

5th International Conference on **ARTIFICIAL INTELLIGENCE**
&
5th International Conference on **AUTOMATION & ROBOTICS**
April 16-17, 2018 | Las Vegas, USA

Unsupervised part-of-speech tagger for Amharic language

Zelalem Fantahun
iCog-Labs, Ethiopia

This paper presents the development of Amharic part-of-speech (POS) tagger using machine learning algorithm. Since Amharic is one of the under-resourced languages, the availability of tagged corpus is the train bottleneck problem for natural language processing, such as POS tagging. The promising direction to tackle this problem is a learning model that does not require annotated dataset. In this study, k-means clustering algorithm is used to group words that have similar features together. This paper also identifies morphological and positional word features that can distinguish the words into different clusters. The experimental result shows that the prototype achieves 81% accuracy.

zelalemfantahun@gmail.com