**Abstract**

Cyberbullying refers to the use of technology to humiliate and slander behaviour towards other people with the use of computers or mobile phones. It takes form of hate messages sent through social media and e-mails. Recently it has become a serious social problem in many countries. Since adolescents are more inclined to social media than any other age groups, they are proven to be more vulnerable to cyberbullying attacks that is why there is a need for intelligent systems to identify the risks of cyberbullying incidents automatically. Therefore, we propose a method to detect cyberbullying entries on social media. We present the construction and annotation of a corpus of Filipino social media posts gathered from Facebook, Twitter and AskFM. The data that we obtained were annotated with fine-grained cyberbullying-related text categories, such as insults, threats, sexual talks, defense and curse. We also defined the roles in a cyberbullying contexts such as the harasser, victim and bystander. Apart from describing our dataset construction and annotation, we present proof-of-concept experiments on the automatic identification of cyberbullying events and fine-grained cyberbullying categories.

Keywords: Cyberbullying, Detection, Implications, Social Media