021

ISBN: 978-1-7138-2972-0

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© (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

QADI: Arabic Dialect Identification in the Wild
Ahmed Abdelali, Hamdy Mubarak, Younes Samih, Sabit Hassan and Kareem Darwish 1

DiaLex: A Benchmark for Evaluating Multidialectal Arabic Word Embeddings
Muhammad Abdul-Mageed, Shady Elbassuoni, Jad Doughman, AbdelRahim Elmadany, El Moatez Billah Nagoudi, Yorgo Zoughby, Ahmad Shaher, Iskander Gaba, Ahmed Helal and Mohammed El-Razzaz 11

Benchmarking Transformer-based Language Models for Arabic Sentiment and Sarcasm Detection
Ibrahim Abu Farha and Walid Magdy 21

What does BERT Learn from Arabic Machine Reading Comprehension Datasets?
Eman Albilali, Nora Altwairesh and Manar Hosny 32

Kawarith: an Arabic Twitter Corpus for Crisis Events
Alaa Alharbi and Mark Lee 42

Arabic Compact Language Modelling for Resource Limited Devices
Zaid Alyafeai and Irfan Ahmad 53

Arabic Emoji Sentiment Lexicon (Arab-ESL): A Comparison between Arabic and European Emoji Sentiment Lexicons
Shatha Ali A. Hakami, Robert Hendley and Phillip Smith 60

ArCOV19-Rumors: Arabic COVID-19 Twitter Dataset for Misinformation Detection
Fatima Haouari, Maram Hasanain, Reem Suwaileh and Tamer Elsayed 72

ArCOV-19: The First Arabic COVID-19 Twitter Dataset with Propagation Networks
Fatima Haouari, Maram Hasanain, Reem Suwaileh and Tamer Elsayed 82

The Interplay of Variant, Size, and Task Type in Arabic Pre-trained Language Models
Go Inoue, Bashar Alhafni, Nurpeiis Baimukan, Houda Bouamor and Nizar Habash 92

Automatic Difficulty Classification of Arabic Sentences
Nouran Khallaf and Serge Sharoff 105

Dynamic Ensembles in Named Entity Recognition for Historical Arabic Texts
Muhammad Majadly and Tomer Sagi 115

Arabic Offensive Language on Twitter: Analysis and Experiments
Hamdy Mubarak, Ammar Rashed, Kareem Darwish, Younes Samih and Ahmed Abdelali 126

Adult Content Detection on Arabic Twitter: Analysis and Experiments
Hamdy Mubarak, Sabit Hassan and Ahmed Abdelali 136

UL2C: Mapping User Locations to Countries on Arabic Twitter
Hamdy Mubarak and Sabit Hassan 145

Let-Mi: An Arabic Levantine Twitter Dataset for Misogynistic Language
Hala Mulki and Bilal Ghanem 154

Empathetic BERT2BERT Conversational Model: Learning Arabic Language Generation with Little Data
Tarek Naous, Wissam Antoun, Reem Mahmoud and Hazem Hajj 164

ALUE: Arabic Language Understanding Evaluation
Haitham Seelawi, Ibraheem Tuffaha, Mahmoud Gzawi, Wael Farhan, Bashar Talafha, Riham Badawi, Zyad Sober, Oday Al-Dweik, Abed Alhakim Freihat and Hussein Al-Natsheh 173

Quranic Verses Semantic Relatedness Using AraBERT
Abdullah Alsaleh, Eric Atwell and Abdulrahman Altahhan 185

AraELECTRA: Pre-Training Text Discriminators for Arabic Language Understanding
Wissam Antoun, Fady Baly and Hazem Hajj 191

AraGPT2: Pre-Trained Transformer for Arabic Language Generation
Wissam Antoun, Fady Baly and Hazem Hajj 196

QuranTree.jl: A Julia Package for Quranic Arabic Corpus
Al-Ahmadgaid Asaad 208

Automatic Romanization of Arabic Bibliographic Records
Fadhl Eryani and Nizar Habash 213

SERAG: Semantic Entity Retrieval from Arabic Knowledge Graphs
Saher Esmeir 219

Introducing A large Tunisian Arabizi Dialectal Dataset for Sentiment Analysis
Chayma Fourati, Hatem Haddad, Abir Messaoudi, Moez BenHajhmida, Aymen Ben Elhaj Mabrouk and Malek Naski 226

AraFacts: The First Large Arabic Dataset of Naturally Occurring Claims
Zien Sheikh Ali, Watheq Mansour, Tamer Elsayed and Abdulaziz Al-Ali 231

Improving Cross-Lingual Transfer for Event Argument Extraction with Language-Universal Sentence Structures
Minh Van Nguyen and Thien Huu Nguyen 237

NADI 2021: The Second Nuanced Arabic Dialect Identification Shared Task
Muhammad Abdul-Mageed, Chiyu Zhang, AbdelRahim Elmadany, Houda Bouamor and Nizar Habash 244

