

The Third Arabic Natural Language Processing Workshop

(WANLP-2017)

co-located with EACL 2017, Valencia, Spain

Important Dates

~~Nov 8, 2016: First Call for Workshop Papers~~

~~Dec 7, 2016: Second Call for Workshop Papers~~

Jan 28, 2017 (Midnight PST): Workshop Paper Due Date (Deadline Extended)

~~Feb 11, 2017: Notification of Acceptance~~

~~Feb 21, 2017 (Midnight PST): Feb 22, 2017 (Midnight PST): Camera-ready papers due~~

~~Feb 23, 2017 Workshop Schedule~~

Monday April 3, 2017: Workshop Date (one day workshop)

Proceedings

[Proceedings of the Third Arabic Natural Language Processing Workshop](#)

Workshop Description

Arabic is a challenging language for the field of computational linguistics. This is due to many factors including its complex and rich morphology, its high degree of ambiguity as well as the presence of a number of dialects that vary quite widely. Arabic is also a language with important geopolitical connections. It is spoken by over 400 million people in countries with varying degrees of prosperity and stability. It is the primary language of the latest world refugee problem affecting the Middle East and Europe. The opportunities that are made possible by working on this language and its dialects cannot be underestimated in their consequence on the Arab World, the Mediterranean Region and the rest of the World.

There has been a lot of progress in the last 15 years in the area of Arabic Natural Language Processing (NLP). Many Arabic NLP (or Arabic NLP-related) workshops and

conferences have taken place, both in the Arab World and in association with international conferences. Examples include the following:

- The First Arabic Natural Language Processing Workshop at EMNLP 2014, and The Second Arabic Natural Language Processing Workshop at ACL 2015.
- The 1st and 2nd Workshop on Arabic Corpora and Processing Tools (LREC 2014, LREC 2016).
- The conference on Arabic Language Resources and Tools (MEDAR-2009, NEMLAR-2004).
- The workshop on Computational Approaches to Semitic Languages (LREC 2010, EACL 2009, ACL 2007, ACL 2005, ACL 2002, ACL 1998).
- The workshop on Computational Approaches to Arabic Script-based Languages (MTSummit XII 2009, LSA 2007, COLING 2004).
- The International Symposium on Computer and Arabic Language (ISCAL 2009, ISCAL 2007)

This workshop follows in the footsteps of these efforts to provide a forum for researchers to share and discuss their ongoing work. This workshop is timely given the continued rise in research projects focusing on Arabic NLP.

We invite submissions on topics that include, but are not limited to, the following:

- Basic core technologies: morphological analysis, disambiguation, tokenization, POS tagging, named entity detection, chunking, parsing, semantic role labeling, sentiment analysis, Arabic dialect modeling, etc.
- Applications: machine translation, speech recognition, speech synthesis, optical character recognition, pedagogy, assistive technologies, social media, etc.
- Resources: dictionaries, annotated data, corpus, etc.

Submissions may include work in progress as well as finished work. Submissions must have a clear focus on specific issues pertaining to the Arabic language whether it is standard Arabic, dialectal, or mixed. Papers on other languages sharing problems faced by Arabic NLP researchers such as Semitic languages or languages using Arabic script are welcome. Additionally, papers on efforts using Arabic resources but targeting other languages are also welcome. Descriptions of commercial systems are welcome, but authors should be willing to discuss the details of their work.

Paper Submission Instructions

Paper Length: Submissions are expected to be up to **8 pages** long plus any number of pages for references. Final versions of long papers will be given one additional page of content (**up to 9 pages**) so that reviewers' comments can be taken into account.

Submission Format: Submissions must be in PDF and prepared using LaTeX. The format must conform to the [official style guidelines for EACL 2017](#). **The only accepted format is the [EACL-2017 LaTeX template](#) and no other format will be accepted** such as Microsoft Word or Open Office etc...

Submission Website: <https://www.softconf.com/eacl2017/ArabicNLP/>

Blind Reviewing Policy: The workshop follows a blind reviewing policy. The authors should omit their names and affiliations from the paper and avoid self-references that reveal their identity. Papers that do not conform to these requirements will be rejected without review.

Multiple Submission Policy: Papers that have been or will be submitted to other meetings or publications must indicate this at submission time. Authors must inform organizers immediately once a paper is to be withdrawn from the workshop for any reason. Attempting to publish the same paper or with a large overlap (50%) may lead to rejection of the paper even after an acceptance notification have gone out.

