ACL 2019

The Fourth Arabic Natural Language Processing Workshop
WANLP 2019

Proceedings of the Workshop

August 1, 2019
Florence, Italy

©2019 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-950737-32-1

Preface

Assalamu 3alaykum, benvenuti a tutti! Welcome to the The Fourth Arabic Natural Language Processing Workshop (WANLP 2019) held at ACL 2019 in Florence, Italy.

A number of Arabic NLP (or Arabic NLP-related) workshops and conferences have taken place in the last few years, both in the Arab World and in association with international conferences. The Arabic NLP workshop at ACL 2019 follows in the footsteps of these previous efforts to provide a forum for researchers to share and discuss their ongoing work. This particular workshop is the fourth in a series, following the First Arabic NLP workshop held at EMNLP 2014 in Doha, Qatar; the Second Arabic NLP workshop held at ACL 2015 in Beijing, China; the Third Arabic NLP workshop held at EACL 2017 in Valencia, Spain. This workshop included a shared task on Arabic dialect identification. As opposed to previous shared tasks which focused on regional level dialect labeling, this shared task is the first to target a large set of dialect labels at the city and country levels.

We received 49 main workshop submissions, out of which 22 were accepted, 23 were rejected, and 4 were withdrawn. All main workshop submissions were reviewed by at least three reviewers. The shared task was also a success with 17 teams participating. The shared task system descriptions (short) papers were reviewed by two reviewers each and will be included in the proceedings and presented during the workshop as posters. A long paper describing the shared task was submitted as part of the main workshop and got accepted for publication.

The acceptance rate of 49% in the main workshop, the quantity of the main workshop submissions (highest number of submissions in this workshop series), the shared task success, and the high quality of all contributions are strong indicators that there is a continued need for this kind of dedicated Arabic NLP workshop.

We would like to acknowledge all the hard work of the submitting authors and thank the reviewers for the valuable feedback they provided. We hope these proceedings will serve as a valuable reference for researchers and practitioners in the field of Arabic NLP and NLP in general.

Wassim El-Hajj, General Chair, on behalf of the organizers of the workshop.

Organizers

General chair

Wassim El-Hajj, American University of Beirut, Lebanon

Program Chairs

Lamia Hadrich Belguith, Sfax University, Tunisia
Fethi Bougares, University of Le Mans, France
Walid Magdy, University of Edinburgh, Scotland
Imed Zitouni, Microsoft

Publication Chairs

Nadi Tomeh, LIPN, Université Paris 13, Sorbonne Paris Cité
Mahmoud El-Haj, Lancaster University, England

Publicity Chair

Wajdi Zaghouani, Hamad Bin Khalifa University, Qatar

Ex-General Chair / Advisor

Nizar Habash, New York University Abu Dhabi, UAE

Advisory Committee

Hend Al-Khalifa, King Saud University, KSA
Houda Bouamor, Fortia Financial Solutions, France
Fethi Bougares, University of Le Mans, France
Kareem Darwish, Qatar Computing Research Institute, Qatar
Mona Diab, The George Washington University, USA
Mahmoud El-Haj, Lancaster University, England
Wassim El-Hajj, American University of Beirut, Lebanon
Nizar Habash, New York University Abu Dhabi, UAE
Nadi Tomeh, LIPN, Université Paris 13, Sorbonne Paris Cité
Wajdi Zaghouani, Hamad Bin Khalifa University , Qatar

Invited Keynote Speaker

Ahmed Ali, Qatar Computing Research Institute (QCRI)

