Abstract

Nahla A Belal

Enhancing Root Extractors Using Light Stemmers

The rise of Natural Language Processing (NLP) opened new possibilities for various applications that were not applicable before. A morphological-rich language such as Arabic introduces a set of features, such as roots, that would assist the progress of NLP. Many tools were developed to capture the process of root extraction (stemming). Stemmers have improved many NLP tasks without explicit knowledge about its stemming accuracy. In this paper, a study is conducted to evaluate various Arabic stemmers. The study is done as a series of comparisons using a manually annotated dataset, which shows the efficiency of Arabic stemmers, and points out potential improvements to existing stemmers. The paper also presents enhanced root extractors by using light stemmers as a preprocessing phase.