Analyzing Tagging Accuracy of Part-of-speech Taggers

Nyein Pyae Pyae Khin
nyeinpyaepyaekhin@gmail.com
University of Computer Studies, Mandalay, UCSM

Than Nwe Aung
mdytna@gmail.com
University of Computer Studies, Mandalay, UCSM

Automated part-of-speech (POS) tagging has been a very active research area for many years and is the foundation of natural language processing systems. Natural Language Toolkit (NLTK) library in the Python environment provides the necessary tools for tagging, but doesn't actually tell us what methods work the best. Therefore, this work analyzes the performance of part-of-speech taggers, namely the NLTK Default tagger, Regex tagger and N-gram taggers (Unigram, Bigram and Trigram) on a particular corpus. The corpora we have used for the analysis are; Brown, Penn Treebank and CoNLL2000. We have applied all taggers to these three corpora, resultantly we have shown that whereas Unigram tagger does the best tagging in all corpora, the combination of taggers does better if it is correctly ordered.