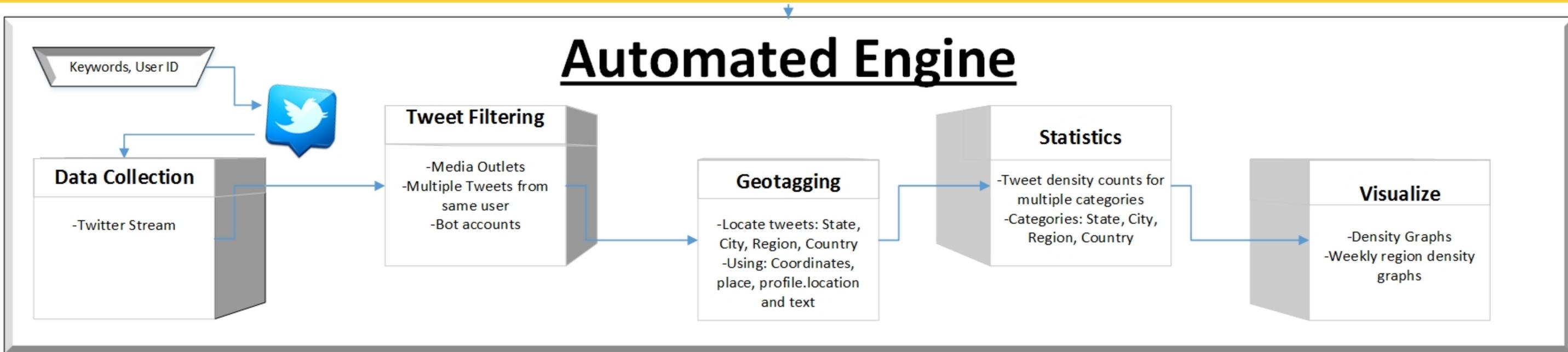


Automated Twitter Collection, Analysis & Visualization System



Purpose of Project

The purpose of this project is to mine the data that twitter provides every single day via tweets. These tweets can be analyzed using this Automated engine designed for Twitter Collection, analysis and Visualization.

Why Twitter Data?

Social media is a continuously growing and massive communication outlet for much of the population. While many see those millions of tweets every day as simply tweets, we see those tweets as data. Using a twitter API we are able to obtain tweets in real time simply using keywords. We can then analyze the data using our framework including: Filtering, Geotagging, gathering statistics and drawing conclusions based on the users parameters for the search. There is a great power within social media for analysis the automated engine we have designed will allow users to analyze trends that they pick.

References

[1] Feng Wang, Haiyan Wang, Kuai Xu, Ross Raymond, Jamie Chou, Shaun Fuller, Anton Debruyne, "Regional. Level Influenza Study with Geo-Tagged Twitter Data", Journal of Medical Systems, Vol. 40, No. 180, August 2016

How it works

As shown above, the process has 5 main steps:

1. Using a twitter stream, tweets are gathered in real time based upon search criteria.
2. The tweet files are then passed through into a filtering module based off a variety of parameters. Whether they be restricting users to a single tweet within a specified timeframe or removing all tweets from certain users. After the filtration comes the geotagging module.
3. The geotagging module takes in the filtered tweets and attempts to obtain the location of the tweet. Geotagging uses tools things such as coordinates and profile.location.
4. After using the resolvers to obtain locations for as many tweets as possible, we can then obtain multiple statistics from the data. One example would be obtaining a state count which gives us every state associated with the count of tweets from that state.
5. The last step would be to draw graphical representations for the statistics provided.

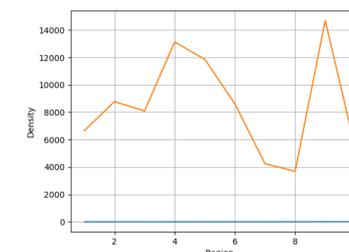
Acknowledgements

This project is supported by NSF and NCUIRE.

Applications

-Flu Influenza Tracking and Predicting: Knowing when and how much the Flu will outbreak is a very important thing. It gives people a cautious mindset and allows people to be prepared. CDC data on Flu outbreaks has quite the delay. Using this tool we can and have successfully monitored the Flu trend on twitter and have been able to predict CDC results with almost 84% accuracy over an 11 week period.

-A prime example of the power behind this engine Is who talks about Guns more? Texans or Californians. Using these tools you could obtain a state count and compare the amount to the population of the states thus giving you a comparison between the two states.



Density graph for 12/1/2016 – 12/31/2016. These regions depicts 50 states grouped into 10 clusters.

Conclusion

It is feasible to build such a tool that we described. A tool such that you can portray and predict to a certain level of accuracy patterns and trends given to us by social media outlets. The Flu Influenza tracking showed that this tool can do this and has data to back up that it did accomplish this goal.