

Fourth Workshop on Arabic Natural Language Processing (WANLP 2019)

Florence, Italy
1 August 2019

ISBN: 978-1-5108-9112-8

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© (2019) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2019)

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>Incremental Domain Adaptation for Neural Machine Translation in Low-Resource Settings</i> Marimuthu Kalimuthu, Michael Barz and Daniel Sonntag | 1 |
| <i>Morphology-aware Word-Segmentation in Dialectal Arabic Adaptation of Neural Machine Translation</i> Ahmed Tawfik, Mahitab Emam, Khaled Essam, Robert Nabil and Hany Hassan | 11 |
| <i>POS Tagging for Improving Code-Switching Identification in Arabic</i> Mohammed Attia, Younes Samih, Ali Elkahky, Hamdy Mubarak, Ahmed Abdelali and Kareem Darwish | 18 |
| <i>Syntax-Ignorant N-gram Embeddings for Sentiment Analysis of Arabic Dialects</i> Hala Mulki, Hatem Haddad, Mourad Gridach and Ismail Babaoğlu | 30 |
| <i>ArbEngVec : Arabic-English Cross-Lingual Word Embedding Model</i> Raki Lachraf, El Moatez Billah Nagoudi, Youcef Ayachi, Ahmed Abdelali and Didier Schwab .. | 40 |
| <i>Homograph Disambiguation through Selective Diacritic Restoration</i> Sawsan Alqahtani, Hanan Aldarmaki and Mona Diab | 49 |
| <i>Arabic Named Entity Recognition: What Works and What's Next</i> Liyuan Liu, Jingbo Shang and Jiawei Han | 60 |
| <i>hULMonA: The Universal Language Model in Arabic</i> Obeida ElJundi, Wissam Antoun, Nour El Droubi, Hazem Hajj, Wassim El-Hajj and Khaled Shaban 68 | |
| <i>Neural Models for Detecting Binary Semantic Textual Similarity for Algerian and MSA</i> Wafia Adouane, Jean-Philippe Bernardy and Simon Dobnik | 78 |
| <i>Constrained Sequence-to-sequence Semitic Root Extraction for Enriching Word Embeddings</i> Ahmed El-Kishky, Xingyu Fu, Aseel Addawood, Nahil Sobh, Clare Voss and Jiawei Han | 88 |
| <i>En-Ar Bilingual Word Embeddings without Word Alignment: Factors Effects</i> Taghreed Alqaisi and Simon O'Keefe | 97 |
| <i>Neural Arabic Question Answering</i> Hussein Mozannar, Elie Maamary, Karl El Hajal and Hazem Hajj | 108 |
| <i>Segmentation for Domain Adaptation in Arabic</i> Mohammed Attia and Ali Elkahky | 119 |
| <i>Assessing Arabic Weblog Credibility via Deep Co-learning</i> Chadi Helwe, Shady Elbassuoni, Ayman Al Zaatari and Wassim El-Hajj | 130 |
| <i>Morphologically Annotated Corpora for Seven Arabic Dialects: Taizi, Sanaani, Najdi, Jordanian, Syrian, Iraqi and Moroccan</i> Faisal Alshargi, Shahd Dibas, Sakhar Alkhereyf, Reem Faraj, Basmah Abdulkareem, Sane Yagi, Ouafaa Kacha, Nizar Habash and Owen Rambow | 137 |
| <i>Construction and Annotation of the Jordan Comprehensive Contemporary Arabic Corpus (JCCA)</i> Majdi Sawalha, Faisal Alshargi, Abdallah AlShdaifat, Sane Yagi and Mohammad A. Qudah .. | 148 |