Program Committee

Abdelali Ahmed, Qatar Computing Research Institute, Qatar
Abdul-Mageed, Muhammad , UBC, Canada
Afli Haithem, Cork Institute of Technology, Ireland
Al Sallab Ahmad, Faculty of Enginneeing, Cairo university
Ali Ahmed, Qatar Computing Research Institute, Qatar
Alkhalifa Hend, King Saud University, Saudi Arabia
Alowsiheq Areeb, Imam University, KSA
Al-Twairesh Nora, King Saud University, Saudi Arabia
Alzahrani Salha, Taif University, Saudi Arabia
Aransa Walid, University du Maine, Le Mans, France
Attia Mohammed, George Washington University
B. Al-Said Almoataz, Cairo University, Egypt
Badaro Gilbert, American University of Beirut
Barrón-Cedeño Alberto, Qatar Computing Research Institute, Qatar
Ben-Hamadou Abdelmajid, University of Sfax, Tunisia
Bouamor Houda, Fortia Financial Solutions, France
Bougares Fethi, Le Mans University, France
Bouzoubaa Karim, Mohammad V University, Morocco
Buckwalter Tim, University of Maryland, USA
Cavalli-Sforza Violetta, Al Akhawayn University, Morocco
Chafik Aloulou, Univeristé de Sfax, Tunisia
Choukri Khalid, ELDA, European Language Resource Association, France
Darwish Kareem, Qatar Computing Research Institute, Qatar
Dayel Abeer, King Saud University, Saudi Arabia
Diab Mona, George Washington University, USA
Dichy Joseph, Université Lyon 2 , France
El-Haj Mahmoud, Lancaster University, UK
El-Hajj Wassim, American University of Beirut, Lebanon
Elbassuoni Shady, American University of Beirut
Elkahky Ali, Google AI
Ellouze Mariem, University of Sfax, Tunisia
Elmadany AbdelRahim, Jazan Univeristy
Elmahdy Mohamed, Qatar University, Qatar
Elsayed Tamer, Qatar University, Qatar
Emam Ossama, IBM, USA
Eskander Ramy, Columbia University, USA
Fahmy Aly, Cairo University, Egypt
Farghaly Ali, Monterey Peninsula College, USA
Gargouri Bilel, University of Sfax, Tunisia
Ghannay Sahar, LIUM Laboratory, France
Ghneim Nada, Higher Institute for Applied Sciences and Technology, Syria
Habash Nizar, New York University Abu Dhabi, UAE

Haddad Bassam, University of Petra, Jordan
Hadrich Belguith Lamia, University of Sfax, Tunisia
Hajj Hazem, American University of Beirut, Lebanon
Hamada Salwa, Cairo University, Egypt
Hasanain Maram, Qatar University
Jarrar Mustafa, Bir Zeit University, Palestine
Khadivi Shahram, Tehran Polytechnic, Iran
Maamouri Mohamed, Linguistic Data Consortium, USA
Magdy Walid, University of Edinburgh, Scotland
Mazroui Azzeddine, University Mohamed I, Morocco
Mechti Seif, University of Sfax, Tunisia
Medhaffar salima, Le Mans University, France
Megerdoomian Karine, The MITRE Corporation, USA
Mohamed Emad, Suez Canal University, Egypt
Mourad Ghassan, Lebanese University, Lebanon
Mubarak Hamdy, Qatar Computing Research Institute, Qatar
Nakov Preslav, Qatar Computing Research Institute, Qatar
Nasr Alexis, University of Marseille, France
Nwesri Abdelsalam, University of Tripoli, Libya
Oflazer Kemal, Carnegie Mellon University Qatar, Qatar
Rafea Ahmed, The American University in Cairo, Egypt
Rambow Owen, Columbia University, USA
Rashwan Mohsen, RDI, Egypt
Refaee Eshrag, Jazan University, Saudi Arabia
Salameh Mohammad, Carnegie Mellon University, Qatar
Sawaf Hassan, eBay Inc., USA
Shaalan Khaled, The British University in Dubai, UAE
Shaban Khaled, Qatar University, Qatar
Smrž Otakar, ÚFAL, Czech Republic
Suwaileh Reem, Qatar University, Qatar
Tomeh Nadi, University Paris 13, France
Trigui Omar, University of Sousse, Tunisia
Vogel Stephan, Qatar Computing Research Institute, Qatar
Wray Samantha, Qatar Computing Research Institute, Qatar
Zaghouani Wajdi, Hamad Bin Khalifa University, Qatar
Zerrouki Taha, University of Bouira, Algeria
Zitouni Imed, Microsoft Research, USA
Zribi Ines, Sfax University, Tunisia

