Abstract

Stemming is considered as a pre-processing step in many applications: text mining, information retrieval, machine translation etc. The Arabic language has many special cases or properties that affect stemming or any automatic method, it depends on both inflectional and derivational morphology to produce the various forms of the language words. Many researchers have proposed algorithms to solve the problems of stemming. This paper aims to make a comparison study among the existing Arabic stemmers, the comparison study is based on the methodologies, the usage, main idea, algorithm, the affixes, limitations, output, and the stemmers’ sensitivity for both diacritics and context.

References


Conference on Statistical Analysis of Textual Data.


Index Terms

Computer Science Pattern Recognition

Keywords

Arabic Stemmers, Arabic Morphological Analyzer.