World Academy of Science, Engineering and Technology
International Journal of Psychological and Behavioral Sciences
Vol:14, No:10, 2020

Spanish Language Violence Corpus: An Analysis of Offensive Language in Twitter

Authors : Beatriz Botella-Gil, Patricio Martínez-Barco, Lea Canales

Abstract : The Internet and ICT are an integral element of and omnipresent in our daily lives. Technologies have changed the way we see the world and relate to it. The number of companies in the ICT sector is increasing every year, and there has also been an increase in the work that occurs online, from sending e-mails to the way companies promote themselves. In social life, ICT's have gained momentum. Social networks are useful for keeping in contact with family or friends that live far away. This change in how we manage our relationships using electronic devices and social media has been experienced differently depending on the age of the person. According to currently available data, people are increasingly connected to social media and other forms of online communication. Therefore, it is no surprise that violent content has also made its way to digital media. One of the important reasons for this is the anonymity provided by social media, which causes a sense of impunity in the victim. Moreover, it is not uncommon to find derogatory comments, attacking a person's physical appearance, hobbies, or beliefs. This is why it is necessary to develop artificial intelligence tools that allow us to keep track of violent comments that relate to violent events so that this type of violent online behavior can be deterred. The objective of our research is to create a guide for detecting and recording violent messages. Our annotation guide begins with a study on the problem of violent messages. First, we consider the characteristics that a message should contain for it to be categorized as violent. Second, the possibility of establishing different levels of aggressiveness. To download the corpus, we chose the social network Twitter for its ease of obtaining free messages. We chose two recent, highly visible violent cases that occurred in Spain. Both of them experienced a high degree of social media coverage and user comments. Our corpus has a total of 633 messages, manually tagged, according to the characteristics we considered important, such as, for example, the verbs used, the presence of exclamations or insults, and the presence of negations. We consider it necessary to create wordlists that are present in violent messages as indicators of violence, such as lists of negative verbs, insults, negative phrases. As a final step, we will use automatic learning systems to check the data obtained and the effectiveness of our guide.

Keywords : human language technologies, language modelling, offensive language detection, violent online content
Conference Title : ICAC 2020 : International Conference on Aggression and Cyberbullying
Conference Location : London, United Kingdom
Conference Dates : October 22-23, 2020

Open Science Index, Psychological and Behavioral Sciences Vol:14, No:10, 2020 publications.waset.org/abstracts/114711.pdf