Table of Contents

Incremental Domain Adaptation for Neural Machine Translation in Low-Resource Settings
Marimuthu Kalimuthu, Michael Barz and Daniel Sonntag 1

Morphology-aware Word-Segmentation in Dialectal Arabic Adaptation of Neural Machine Translation
Ahmed Tawfik, Mahitab Emam, Khaled Essam, Robert Nabil and Hany Hassan 11

POS Tagging for Improving Code-Switching Identification in Arabic
Mohammed Attia, Younes Samih, Ali Elkahky, Hamdy Mubarak, Ahmed Abdelali and Kareem Darwish 18

Syntax-Ignorant N-gram Embeddings for Sentiment Analysis of Arabic Dialects
Hala Mulki, Hatem Haddad, Mourad Gridach and Ismail Babaoğlu 30

ArbEngVec : Arabic-English Cross-Lingual Word Embedding Model
Raki Lachraf, El Moatez Billah Nagoudi, Youcef Ayachi, Ahmed Abdelali and Didier Schwab . 40

Homograph Disambiguation through Selective Diacritic Restoration
Sawsan Alqahtani, Hanan Aldarmaki and Mona Diab 49

Arabic Named Entity Recognition: What Works and What's Next
Liyuan Liu, Jingbo Shang and Jiawei Han 60

hULMonA: The Universal Language Model in Arabic
Obeida ElJundi, Wissam Antoun, Nour El Droubi, Hazem Hajj, Wassim El-Hajj and Khaled Shaban 68

Neural Models for Detecting Binary Semantic Textual Similarity for Algerian and MSA
Wafia Adouane, Jean-Philippe Bernardy and Simon Dobnik 78

Constrained Sequence-to-sequence Semitic Root Extraction for Enriching Word Embeddings
Ahmed El-Kishky, Xingyu Fu, Aseel Addawood, Nahil Sobh, Clare Voss and Jiawei Han 88

En-Ar Bilingual Word Embeddings without Word Alignment: Factors Effects
Taghreed Alqaisi and Simon O'Keefe 97

Neural Arabic Question Answering
Hussein Mozannar, Elie Maamary, Karl El Hajal and Hazem Hajj 108

Segmentation for Domain Adaptation in Arabic
Mohammed Attia and Ali Elkahky 119

Assessing Arabic Weblog Credibility via Deep Co-learning
Chadi Helwe, Shady Elbassuoni, Ayman Al Zaatari and Wassim El-Hajj 130

Morphologically Annotated Corpora for Seven Arabic Dialects: Taizi, Sanaani, Najdi, Jordanian, Syrian, Iraqi and Moroccan
Faisal Alshargi, Shahd Dibas, Sakhar Alkhereyf, Reem Faraj, Basmah Abdulkareem, Sane Yagi, Ouafaa Kacha, Nizar Habash and Owen Rambow 137

Construction and Annotation of the Jordan Comprehensive Contemporary Arabic Corpus (JCCA)
Majdi Sawalha, Faisal Alshargi, Abdallah AlShdaifat, Sane Yagi and Mohammad A. Qudah . . 148

Translating Between Morphologically Rich Languages: An Arabic-to-Turkish Machine Translation System
İlknur Durgar El-Kahlout, Emre Bektaş, Naime Şeyma Erdem and Hamza Kaya 158

Improved Generalization of Arabic Text Classifiers
Alaa Khaddaj, Hazem Hajj and Wassim El-Hajj 167

OSIAN: Open Source International Arabic News Corpus - Preparation and Integration into the CLARIN-infrastructure
Imad Zeroual, Dirk Goldhahn, Thomas Eckart and Abdelhak Lakhouaja 175