Workshop Schedule (The workshop is a **one-day event**)

9:00 - 9:10	Opening remarks (Nizar Habash)
9:10 - 10:00	Keynote (Stephan Vogel)
10:00 - 11:00	Session 1 : Enabling Technologies Session
	10:00 - 10:20 Identification of Languages in Algerian Arabic Multilingual Documents (Wafia Adouane and Simon Dobnik)
	10:20 - 10:40 Arabic Diacritization: Stats, Rules, and Hacks (Kareem Darwish, Hamdy Mubarak and Ahmed Abdelali)
	10:40 - 11:00 Semantic Similarity of Arabic Sentences with Word Embeddings (El Moatez Billah NAGOUDI and Didier Schwab)
11:00 - 11:30	Coffee Break
11:30 - 12:50	Session 2 : Dialects Session
	11:30 - 11:50 Morphological Analysis for the Maltese Language: The challenges of a hybrid system (Claudia Borg and Albert Gatt)
	11:50 - 12:10 A Morphological Analyzer for Gulf Arabic Verbs (Salam Khalifa, Sara Hassan and Nizar Habash)
	12:10 - 12:30 A Neural Architecture for Dialectal Arabic Segmentation (Younes Samih, Mohammed Attia, Mohamed Eldesouki, Ahmed Abdelali, Hamdy Mubarak, Laura Kallmeyer and Kareem Darwish)
12:30 - 12:50 Sentiment analysis of Tunisian dialects: Linguistic Ressources and Experiments (Salima Medhaffar, Fethi Bougares, Yannick Estève and Lamia Hadrich-Belguith)	
12:50 - 14:30	Lunch
14:30 - 15:30	Session 3 : Applications Session
	14:30 - 14:50 CAT: Credibility Analysis of Arabic Content on Twitter (Rim EL Ballouli, Wassim El-Hajj, Ahmad Ghandour, Shady Elbassuoni, Hazem Hajj and Khaled Shaban)
	14:50 - 15:10 A New Error Annotation for Dyslexic texts in Arabic (Maha Alamri and William J Teahan)

	<table border="1"> <tr> <td data-bbox="391 63 505 165">15:10 - 15:30</td> <td data-bbox="505 63 1442 165">An Unsupervised Speaker Clustering Technique based on SOM and I-vectors for Speech Recognition Systems (Hany Ahmed, Mohamed Elaraby, Abdullah M. Mousa, Mostafa Elhosiny, Sherif Abdou and Mohsen Rashwan)</td> </tr> </table>	15:10 - 15:30	An Unsupervised Speaker Clustering Technique based on SOM and I-vectors for Speech Recognition Systems (Hany Ahmed, Mohamed Elaraby, Abdullah M. Mousa, Mostafa Elhosiny, Sherif Abdou and Mohsen Rashwan)
15:10 - 15:30	An Unsupervised Speaker Clustering Technique based on SOM and I-vectors for Speech Recognition Systems (Hany Ahmed, Mohamed Elaraby, Abdullah M. Mousa, Mostafa Elhosiny, Sherif Abdou and Mohsen Rashwan)		
15:30 - 16:00	Poster Boaster (~2 min per poster)		
16:00 - 16:30	Coffee Break		
16:30 - 18:00	<p>Poster Session</p> <ul style="list-style-type: none"> • SHAKKIL: An Automatic Diacritization System for Modern Standard Arabic Texts (Amany Fashwan, Sameh Alansary) • Arabic Tweets Treebanking and Parsing: A Bootstrapping Approach (Fahad Albogamy, Allan Ramsay, Hanady Ahmed) • Identifying Effective Translations for Cross-lingual Arabic-to-English User-generated Speech Search (Ahmed Khwileh, Haithem Afli, Gareth Jones, Andy Way) • A Characterization Study of Arabic Twitter Data with a Benchmarking for State-of-the-Art Opinion Mining Models (Ramy Baly, Gilbert Badaro, Georges El-Khoury, Rawan Moukalled, Rita Aoun, Hazem Hajj, Wassim El-Hajj, Nizar Habash and Khaled Shaban) • Robust Dictionary Lookup in Multiple Noisy Orthographies (Lingliang Zhang, Nizar Habash and Godfried Toussaint) • Arabic POS Tagging: Don't Abandon Feature Engineering Just Yet (Kareem Darwish, Hamdy Mubarak, Ahmed Abdelali and Mohamed Eldesouki) • Toward a Web-based Speech Corpus for Algerian Dialectal Arabic Varieties (Soumia Bougrine, Aicha Chorana, Abdallah Lakhdari and Hadda Cherroun) • Not All Segments are Created Equal: Syntactically Motivated Sentiment Analysis in Lexical Space (Muhammad Abdul-Mageed) • An enhanced automatic speech recognition system for Arabic (Mohamed Amine Menacer, Odile Mella, Dominique Fohr, Denis Jouvét, David Langlois and Kamel Smaili) • Universal Dependencies for Arabic (Dima Taji, Nizar Habash and Daniel Zeman) • A Layered Language Model based Hybrid Approach to Automatic Full Diacritization of Arabic (Mohamed Al-Badrashiny, Abdelati Hawwari and Mona Diab) • Arabic Textual Entailment with Word Embeddings (Nada Almarwani and Mona Diab) <p><i>The dimensions of the poster board are 90 cm wide and 120 cm high (view sample image here).</i> <i>Please allocate the top of the poster for the title and authors' names and affiliations as stated in the submitted abstract.</i> <i>The text and illustrations should be bold and large enough to read from a distance of two meters (six feet).</i></p>		

