Arabic Light Stemmer for Better Search Accuracy

Authors : Sahar Khedr, Dina Sayed, Ayman Hanafy

Abstract : Arabic is one of the most ancient and critical languages in the world. It has over than 250 million Arabic native speakers and more than twenty countries having Arabic as one of its official languages. In the past decade, we have witnessed a rapid evolution in smart devices, social network and technology sector which led to the need to provide tools and libraries that properly tackle the Arabic language in different domains. Stemming is one of the most crucial linguistic fundamentals. It is used in many applications especially in information extraction and text mining fields. The motivation behind this work is to enhance the Arabic light stemmer to serve the data mining industry and leverage it in an open source community. The presented implementation works on enhancing the Arabic light stemmer by utilizing and enhancing an algorithm that provides an extension for a new set of rules and patterns accompanied by adjusted procedure. This study has proven a significant enhancement for better search accuracy with an average 10% improvement in comparison with previous works.

Keywords : Arabic data mining, Arabic Information extraction, Arabic Light stemmer, Arabic stemmer

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

Conference Location : Chicago, United States

Conference Dates : December 12-13, 2020