

Processing Dialectal Arabic: Exploiting Variability and Similarity to Overcome Challenges and Discover Opportunities (invited talk)

Mona Diab

Department of Computer Science
George Washington University
mtdiab@email.gwu.edu

Abstract

We recently witnessed an exponential growth in dialectal Arabic usage in both textual data and speech recordings especially in social media. Processing such media is of great utility for all kinds of applications ranging from information extraction to social media analytics for political and commercial purposes to building decision support systems. Compared to other languages, Arabic, especially the informal variety, poses a significant challenge to natural language processing algorithms since it comprises multiple dialects, linguistic code switching, and a lack of standardized orthographies, to top its relatively complex morphology. Inherently, the problem of processing Arabic in the context of social media is the problem of how to handle resource poor languages. In this talk I will go over some of our insights to some of these problems and show how there is a silver lining where we can generalize some of our solutions to other low resource language contexts.

Biography

Mona Diab is an Associate Professor in the Department of Computer Science, George Washington University (GW). She is the founder and Director of the GW NLP lab CARE4Lang. Before joining GW, She was Research Scientist (Principal Investigator) at the Center for Computational Learning Systems (CCLS), Columbia University in New York. She is also co-founder of the CADIM group with Nizar Habash and Owen Rambow, which is one of the leading reference points on computational processing of Arabic and its dialects. Her research interests span several areas in computational linguistics/natural language processing: cross linguistic modeling and multilingual processing, computational lexical semantics, social media processing, information extraction & text analytics, machine translation, resource building, and computational socio-pragmatics. She has a special interest in low resource language processing with a focus on Arabic dialects.