

Abstract

Ingy A Sarhan

Semi-Supervised Pattern Based Algorithm for Arabic Relation Extraction

While several relation extraction algorithms have been developed in the past decade, mainly in the English language, only few researchers target the Arabic language owing to its complexity and rich morphology. This paper proposes a semi-supervised pattern-based bootstrapping technique to extract Arabic semantic relation that lies between entities. In order to enhance the performance to suit the morphologically rich Arabic language, stemming, semantic expansion using synonyms, and an automatic scoring technique to measure the reliability of the generated patterns and extracted relations were used. To further improve performance, a dependency parser was then used to omit negative relations. The proposed system was tested by applying it to two corpora, which differ in both size and genre, scoring a highest F-measure of 75.06%. Furthermore, the effect of adding stemming and synonyms was also experimentally tested. The results show that this bootstrapping methodology achieves higher performance than existing state-of-the-art methods, and can be expanded to include more relations for use in various NLP tasks.