

## **Automated Text Annotation for Social Media Data during Natural Disasters**

**Si Si Mar Win**  
**sisimarwin@gmail.com**

**University of Computer Studies, Mandalay, UCSM**

**Than Nwe Aung**  
**mdytna@gmail.com**

**University of Computer Studies, Mandalay, UCSM**

Nowadays, text annotation plays an important role within real-time social media mining. Social media analysis provides actionable information to its users in times of natural disasters. This paper presents an approach to a real-time two layer text annotation system for social media stream to the domain of natural disasters. The proposed system annotates raw tweets from Twitter into two types such as Informative or Not Informative as first layer. And then it annotates again five information types based on Informative tweets only as second layer. Based on the first and second layer annotation results, this system provides the tweets with user desired informative type in real time. In this system, annotation is done at tweet level by using word and phrase level features with LibLinear classifier. All features are in the form of Ngram nature based on part of speech (POS) tag, Sentiment Lexicon and especially created Disaster Lexicon. The validation of this system is performed based on different disaster related datasets and new Myanmar\_Earthquake\_2016 dataset derived from Twitter. The annotated datasets generated from this work can also be used by interested research communities to study the social media natural disaster related research.