| | |
|---|-----|
| <i>Translating Between Morphologically Rich Languages: An Arabic-to-Turkish Machine Translation System</i> | |
| İlknur Durgar El-Kahlout, Emre Bektaş, Naime Şeyma Erdem and Hamza Kaya | 158 |
| <i>Improved Generalization of Arabic Text Classifiers</i> | |
| Alaa Khaddaj, Hazem Hajj and Wassim El-Hajj | 167 |
| <i>OSIAN: Open Source International Arabic News Corpus - Preparation and Integration into the CLARIN-infrastructure</i> | |
| Imad Zeroual, Dirk Goldhahn, Thomas Eckart and Abdelhak Lakhouaja | 175 |
| <i>Arabic Tweet-Act: Speech Act Recognition for Arabic Asynchronous Conversations</i> | |
| Bushra Algotiml, AbdelRahim Elmadany and Walid Magdy | 183 |
| <i>Mazajak: An Online Arabic Sentiment Analyser</i> | |
| Ibrahim Abu Farha and Walid Magdy | 192 |
| <i>The MADAR Shared Task on Arabic Fine-Grained Dialect Identification</i> | |
| Houda Bouamor, Sabit Hassan and Nizar Habash | 199 |
| <i>ZCU-NLP at MADAR 2019: Recognizing Arabic Dialects</i> | |
| Pavel Přibáň and Stephen Taylor | 208 |
| <i>Simple But Not Naïve: Fine-Grained Arabic Dialect Identification Using Only N-Grams</i> | |
| Sohaila Eltanbouly, May Bashendy and Tamer Elsayed | 214 |
| <i>LIUM-MIRACL Participation in the MADAR Arabic Dialect Identification Shared Task</i> | |
| Saméh Kchaou, Fethi Bougares and Lamia Hadrich-Belguith | 219 |
| <i>Arabic Dialect Identification with Deep Learning and Hybrid Frequency Based Features</i> | |
| Youssef Fares, Zeyad El-Zanaty, Kareem Abdel-Salam, Muhammed Ezzeldin, Aliaa Mohamed, Karim El-Awaad and Marwan Torki | 224 |
| <i>MICHAEL: Mining Character-level Patterns for Arabic Dialect Identification (MADAR Challenge)</i> | |
| Dhaou Ghoul and Gaël Lejeune | 229 |
| <i>Arabic Dialect Identification for Travel and Twitter Text</i> | |
| Pruthwik Mishra and Vandan Mujadia | 234 |
| <i>Mawdoo3 AI at MADAR Shared Task: Arabic Tweet Dialect Identification</i> | |
| Bashar Talafha, Wael Farhan, Ahmed Altakrouri and Hussein Al-Natsheh | 239 |
| <i>Mawdoo3 AI at MADAR Shared Task: Arabic Fine-Grained Dialect Identification with Ensemble Learning</i> | |
| Ahmad Ragab, Haitham Seelawi, Mostafa Samir, Abdelrahman Mattar, Hesham Al-Bataineh, Mohammad Zaghoul, Ahmad Mustafa, Bashar Talafha, Abed Alhakim Freihat and Hussein Al-Natsheh | 244 |
| <i>Hierarchical Deep Learning for Arabic Dialect Identification</i> | |
| Gael de Francony, Victor Guichard, Praveen Joshi, Haithem Afi and Abdessalam Boucekif | 249 |
| <i>ArbDialectID at MADAR Shared Task 1: Language Modelling and Ensemble Learning for Fine Grained Arabic Dialect Identification</i> | |
| Kathrein Abu Kwaik and Motaz Saad | 254 |

| | |
|---|-----|
| <i>The SMarT Classifier for Arabic Fine-Grained Dialect Identification</i> Karima Meftouh, Karima Abidi, Salima Harrat and Kamel Smaili | 259 |
| <i>JHU System Description for the MADAR Arabic Dialect Identification Shared Task</i> Tom Lippincott, Pamela Shapiro, Kevin Duh and Paul McNamee | 264 |
| <i>ST MADAR 2019 Shared Task: Arabic Fine-Grained Dialect Identification</i> Mourad Abbas, Mohamed Lichouri and Abed Alhakim Freihat | 269 |
| <i>A Character Level Convolutional BiLSTM for Arabic Dialect Identification</i> Mohamed Elaraby and Ahmed Zahran | 274 |
| <i>No Army, No Navy: BERT Semi-Supervised Learning of Arabic Dialects</i> Chiyu Zhang and Muhammad Abdul-Mageed | 279 |
| <i>Team JUST at the MADAR Shared Task on Arabic Fine-Grained Dialect Identification</i> Bashar Talafha, Ali Fadel, Mahmoud Al-Ayyoub, Yaser Jararweh, Mohammad AL-Smadi and Patrick Juola | 285 |
| <i>QC-GO Submission for MADAR Shared Task: Arabic Fine-Grained Dialect Identification</i> Younes Samih, Hamdy Mubarak, Ahmed Abdelali, Mohammed Attia, Mohamed Eldesouki and Kareem Darwish | 290 |