

Jordan University of Science and Technology

Human Annotated Arabic Dataset of Book Reviews for Aspect Based Sentiment Analysis

Authors: Mohammad AL-Smadi, Omar Al-Qawasmeh, Bashar Talafha and Muhannad Quwaider

Abstract: With the prominent advances in Web interaction and the enormous growth in user-generated content, sentiment analysis has gained more interest in commercial and academic purposes. Recently, sentiment analysis of Arabic user-generated content is increasingly viewed as an important research field. However, the majority of available approaches target the overall polarity of the text. To the best of our knowledge, there is no available research on aspect-based sentiment analysis (ABSA) of Arabic text. This can be explained due to the lack of publically available datasets prepared for ABSA, and to the slow progress in sentiment analysis of Arabic text research in general. This paper fosters the domain of Arabic ABSA, and provides a benchmark human annotated Arabic dataset (HAAD). HAAD consists of books reviews in Arabic which have been annotated by humans with aspect terms and their polarities. Nevertheless, the paper reports a baseline results and a common evaluation technique to facilitate future evaluation of research and methods.