

GEOGRAPHY AND TIMESTAMP OF SPAM TWEETS.

CSCE 5290

NATURAL LANGUAGE PROCESSING

PROJECT PROPOSAL

GROUP 7

Github source:

<https://github.com/anith462/Anis462>

GROUP MEMBER'S:

1. Anitha Nari - 11550658
2. Vijayalakshmi pepala - 11656632
3. Monisha Mahitha boddu - 11513712
4. Mounika nuchu - 11653658

Motivation:

Now a days there is lots of usage of tweets, due to increase in technology along with the not only tweets but also spam risk has been increased accordingly. To locate the source of the spam using NLP techniques such as NLTK, text cleansing, along with sentiment analysis in order to reduce the further spam risks.

Significance:

Initially spams used to be about advertising according to the user mostly used categories. but now some spammers infect computers these days in order to increase traffic to their spam websites. They link their URL to the informative tweet, even though the content of both is unrelated. Spammed tweets contain the URL and the tweet message. So, it is important to locate the spam tweets to be able to stop and control the spam.

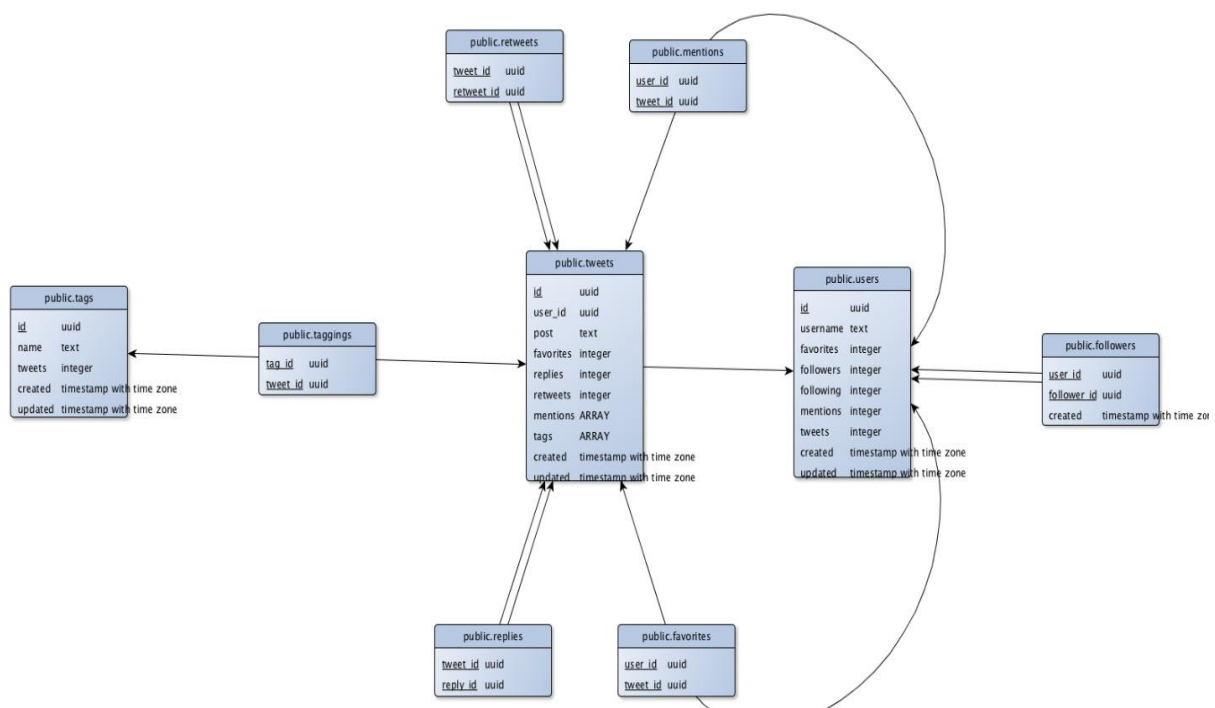


Fig 1. Tweet structure

This figure shows the structure of tweet, The tables which are used in tweets are users, tweets, favorites, followers, replies, retweets.

Objectives:

Here using NLP and its techniques we want to produce an effective result that can get the source details.

Features:

The URL of any product, commercial, or other content that has no relation to the posted tweet is typically included in spammers' mentions of their enlightening topical tweet. Drug sales and downloads are occasionally available on the URLs of malware, phishing, and other malicious software that obstruct client encounters and occasionally close down the entire network. It's important to block those and for doing so we need to find the source. Here we're using NLP and CNN techniques for determining spams among tweets and to find its location.

References:

1. <https://github.com/shuber/postgres-twitter>
2. https://www.google.com/imgres?imgurl=https%3A%2F%2Fwww.mdpi.com%2Fsymmetry%2Fsymmetry-12-01054%2Farticle_deploy%2Fhtml%2Fimages%2Fsymmetry-12-01054-g002.png&imgrefurl=https%3A%2F%2Fwww.mdpi.com%2F2073-8994%2F12%2F6%2F1054&tbnid=-w-6jB-C5VS6VM&vet=12ahUKEwui3e6X4ar9AhXVAFkFHZuUD5MQMygCegQIARAp..i&docid=hNCDeG7q65SD1M&w=2274&h=1977&q=tweets%20sorted%20by%20geography%20and%20timestamp%20project%20in%20nlp&ved=2ahUKEwui3e6X4ar9AhXVAFkFHZuUD5MQMygCegQIARAp