

ABSTRACT

CORRECTION OF TYPICAL ERRORS IN INDONESIAN LANGUAGE USING MACHINE LEARNING

Oleh
Alma Alfiatul Inayah
17102050

Spelling errors which are caused by typing errors are often found in writing. Typing errors cause a word to have ambiguous meaning or not in accordance with the KBBI. Errors in writing affect the process of creating a paper, it needs to be revised because the writing doesn't follow the EYD (Enhanced Spelling Bahasa Indonesia). Many studies have produced an automatic correction system that can correct typing errors. Levenshtein Distance, Soundex, Recurrent Neural Network, FastText are examples of algorithms used to correct typing errors. The difference between the algorithm lies in how it works. The Levenshtein Distance works by calculating the shortest distance between two strings. Soundex corrects typos based on the sound. Recurrent Neural Network will correct typos in predicting misspelled words. FastText will correct the wrong word based on the similarity of the vector. Based on this, a study was conducted to compare the accuracy and the processing speed of the four algorithms mentioned in correcting typing errors according to the context of the sentence. As a test material, there are 2 datasets used to build the model, which are 36,517 basic words from KBBI and a collection of corpora novels. Based on testing using a sample of 100 sentences and 1000 sentences The algorithm with the best accuracy rate is generated on the FastText model of 44% for 100 sentences and 35% for 1000 sentences, with a processing speed of 0.00771513 seconds per sentence.

Keywords: *Typing Error Detection, Typing Correction, Levenshtein Distance, Soundex, Recurrent Neural Network, Fasttext.*