|  |
| --- |
| Literature Review: Twitter Sentiment Analysis |
| CKME136 |
| Hiren Dossani |

Contents

[Introduction 2](#_Toc11331475)

[Research Papers 2](#_Toc11331476)

[Amazon Web Services (https://docs.aws.amazon.com/index.html) 2](#_Toc11331477)

[Like It or Not: A Survey of Twitter Sentiment Analysis Methods (https://dl.acm.org/citation.cfm?id=2938640) 2](#_Toc11331478)

[A real-time Twitter sentiment analysis using an unsupervised method (https://dl.acm.org/citation.cfm?id=3102282) 3](#_Toc11331479)

[Dataset 3](#_Toc11331480)

[Twitter DataSet 3](#_Toc11331481)

[Approach 4](#_Toc11331482)

[Create repository 4](#_Toc11331483)

[Create Lambda function to collect raw tweets 4](#_Toc11331484)

[Save Raw Tweets 4](#_Toc11331485)

[Trigger Lambda Function To Preprocess Data 5](#_Toc11331486)

[Process Clean Data 5](#_Toc11331487)

[Save the results 5](#_Toc11331488)

[Query Results 5](#_Toc11331489)

[Let Data Tell the Story 5](#_Toc11331490)

# Introduction

With the evolution of Social Networks (SNs) such as Twitter, millions of users can interact, share interests, activities, contents or exchange experiences and opinions. Sharing opinions is an active research topic in the framework of sentiment analysis and opinion mining.

Social media interactions between organizations and customers deepen brand awareness. These conversations are a low-cost way to acquire leads, improve website traffic, develop customer relationships, and improve customer service.

In this paper, I present a real-time implementation of a system that can discover and track opinions on Twitter using AWS Cloud technologies.

# Research Papers

## Amazon Web Services (<https://docs.aws.amazon.com/index.html>)

Amazon managed services such as Amazon Translate, Amazon S3, Amazon Athena, Amazon Lambda, Amazon EC2 and Amazon QuickSight to perform language translation and natural language processing on the tweets flowing through the system and to create data visualization.

## Like It or Not: A Survey of Twitter Sentiment Analysis Methods (<https://dl.acm.org/citation.cfm?id=2938640>)

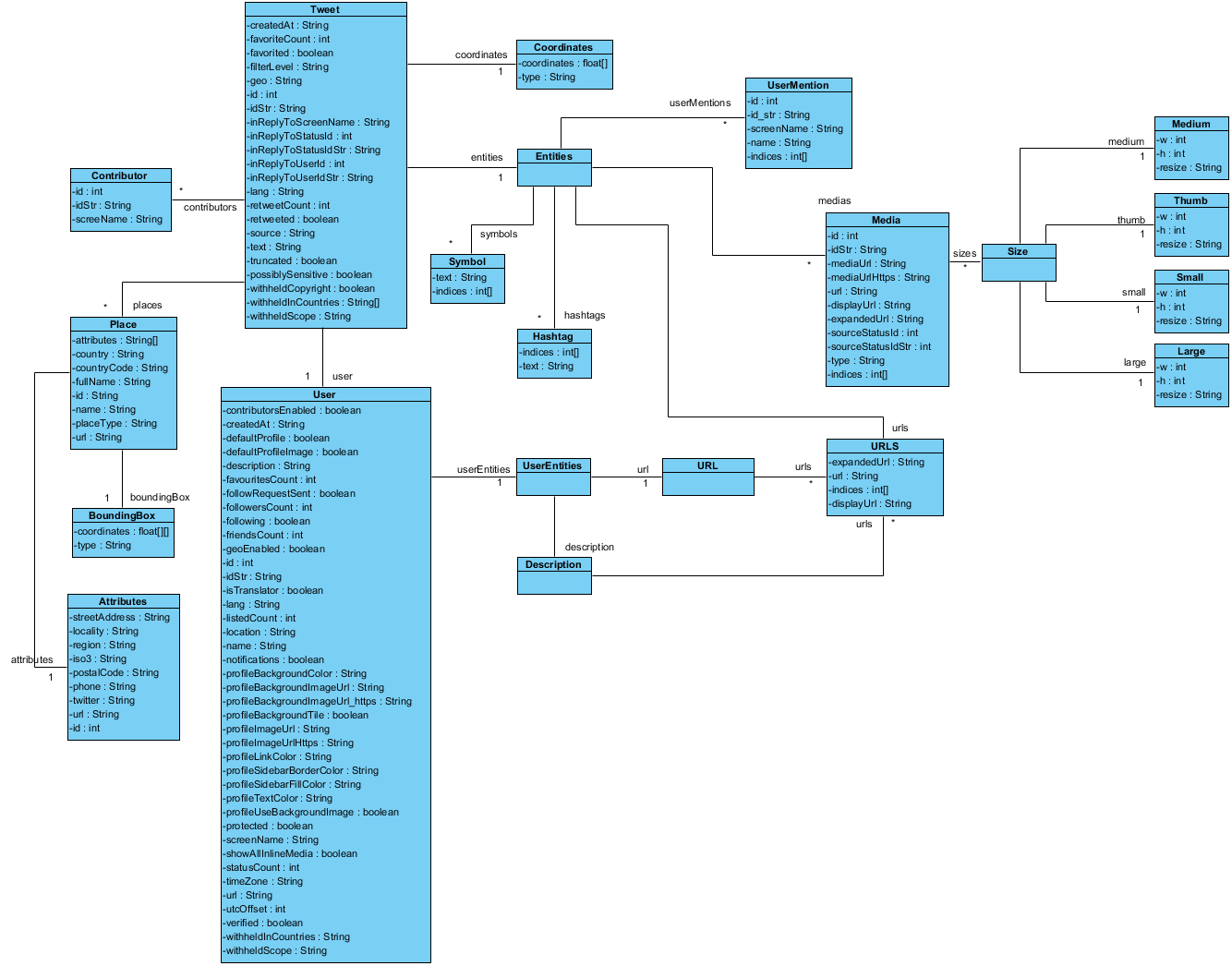
Sentiment analysis in Twitter is a field that has recently attracted research interest. Twitter is one of the most popular microblog platforms on which users can publish their thoughts and opinions. Sentiment analysis in Twitter tackles the problem of analyzing the tweets in terms of the opinion they express. This survey provides an overview of the topic by investigating and briefly describing the algorithms that have been proposed for sentiment analysis in Twitter