Arabic Tweet-Act: Speech Act Recognition for Arabic Asynchronous Conversations
Bushra Algotiml, AbdelRahim Elmadany and Walid Magdy 183

Mazajak: An Online Arabic Sentiment Analyser
Ibrahim Abu Farha and Walid Magdy 192

The MADAR Shared Task on Arabic Fine-Grained Dialect Identification
Houda Bouamor, Sabit Hassan and Nizar Habash 199

ZCU-NLP at MADAR 2019: Recognizing Arabic Dialects
Pavel Přibáň and Stephen Taylor 208

Simple But Not Naïve: Fine-Grained Arabic Dialect Identification Using Only N-Grams
Sohaila Eltanbouly, May Bashendy and Tamer Elsayed 214

LIUM-MIRACL Participation in the MADAR Arabic Dialect Identification Shared Task
Saméh Kchaou, Fethi Bougares and Lamia Hadrich-Belguith 219

Arabic Dialect Identification with Deep Learning and Hybrid Frequency Based Features
Youssef Fares, Zeyad El-Zanaty, Kareem Abdel-Salam, Muhammed Ezzeldin, Aliaa Mohamed, Karim El-Awaad and Marwan Torki 224

MICHAEL: Mining Character-level Patterns for Arabic Dialect Identification (MADAR Challenge)
Dhaou Ghoul and Gaël Lejeune 229

Arabic Dialect Identification for Travel and Twitter Text
Pruthwik Mishra and Vandan Mujadia 234

Mawdoo3 AI at MADAR Shared Task: Arabic Tweet Dialect Identification
Bashar Talafha, Wael Farhan, Ahmed Altakrouri and Hussein Al-Natsheh 239

Mawdoo3 AI at MADAR Shared Task: Arabic Fine-Grained Dialect Identification with Ensemble Learning
Ahmad Ragab, Haitham Seelawi, Mostafa Samir, Abdelrahman Mattar, Hesham Al-Bataineh, Mohammad Zaghloul, Ahmad Mustafa, Bashar Talafha, Abed Alhakim Freihat and Hussein Al-Natsheh 244

Hierarchical Deep Learning for Arabic Dialect Identification
Gael de Francony, Victor Guichard, Praveen Joshi, Haithem Afli and Abdessalam Bouchekif . . 249

ArbDialectID at MADAR Shared Task 1: Language Modelling and Ensemble Learning for Fine Grained Arabic Dialect Identification
Kathrein Abu Kwaik and Motaz Saad 254

The SMarT Classifier for Arabic Fine-Grained Dialect Identification
Karima Meftouh, Karima Abidi, Salima Harrat and Kamel Smaili 259

JHU System Description for the MADAR Arabic Dialect Identification Shared Task
Tom Lippincott, Pamela Shapiro, Kevin Duh and Paul McNamee 264

ST MADAR 2019 Shared Task: Arabic Fine-Grained Dialect Identification
Mourad Abbas, Mohamed Lichouri and Abed Alhakim Freihat 269

A Character Level Convolutional BiLSTM for Arabic Dialect Identification
Mohamed Elaraby and Ahmed Zahran 274

No Army, No Navy: BERT Semi-Supervised Learning of Arabic Dialects
Chiyu Zhang and Muhammad Abdul-Mageed 279

Team JUST at the MADAR Shared Task on Arabic Fine-Grained Dialect Identification
Bashar Talafha, Ali Fadel, Mahmoud Al-Ayyoub, Yaser Jararweh, Mohammad AL-Smadi and Patrick Juola 285

QC-GO Submission for MADAR Shared Task: Arabic Fine-Grained Dialect Identification
Younes Samih, Hamdy Mubarak, Ahmed Abdelali, Mohammed Attia, Mohamed Eldesouki and Kareem Darwish 290

Conference Program

Thursday, August 1, 2019

8:30–8:40 *Opening Remarks*
Wassim El-Hajj

8:40–9:30 *Invited Keynote*
Ahmed Ali

9:30–10:20 Session 1: Machine Translation

9:30–9:55 *Incremental Domain Adaptation for Neural Machine Translation in Low-Resource Settings*
Marimuthu Kalimuthu, Michael Barz and Daniel Sonntag

