Introduction

Twitter Analyzer is an application created to analyze twitter tweets against each other to find the most similar tweets. A user will be prompted to enter different tweets and will receive the latest tweet, this tweet is then compared to a list of up to one hundred tweets using the same search word and any previous search word the user has inputted. Using the natural language processing library the tweet is analyzed against this list to find the most similar tweet to itself and a percentage and similarity is displayed. As the number of tweets being compared to each other grows, the number of comparisons grows exponentially thus using more and more cpu. To solve this issue, Amazon Web Service(AWS) is used to setup a scalable cloud to increase when the usage is increased.

API and libraries used

Twitter API - <https://developer.twitter.com/en/docs/api-reference-index>

twitter API is primarily used to receive a large input of data, using a keyword the API will fetch the latest 100 tweets with that word, an issue occurred because of the API limit of 450 calls per 15 mins, this however is rectified by creating multiple accounts and thus increasing the amount of calls that can be made

Natural - <https://github.com/NaturalNode/natural>

Natural is a library that gives access to language processing tools, the tool used in twiter Analyzer is the JaroWinklerDistance. Jaro–Winkler string distance measuring algorithm is used to measure two strings together to see how similar the strings are to each other.

Technical description

Scaling and performance

Testing and limitations

Possible extensions

References

Appendix