Contribution ID: **81** Type: **Poster**

Mining Tweets on Sexual Violence in South Africa

Society or rather frustrated victims of Gender based violence have taken to Twitter as their most dependable social media platform where this situation is mostly reported.
The manifestation of this issue on Twitter is represented in keywords like rape, sexual abuse, sexual assault and sexual harassment. These keywords are held in a metadata tag known as "hashtag", which is verifiable through its intuitive and dynamic tagging which facilitates the simplicity of searching through specific text or content. To our knowledge, this issue has long existed both physically and digitally yet, no effective solution had been devised that provides information to help tackle this problem of analysing and visualising sexual violence tweet in real-time as well as the sentiment analysis.
An effective analysis of sexual violence related tweets can expose information that were previously unknown about the issue hence, leading to better mitigating strategies. However, its value for sentiment analysis has not thoroughly gained scholarly interest or in some instance, scantly studied – with gender related workshops and community-based awareness campaigns preferred for tackling this issue. We present a focused approach into mining Twitter data on sexual violence through its Streaming API by developing a real-time, extractive and interactive tool whose server side is built on Node in an attempt to guarantee speed with its robust modules and utility libraries. Our tool is a web-based application comprising of Elasticsearch and Kibana. Elasticsearch functions as a storage repository and search engine, while Kibana facilitates intuitive dashboard creation and real-time analysis of Elasticsearch index. The Indexer, which is an application built with Node.js, connects with the Twitter Streaming API to get tweets found in the boundaries of sexual violence in South Africa. The study investigated and established the bias of Twitter users based on a demographic indicator like Gender and Educational status. The extraction of Gender from tweet was through neural network with Google's TensorFlow and the Educational status was inferred from tweets using supervised machine learning. This variable reveals the exact people posting on the issue. Correlating this with reported patterns of sexual violence can help determine whether Twitter data is sufficiently free of demographic bias to allow it to be used for further analysis.

Presenter Biography

Primary author: Mr OYASOR, Jude (University of the Witwatersrand)

Presenter: Mr OYASOR, Jude (University of the Witwatersrand)

Session Classification: Poster session

Track Classification: HPC Techniques and Computer Science