

The Problem of Pos Tagging and Stemming for Agglutinative Languages



B.B.Elov, Eşref Adalı, Sh.Khamroeva, O.Abdullayeva, Z.Xusainova, N.Xudayberganov

Issue 8 th International Conference on Computer Science and Engineering UBMK 2023,
Mehmet Akif Ersoy University, Burdur – Turkey

Article Number 10286792

Number of page(s) 12

Section Information technology

DOI <https://doi.org/10.1109/UBMK59864.2023.10286792>

Published online 24 October 2023

Abstract

The number of possible word forms in agglutinative languages is theoretically unlimited. This, in turn, creates the problem of POS tagging (part-of-speech) of out-of-vocabulary (OOV) words in agglutinative languages. In agglutinative languages, words are formed by adding suffixes to the stem. Due to the occurrence of phonetic harmony and disharmony while adding suffixes to the stem, it is necessary to analyze both phonetic and morphological changes. When solving many NLP tasks, it is necessary to reduce word forms to the stem (stemming). Removing all inflectional affixes from a word and lemmatizing the rest of the word is considered one of the important tasks of natural language processing (NLP), and this process is called stemming. The stemming process is important in information retrieval (IR) systems.

<https://doi.org/10.1109/UBMK59864.2023.10286792>