

ABSTRACT

Syllabification is a process from word to a series of syllable. Syllabification can be used in Natural Language Processing (NLP) such as speech recognition, text-to-speech, rhyme detection, and many more. Syllabification of the Indonesian language will refer to the General Guidelines for Indonesian Language Spelling (PUEBI). The corpus for this research contains the main word and the syllable. The corpus for this research is using “*50k KBBI 5-k fold*” and combined with “*103k Named Entity 5-k fold*”. The evaluation for this model is using Word Error Rate (WER). WER of previous deep learning model for syllabification is still high with 3.75% WER. The objective of this research is to lower the WER for deep learning using Transformer and Syllable Tagging because it can accept long contextual dependency. The evaluation result of this model is 3.68% WER and can be used universally for Indonesian words because the margin between Formal Words and Named Entity Words is close with the average result. Thus, this model currently the better model for Indonesian syllabification deep learning model according from the average WER is lower than the other deep learning model.

Index Terms: syllabification, indonesian language, deep learning, transformer