

## Part of speech tagging for Twitter conversations using Conditional Random Fields model

A. R. F. S. Fanoon  
Postgraduate Institute of Science,  
University of Peradeniya, Sri Lanka  
fanoonarfs@gmail.com

G. A. I. Uwanthika  
Postgraduate Institute of Science,  
University of Peradeniya, Sri Lanka  
isuriuwanthika@gmail.com

### Abstract

*Part-of-Speech Tagging is the technology of assigning the appropriate parts-of-speech to a word. Part-of-speech tagging is very useful in information retrieval, information extraction, and speech processing. This research presents a part-of-speech tagging, especially for twitter text data. The process of part-of-speech tagging for twitter conversation is a difficult task. Several approaches have been made to develop an accurate tagging system but most of them are relevant to news text data and web contents. Therefore, this research intends to develop a part-of-speech tagger model for twitter speech. using CRF toolkit. The system was developed for nearly 1000 twitter conversations employing Conditional Random Field stochastic model. The data for twitter speech was downloaded from the internet. A POS-tagged text corpus, template file and CoNLL file for both training and testing database were prepared accordingly. The training was carried out for both the unigram model as well as the bigram model. The performance of the system over these models was obtained through this examination which showed a significant efficiency, calculated from the number of correctly tagged words and the total number of words.*

**Keywords:** *Bigram model, Conditional Random Field, Tagging, Unigram model*