9:55–10:20 *Morphology-aware Word-Segmentation in Dialectal Arabic Adaptation of Neural Machine Translation*
Ahmed Tawfik, Mahitab Emam, Khaled Essam, Robert Nabil and Hany Hassan

10:20–10:30 ***The MADAR Shared Task Overview***

The MADAR Shared Task on Arabic Fine-Grained Dialect Identification
Houda Bouamor, Sabit Hassan and Nizar Habash

10:30–11:00 ***Coffee Break***

11:00–12:40 Session 2: Selected Topics

11:00–11:25 *POS Tagging for Improving Code-Switching Identification in Arabic*
Mohammed Attia, Younes Samih, Ali Elkahky, Hamdy Mubarak, Ahmed Abdelali and Kareem Darwish

11:25–11:50 *Syntax-Ignorant N-gram Embeddings for Sentiment Analysis of Arabic Dialects*
Hala Mulki, Hatem Haddad, Mourad Gridach and Ismail Babaoğlu

11:50–12:15 *ArbEngVec : Arabic-English Cross-Lingual Word Embedding Model*
Raki Lachraf, El Moatez Billah Nagoudi, Youcef Ayachi, Ahmed Abdelali and Didier Schwab

12:15–12:40 *Homograph Disambiguation through Selective Diacritic Restoration*
Sawsan Alqahtani, Hanan Aldarmaki and Mona Diab

Thursday, August 1, 2019 (continued)

12:40–14:00 ***Lunch***

14:00–14:50 **Session 3: Applications**

14:00–14:25 *Arabic Named Entity Recognition: What Works and What's Next*
Liyuan Liu, Jingbo Shang and Jiawei Han

14:25–14:50 *hULMonA: The Universal Language Model in Arabic*
Obeida ElJundi, Wissam Antoun, Nour El Droubi, Hazem Hajj, Wassim El-Hajj and Khaled Shaban

14:50–15:30 *Workshop Poster Boaster*
2.5 minutes per poster

15:30–16:00 ***Coffee Break***

16:00–18:00 **Poster Session**

+ ***Posters: Main Workshop Papers***

Neural Arabic Question Answering
Hussein Mozannar, Elie Maamary, Karl El Hajal and Hazem Hajj

Neural Models for Detecting Binary Semantic Textual Similarity for Algerian and MSA
Wafia Adouane, Jean-Philippe Bernardy and Simon Dobnik

Constrained Sequence-to-sequence Semitic Root Extraction for Enriching Word Embeddings
Ahmed El-Kishky, Xingyu Fu, Aseel Addawood, Nahil Sobh, Clare Voss and Jiawei Han

En-Ar Bilingual Word Embeddings without Word Alignment: Factors Effects
Taghreed Alqaisi and Simon O'Keefe

Segmentation for Domain Adaptation in Arabic
Mohammed Attia and Ali Elkahky

Assessing Arabic Weblog Credibility via Deep Co-learning
Chadi Helwe, Shady Elbassuoni, Ayman Al Zaatari and Wassim El-Hajj

Morphologically Annotated Corpora for Seven Arabic Dialects: Taizi, Sanaani, Najdi, Jordanian, Syrian, Iraqi and Moroccan
Faisal Alshargi, Shahd Dibas, Sakhar Alkhereyf, Reem Faraj, Basmah Abdulkareem, Sane Yagi, Ouafaa Kacha, Nizar Habash and Owen Rambow

Construction and Annotation of the Jordan Comprehensive Contemporary Arabic Corpus (JCCA)
Majdi Sawalha, Faisal Alshargi, Abdallah AlShdaifat, Sane Yagi and Mohammad A. Qudah

Translating Between Morphologically Rich Languages: An Arabic-to-Turkish Machine Translation System
İlknur Durgar El-Kahlout, Emre Bektaş, Naime Şeyma Erdem and Hamza Kaya

