Abstract

The need of word processor that have spell correction fasility in Bahasa Indonesia is very important. If an electronic dictionary is available that covers the possible input words, the choice of dictionary representation and the technique how to find good corrections candidates for a garbled input word is very important. In this final project, we use an acyclic deterministic finite automata (ADFA) as an effective representation of the lexicon that allows fast checking for the existence of a word in the lexicon and the Levenshtein-automata technique to find good corrections candidates. A Levenshtein-automaton represents an acyclic DFA, which encodes the set of all string, whose edit distance from the input string does not exceed the given bound k (levenshtein-distance) [11]. Because of the morphological structure of Bahasa Indonesia words is complicated then if a word is absent from the lexicon, the stemming is needed. Stemming is a process to find base/common word from word that have preffix and suffix. The testing result shows that the accuration of stemming function for checking texts is 96,75768 % and the validity of candidate corrections is 82,37623 %.

Keywords: Spell correction, Levenshtein-automata, Levenshtein-distance, ADFA, stemming.