Adapting MARBERT for Improved Arabic Dialect Identification: Submission to the NADI 2021 Shared Task
Badr AlKhamissi, Mohamed Gabr, Muhammad ElNokrashy and Khaled Essam 260

Country-level Arabic Dialect Identification Using Small Datasets with Integrated Machine Learning Techniques and Deep Learning Models
Maha J. Althobaiti 265

BERT-based Multi-Task Model for Country and Province Level MSA and Dialectal Arabic Identification
Abdellah El Mekki, Abdelkader El Mahdaouy, Kabil Essefar, Nabil El Mamoun, Ismail Berrada and Ahmed Khoumsi 271

Country-level Arabic Dialect Identification using RNNs with and without Linguistic Features
Elsayed Issa, Mohammed AlShakhori1, Reda Al-Bahrani and Gus Hahn-Powell 276

Arabic Dialect Identification based on a Weighted Concatenation of TF-IDF Features
Mohamed Lichouri, Mourad Abbas, Khaled Lounnas, Besma Benaziz and Aicha Zitouni 282

Machine Learning-Based Approach for Arabic Dialect Identification
Hamada Nayel, Ahmed Hassan, Mahmoud Sobhi and Ahmed El-Sawy 287

Dialect Identification in Nuanced Arabic Tweets Using Farasa Segmentation and AraBERT
Anshul Wadhawan 291

Overview of the WANLP 2021 Shared Task on Sarcasm and Sentiment Detection in Arabic
Ibrahim Abu Farha, Wajdi Zaghouani and Walid Magdy 296

WANLP 2021 Shared-Task: Towards Irony and Sentiment Detection in Arabic Tweets using Multi-headed-LSTM-CNN-GRU and MaRBERT
Reem Abdel-Salam 306

Sarcasm and Sentiment Detection In Arabic Tweets Using BERT-based Models and Data Augmentation
Abeer Abuzayed and Hend Al-Khalifa 312

Multi-task Learning Using a Combination of Contextualised and Static Word Embeddings for Arabic Sarcasm Detection and Sentiment Analysis
Abdullah I. Alharbi and Mark Lee 318

ArSarcasm Shared Task: An Ensemble BERT Model for SarcasmDetection in Arabic Tweets
Laila Bashmal and Daliyah AlZeer 323

Sarcasm and Sentiment Detection in Arabic: investigating the interest of character-level features
Dhaou Ghoul and Gaël Lejeune 329

Deep Multi-Task Model for Sarcasm Detection and Sentiment Analysis in Arabic Language
Abdelkader El Mahdaouy, Abdellah El Mekki, Kabil Essefar, Nabil El Mamoun, Ismail Berrada and Ahmed Khoumsi 334

A Contextual Word Embedding for Arabic Sarcasm Detection with Random Forests
Hazem Elgabry, Shimaa Attia, Ahmed Abdel-Rahman, Ahmed Abdel-Ate and Sandra Girgis . . 340

SarcasmDet at Sarcasm Detection Task 2021 in Arabic using AraBERT Pretrained Model
Dalya Faraj, Dalya Faraj and Malak Abdullah 345

Sarcasm and Sentiment Detection in Arabic language A Hybrid Approach Combining Embeddings and Rule-based Features
Kamel Gaanoun and Imade Benelallam 351

Combining Context-Free and Contextualized Representations for Arabic Sarcasm Detection and Sentiment Identification
Amey Hengle, Atharva Kshirsagar, Shaily Desai and Manisha Marathe 357

Leveraging Offensive Language for Sarcasm and Sentiment Detection in Arabic
Fatemah Husain and Ozlem Uzuner 364

The IDC System for Sentiment Classification and Sarcasm Detection in Arabic
Abraham Israeli, Yotam Nahum, Shai Fine and Kfir Bar 370

Preprocessing Solutions for Detection of Sarcasm and Sentiment for Arabic
Mohamed Lichouri, Mourad Abbas, Besma Benaziz, Aicha Zitouni and Khaled Lounnas 376

iCompass at Shared Task on Sarcasm and Sentiment Detection in Arabic
Malek Naski, Abir Messaoudi, Hatem Haddad, Moez BenHajhmida, Chayma Fourati and Aymen Ben Elhaj Mabrouk 381

Machine Learning-Based Model for Sentiment and Sarcasm Detection
Hamada Nayel, Eslam Amer, Aya Allam and Hanya Abdallah 386

DeepBlueAI at WANLP-EACL2021 task 2: A Deep Ensemble-based Method for Sarcasm and Sentiment Detection in Arabic
Bingyan Song, Chunguang Pan, Shengguang Wang and Zhipeng Luo 390

AraBERT and Farasa Segmentation Based Approach For Sarcasm and Sentiment Detection in Arabic Tweets
Anshul Wadhawan 395