Improved Generalization of Arabic Text Classifiers
Alaa Khaddaj, Hazem Hajj and Wassim El-Hajj

OSIAN: Open Source International Arabic News Corpus - Preparation and Integration into the CLARIN-infrastructure
Imad Zeroual, Dirk Goldhahn, Thomas Eckart and Abdelhak Lakhouaja

Arabic Tweet-Act: Speech Act Recognition for Arabic Asynchronous Conversations
Bushra Algotiml, AbdelRahim Elmadany and Walid Magdy

Mazajak: An Online Arabic Sentiment Analyser
Ibrahim Abu Farha and Walid Magdy

+ ***Posters: Shared Task Papers***

ZCU-NLP at MADAR 2019: Recognizing Arabic Dialects
Pavel Přibáň and Stephen Taylor

Simple But Bot Naïve: Fine-Grained Arabic Dialect Identification Using Only N-Grams
Sohaila Eltanbouly, May Bashendy and Tamer Elsayed

LIUM-MIRACL Participation in the MADAR Arabic Dialect Identification Shared Task
Saméh Kchaou, Fethi Bougares and Lamia Hadrich-Belguith

Arabic Dialect Identification with Deep Learning and Hybrid Frequency Based Features
Youssef Fares, Zeyad El-Zanaty, Kareem Abdel-Salam, Muhammed Ezzeldin, Aliaa Mohamed, Karim El-Awaad and Marwan Torki

MICHAEL: Mining Character-level Patterns for Arabic Dialect Identification (MADAR Challenge)
Dhaou Ghoul and Gaël Lejeune

Arabic Dialect Identification for Travel and Twitter Text
Pruthwik Mishra and Vandan Mujadia

Mawdoo3 AI at MADAR Shared Task: Arabic Tweet Dialect Identification
Bashar Talafha, Wael Farhan, Ahmed Altakrouri and Hussein Al-Natsheh

Mawdoo3 AI at MADAR Shared Task: Arabic Fine-Grained Dialect Identification with Ensemble Learning
Ahmad Ragab, Haitham Seelawi, Mostafa Samir, Abdelrahman Mattar, Hesham Al-Bataineh, Mohammad Zaghloul, Ahmad Mustafa, Bashar Talafha, Abed Alhakim Freihat and Hussein Al-Natsheh

Hierarchical Deep Learning for Arabic Dialect Identification
Gael de Francony, Victor Guichard, Praveen Joshi, Haithem Afli and Abdessalam Bouchekif

ArbDialectID at MADAR Shared Task 1: Language Modelling and Ensemble Learning for Fine Grained Arabic Dialect Identification
Kathrein Abu Kwaik and Motaz Saad

The SMarT Classifier for Arabic Fine-Grained Dialect Identification
Karima Meftouh, Karima Abidi, Salima Harrat and Kamel Smaili

JHU System Description for the MADAR Arabic Dialect Identification Shared Task
Tom Lippincott, Pamela Shapiro, Kevin Duh and Paul McNamee

ST MADAR 2019 Shared Task: Arabic Fine-Grained Dialect Identification
Mourad Abbas, Mohamed Lichouri and Abed Alhakim Freihat

A Character Level Convolutional BiLSTM for Arabic Dialect Identification
Mohamed Elaraby and Ahmed Zahran

No Army, No Navy: BERT Semi-Supervised Learning of Arabic Dialects
Chiyu Zhang and Muhammad Abdul-Mageed

Team JUST at the MADAR Shared Task on Arabic Fine-Grained Dialect Identification
Bashar Talafha, Ali Fadel, Mahmoud Al-Ayyoub, Yaser Jararweh, Mohammad AL-Smadi and Patrick Juola

QC-GO Submission for MADAR Shared Task: Arabic Fine-Grained Dialect Identification
Younes Samih, Hamdy Mubarak, Ahmed Abdelali, Mohammed Attia, Mohamed Eldesouki and Kareem Darwish