

Jordan University of Science and Technology

Arabic Named Entity Disambiguation Using Linked Open Data

Authors: Omar Al-Qawasmeh, Mohammad AL-Smadi, Nisreen Fraihat

Abstract: This research aims at tackling the problem of Arabic Named-Entity Disambiguation (ANED) through an enhanced approach of information extraction from Arabic Wikipedia and Linked Open Data (LOD). The approach uses query label expansion and text similarity techniques to disambiguate entities of the types: person, location, and organization. A reference dataset for ANED has been prepared and annotated with over 10K entity mentions. The reference dataset was used in evaluating the proposed ANED approach. Results show that the accuracy of ANED approach is 84% on the overall Dataset. Moreover, the proposed approach was capable to disambiguate location entities with accuracy of 94%, person entities with 76%, and organization entities with 78%.