FILIET: An Information Extraction System

For Filipino Disaster-Related Tweets

Ralph Vincent J. Regalado, Kyle Mc Hale B. Dela Cruz, John Paul F. Garcia, Kristine Ma. Dominique F. Kalaw, and Vilson E. Lu

Center for Language Technologies

De La Salle University, Manila

ralph.regalado@delasalle.ph, {kyle\_dela\_cruz, john\_paul\_garcia, kristine\_kalaw, vilson\_lu}@dlsu.edu.ph

**Abstract**: The Philippines is considered the social media capital of the world, and the role of social media has become even more pronounced in the country during disasters. Twitter is among the many forms of social media. As experienced, information and data shared through Twitter have helped individuals, institutions, and organizations (government, public, and private) during emergency response, in making decisions, conducting relief efforts, and practically mobilizing people to humanitarian causes. However, extracting the most relevant information from Twitter is a challenge because natural languages do not have a particular structure immediately useful when programming. Another problem that information extraction faces is that some languages, like Filipino, is morphologically rich, making it even more difficult to extract information. Therefore, the goal of this research is to create the Filipino Information Extraction Tool for Twitter (FILIET), a system that extracts relevant information from Filipino disaster-related tweets. The system consists of several modules but the experiments outlined in this paper focuses on that of the Category Classifier module where the tweets are classified into either of the four categories – Caution and Advice (CA), Casualty and Damage (CD), Donations (D), and Others (O). The experiments are to test which is the best algorithm to be used for classifying the tweets. To improve the results of the tests, it is important to increase the instances of the corpus.

**Key Words**: information extraction; disaster management; Twitter