Abstract

The objective of this research is to produce a lemmatization algorithm for Indonesian language, according to the official definition of lemmatization. Lemmatization process will transform derivational and inflectional words into their base dictionary entry form. Our research method is literature review for various lemmatization and stemming algorithms, and analysis of stemming algorithms for Indonesian language. We are finding the state of the art in this topic and develop a lemmatization algorithm based on the state of the art. The results achieved are a lemmatization algorithm with 98% accuracy. The conclusions are that we are able to develop a dictionary-based and rule-based lemmatization algorithm for Indonesian language and it reaches a moderately high accuracy. We hope this research result can help the development of all aspects concerning the integration of linguistics and technology, especially in Indonesian language.

Keywords: Lemmatization, Linguistics, Indonesian Language