

Third Workshop on Arabic Natural Language Processing (WANLP 2017)

Co-located with EACL 2017

Valencia, Spain
3 April 2017

ISBN: 978-1-5108-3874-1

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© (2017) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2017)

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>Identification of Languages in Algerian Arabic Multilingual Documents</i>	
Wafia Adouane and Simon Dobnik	1
<i>Arabic Diacritization: Stats, Rules, and Hacks</i>	
Kareem Darwish, Hamdy Mubarak and Ahmed Abdelali	9
<i>Semantic Similarity of Arabic Sentences with Word Embeddings</i>	
El Moatez Billah Nagoudi and Didier Schwab	18
<i>Morphological Analysis for the Maltese Language: The challenges of a hybrid system</i>	
Claudia Borg and Albert Gatt	25
<i>A Morphological Analyzer for Gulf Arabic Verbs</i>	
Salam Khalifa, Sara Hassan and Nizar Habash	35
<i>A Neural Architecture for Dialectal Arabic Segmentation</i>	
Younes Samih, Mohammed Attia, Mohamed Eldesouki, Ahmed Abdelali, Hamdy Mubarak, Laura Kallmeyer and Kareem Darwish	46
<i>Sentiment Analysis of Tunisian Dialects: Linguistic Ressources and Experiments</i>	
Salima Medhaffar, Fethi Bougares, Yannick Estève and Lamia Hadrich-Belguith	55
<i>CAT: Credibility Analysis of Arabic Content on Twitter</i>	
Rim El Ballouli, Wassim El-Hajj, Ahmad Ghandour, Shady Elbassuoni, Hazem Hajj and Khaled Shaban	62
<i>A New Error Annotation for Dyslexic texts in Arabic</i>	
Maha Alamri and William J Teahan	72
<i>An Unsupervised Speaker Clustering Technique based on SOM and I-vectors for Speech Recognition Systems</i>	
Hany Ahmed, Mohamed Elaraby, Abdullah M. Mousa, Mostafa Elhosiny, Sherif Abdou and Mohsen Rashwan	79
<i>SHAKKIL: An Automatic Diacritization System for Modern Standard Arabic Texts</i>	
Amany Fashwan and Sameh Alansary	84
<i>Arabic Tweets Treebanking and Parsing: A Bootstrapping Approach</i>	
Fahad Albogamy, Allan Ramsay and Hanady Ahmed	94
<i>Identifying Effective Translations for Cross-lingual Arabic-to-English User-generated Speech Search</i>	
Ahmad Khwileh, Haithem Afli, Gareth Jones and Andy Way	100
<i>A Characterization Study of Arabic Twitter Data with a Benchmarking for State-of-the-Art Opinion Mining Models</i>	
Ramy Baly, Gilbert Badaro, Georges El-Khoury, Rawan Moukalled, Rita Aoun, Hazem Hajj, Wassim El-Hajj, Nizar Habash and Khaled Shaban	110
<i>Robust Dictionary Lookup in Multiple Noisy Orthographies</i>	
Lingliang Zhang, Nizar Habash and Godfried Toussaint	119

<i>Arabic POS Tagging: Don't Abandon Feature Engineering Just Yet</i>	
Kareem Darwish, Hamdy Mubarak, Ahmed Abdelali and Mohamed Eldesouki	130
<i>Toward a Web-based Speech Corpus for Algerian Dialectal Arabic Varieties</i>	
Soumia Bougrine, Aicha Chorana, Abdallah Lakhdari and Hadda Cherroun	138
<i>Not All Segments are Created Equal: Syntactically Motivated Sentiment Analysis in Lexical Space</i>	
Muhammad Abdul-Mageed	147
<i>An enhanced automatic speech recognition system for Arabic</i>	
Mohamed Amine Menacer, Odile Mella, Dominique Fohr, Denis Jouviet, David Langlois and Kamel Smaili	157
<i>Universal Dependencies for Arabic</i>	
Dima Taji, Nizar Habash and Daniel Zeman	166
<i>A Layered Language Model based Hybrid Approach to Automatic Full Diacritization of Arabic</i>	
Mohamed Al-Badrashiny, Abdelati Hawwari and Mona Diab	177
<i>Arabic Textual Entailment with Word Embeddings</i>	
Nada Almarwani and Mona Diab	185