GRAMMAR BASED CORRECTION OF SENTENCES USING MUTUALLY EXCLUSIVE HYPOTHESES GRAPH

M.V.Davydov

Information Systems and Networks Department, Institute of Computer Science and Information Technology, Lviv Polytechnic National University, S. Bandery Str., 12, Lviv, 79013, UKRAINE

The main goal of this article is to study possibility of word correction in Ukrainian sentences based on grammar and context knowledge. The problem of automatic word correction is a common problem for speech recognition, OCR applications and word processors. Known approaches for word correction are based primarily on word input model and dictionary. However the benefit from using context and sentence grammar is not completely studied for this purpose.

A context model based on word form repeat rate and grammar model based on dependency grammar is utilized to imply correct word correction from a set of dictionary based suggestions. The idea of using word context is based on assumption that words repeat in text fragment in the same of other word form. In order to verify this assumption an experiment with legislation, scientific and fiction texts was conducted. The experiment shows that probability of word from repetition in 1000 words fragment is near 80% for legislation, 75% for scientific and 60% for fiction texts. The use of context model based on word repetition rate increases automatic word correction rate from 14% up to 32% for ambiguous corrections.

The use of dependency grammar model for word correction requires processing of sentences induced by word correction suggestions from dictionary. The plurality of word suggestions greatly increases number of grammar parsing variants. Mutually exclusive hypothesis graph is used to combine all word grammar cases into one graph that is used to find the most appropriate word correction that increases probability of correct sentence grammar model. The use of grammar model leads to further increase in automatic word correction rate from 32% to 37% with use of context model and from 14% to 22% without.

The increase in word correction rate in conducted experiments shows prospects of using context and grammar models for automatic word correction. However the used context and grammar models require further improvements to incorporate information about widely used word structures. Knowledge mining from the text and subject area can also help to increase correction rate.

Keywords – natural language processing, dependency grammars, automatic error correction.