

A Supervised Approach for Word Sense Disambiguation Based on Arabic Diacritics

Authors : Alaa Alrakaf, Sk. Md. Mizanur Rahman

Abstract : Since the last two decades' Arabic natural language processing (ANLP) has become increasingly much more important. One of the key issues related to ANLP is ambiguity. In Arabic language different pronunciation of one word may have a different meaning. Furthermore, ambiguity also has an impact on the effectiveness and efficiency of Machine Translation (MT). The issue of ambiguity has limited the usefulness and accuracy of the translation from Arabic to English. The lack of Arabic resources makes ambiguity problem more complicated. Additionally, the orthographic level of representation cannot specify the exact meaning of the word. This paper looked at the diacritics of Arabic language and used them to disambiguate a word. The proposed approach of word sense disambiguation used Diacritizer application to Diacritize Arabic text then found the most accurate sense of an ambiguous word using Naïve Bayes Classifier. Our Experimental study proves that using Arabic Diacritics with Naïve Bayes Classifier enhances the accuracy of choosing the appropriate sense by 23% and also decreases the ambiguity in machine translation.

Keywords : Arabic natural language processing, machine learning, machine translation, Naive bayes classifier, word sense disambiguation

Conference Title : ICSRD 2020 : International Conference on Scientific Research and Development

Conference Location : Chicago, United States

Conference Dates : December 12-13, 2020