

Saudi Twitter Corpus for Sentiment Analysis

Authors : Adel Assiri, Ahmed Emam, Hmood Al-Dossari

Abstract : Sentiment analysis (SA) has received growing attention in Arabic language research. However, few studies have yet to directly apply SA to Arabic due to lack of a publicly available dataset for this language. This paper partially bridges this gap due to its focus on one of the Arabic dialects which is the Saudi dialect. This paper presents annotated data set of 4700 for Saudi dialect sentiment analysis with (K= 0.807). Our next work is to extend this corpus and creation a large-scale lexicon for Saudi dialect from the corpus.

Keywords : Arabic, sentiment analysis, Twitter, annotation

Conference Title : ICLRE 2016 : International Conference on Language Resources and Evaluation

Conference Location : Kuala Lumpur, Malaysia

Conference Dates : February 11-12, 2016