Kanan, Tarek; Ayoub, Souleiman; Al-Dahoud, Ali; Kanaan, Ghassan; Fox, Edward. “Extracting Named Entities Using Named Entity Recognizer for Arabic News Articles”. Proceeding of the International Computer Science and Informatics Conference (ICSIC 2016). Amman Arab University, Amman – Jordan. January 2016.

Abstract:

This paper describes how to extract, for the Arabic language, named entities and topics from news articles. Indeed, there is a lack of high quality tools for Named Entity Recognition (NER) for Arabic; therefore the authors have built an Arabic NER (RenA). NER involves extracting information and identifying types, such as name, organization, and location. For English language there are effective tools for NER, however these are not directly applicable to Arabic language. As a result, a new method and tool (i.e., RenA) have been developed. For NER evaluation purposes a baseline corpus was built for assessment and comparison with other methods and tools, with help from volunteer graduate students who understand Arabic. RenA produces good results, with accurate Name, Organization, and Location extraction from news articles collected from online resources. A comparison between the RenA results with a popular Arabic NER resulted in a noticeable enhancement.