# A real-time Twitter sentiment analysis using an unsupervised method (<https://dl.acm.org/citation.cfm?id=3102282>)

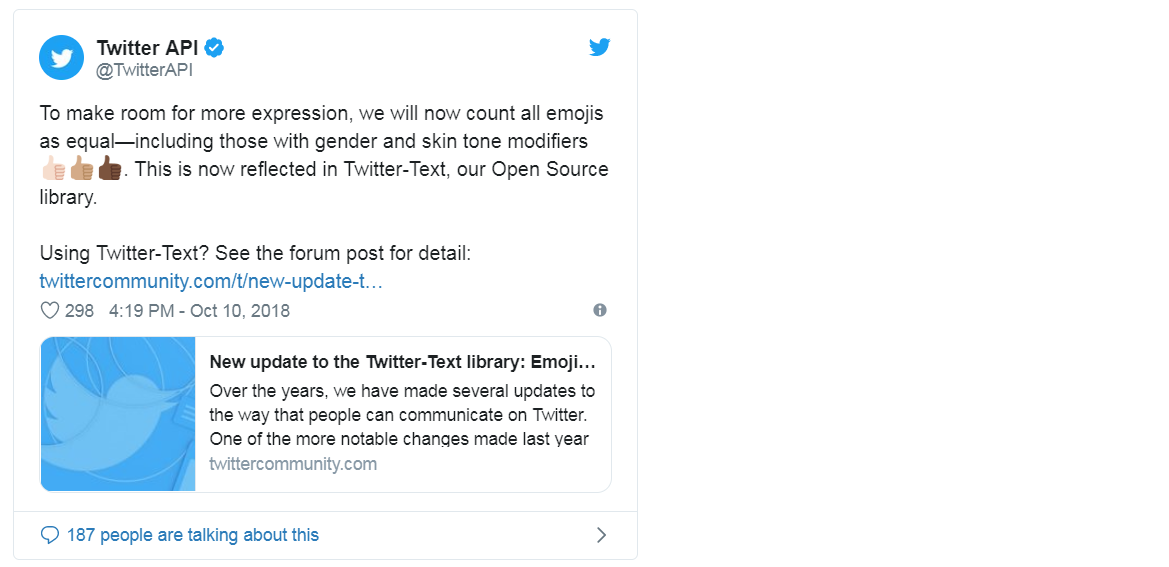
A real-time implementation of a system that can discover and track opinions on Twitter. A recommended approach using machine learning techniques, to analyze opinions and detect tweets polarity.

# Dataset

## Twitter DataSet



The above picture displays all the elements, relationships between elements and element datatype of a single tweet. Not all the data elements are mandatory as seen in a sample tweet below.



For the project, below dataset will be considered

Tweet.created\_at

Tweet.id

Tweet.text

Tweet.user.id

Tweet.user.name

Tweet.user.screenName

Tweet.user.followersCount

Tweet.coordinates

Tweet.place

Tweet.retweetCount

Tweet.entities

Tweet.retweeted

# Approach

### Create repository

Create a version control repository for all the artefacts developed for the project. A github repository is created at : <https://github.com/hdossani/ckme136_capstone_project>

### Create Lambda function to collect raw tweets

Create a lambda function using python programming language to collected raw tweets. Promote the function as a microservice so that users can pass additional parameters for interest and language e.g. AWS, Food, Game, etc. for interest and en, fr, sp for English, French and Spanish language respectively.

### Save Raw Tweets

Save the raw tweets collected in S3 bucket.

### Trigger Lambda Function To Preprocess Data

Saving raw tweets in S3 bucket will trigger lambda function to clean the raw tweets. Clean the:

* URLs
* Hashtags
* Mentions
* Reserved words (RT, FAV)
* Emojis
* Smileys

Save clean data in another S3 bucket.

### Process Clean Data

Process clean data using Amazon Comprehend. Amazon Comprehend is a natural language processing (NLP) service that uses machine learning to find insights and relationships in text.

### Save the results

Save the results of Amazon Comprehend to understand how positive or negative the sentiment is, based on twitter text analysis.

### Query Results

Create Amazon Athena tables to provide SQL interface to data in S3.

### Let Data Tell the Story

Use Amazon QuickSight to create data visualization using several query criteria.