Invited Speaker

The workshop will feature a keynote talk by [Stephan Vogel](#) from [Qatar Computing Research Institute](#) (QCRI).

Workshop Organizers

General Chair	Nizar Habash
Program Chairs	Mona Diab, Kareem Darwish, Wassim El-Hajj, Hend Alkhalifa, and Houda Bouamor
Publication Chairs	Nadi Tomeh and Mahmoud El-Haj
Publicity Chairs	Fethi Bougares and Wajdi Zaghouani

- **Nizar Habash**, New York University Abu Dhabi **Email:** nizar.habash@nyu.edu
- **Mona Diab**, The George Washington University **Email:** mtdiab@email.gwu.edu
- **Kareem Darwish**, Qatar Computing Research Institute **Email:** kdarwish@qf.org.qa
- **Wassim El-Hajj**, American University of Beirut, Lebanon **Email:** wassim.el-hajj@aub.edu.lb
- **Hend Al-Khalifa**, King Saud University **Email:** hendk@ksu.edu.sa
- **Houda Bouamor**, Carnegie Mellon University in Qatar **Email:** hbouamor@cmu.edu
- **Nadi Tomeh**, LIPN, Université Paris 13, Sorbonne Paris Cité **Email:** nadi.tomeh@lipn.univ-paris13.fr
- **Mahmoud El-Haj**, Lancaster University **Email:** m.el-haj@lancaster.ac.uk
- **Fethi Bougares**, University of Le Mans **Email:** fethi.bougares@univ-lemans.fr
- **Wajdi Zaghouani**, Carnegie Mellon University-Qatar **Email:** wajdiz@cmu.edu

Program Committee Members

- Ahmed Abdelali, Qatar Computing Research Institute, Qatar
- Nora Al-Twairesh, King Saud University, Saudi Arabia
- Areeb Alowskiheq, Imam University, KSA
- Salha Alzahrani, Taif University, Saudi Arabia
- Almoataz B. Al-Said, Cairo University, Egypt
- Alberto Barrón-Cedeño, Qatar Computing Research Institute, Qatar
- Fethi Bougares, Le Mans University, France
- Tim Buckwalter, University of Maryland, USA
- Violetta Cavalli-Sforza, Al Akhawayn University, Morocco
- Abeer Dayel, King Saud University, Saudi Arabia
- Tamer Elsayed, Qatar University, Qatar
- Ossama Emam, IBM, USA
- Ramy Eskander, Columbia University, USA
- Nizar Habash, New York University Abu Dhabi, UAE
- Bassam Haddad, University of Petra, Jordan
- Hazem Hajj, American University of Beirut, Lebanon
- Maha Jarallah Althobaiti, Taif University, Saudi Arabia
- Azzeddine Mazroui, University Mohamed I, Morocco
- Karine Megerdooian, The MITRE Corporation, USA
- Ghassan Mourad, Université Libanaise, Lebanon
- Hamdy Mubarak, Qatar Computing Research Institute, Qatar
- Preslav Nakov, Qatar Computing Research Institute, Qatar

- Alexis Nasr, University of Marseille, France
- Kemal Oflazer, Carnegie Mellon University Qatar, Qatar
- Eshrag Refaee, Jazan University, Saudi Arabia
- Mohammad Salameh, Carnegie Mellon University, Qatar
- Hassan Sawaf, eBay Inc., USA
- Khaled Shaalan, The British University in Dubai, UAE
- Khaled Shaban, Qatar University, Qatar
- Otakar Smrž, Džám-e Džam Language Institute, Czech Republic
- Nadi Tomeh, University Paris 13, France
- Wajdi Zaghouani, Carnegie Mellon University, Qatar
- Imed Zitouni, Microsoft Research, USA

The WANLP adheres to the ACL anti-harassment policy. More details can be found in the following link:
https://www.aclweb.org/adminwiki/index.php?title=Anti-Harassment_Policy