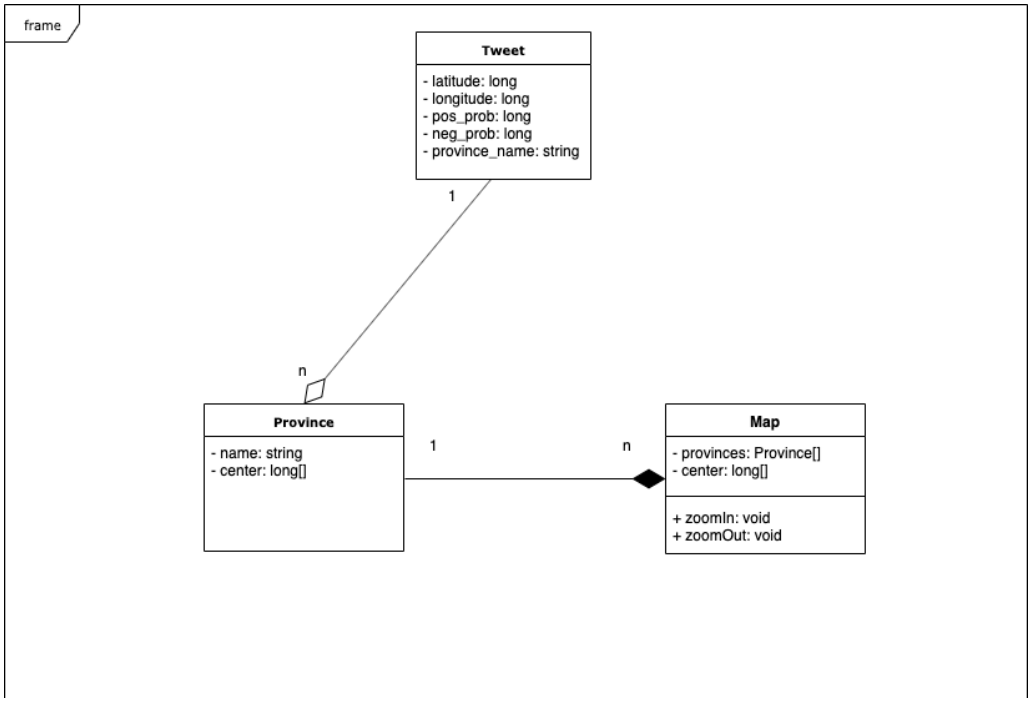
Bi-weekly report#2

(Primary Implementation)

Team members: Jiqing Zhu, 105092951
Jiaxiu Li, 110008532
Weiwei Cao, 105171487
Hang Zhu, 105170587

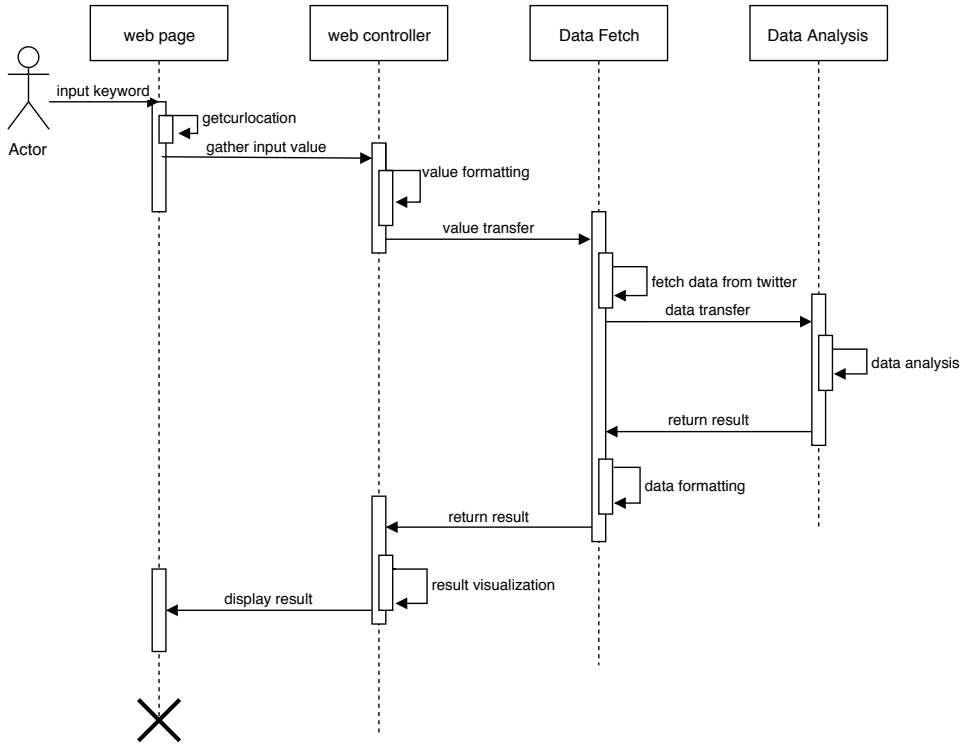
Date: 8th.Nov
To: Dr.Pooya

I. Class diagram



II. Interfaces between back-end and front-end

Users type a keyword in search bar in webpage, and backend will automatically generate current location and bundle (keyword, location) pass to data fetch module. The Data Fetch module is responsible for fetching data from Twitter based on provided value. After data clustering, Data Fetch module passes data to Data Analysis module. Once the data analysis is done, Data Analysis pass the raw data to Data Fetch to do the data formatting to pass to Web module for displaying in Map.



III. API research

a. Social media data extraction

1. Set up application in Twitter Developer platform for accessToken and accessSecretKey

2. Research Tweepy, which is the official API framework of Twitter

3. Design interfaces for data transformation of visualization

4. Implement the demo version of classes, objects, parameters and functions
(The pictures shows the testing data while the key word is #unitedAIRLINES)

b. Text analysis

1.TextBlob is a popular python package for processing nature language. It's based on the NLTK library, which is a very strong and powerful library to deal with natural language. There are many APIs in TextBlob for analyzing and handling text data, you can even build your own text classifier based on TextBlob built-in materials.

```
(base) WeiweiCao-MacBook-Pro:~ caoweiwei$ python
Python 3.7.3 (default, Mar 27 2019, 16:54:48)
[Clang 4.0.1 (tags/RELEASE_401/final)] :: Anaconda, Inc. on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> from textblob import TextBlob
>>> from textblob.sentiments import NaiveBayesAnalyzer
>>> tweet = "I fight for the Constitution, individual liberty and the freedoms that make this country great."
>>> tweet_tb = TextBlob(tweet)
>>> tweet_tb.sentiment
Sentiment(polarity=0.4, subjectivity=0.575)
>>> tweet_tb.tags
[('I', 'PRP'), ('fight', 'VBP'), ('for', 'IN'), ('the', 'DT'), ('Constitution', 'NNP'), ('individual', 'JJ'), ('liberty', 'NN'), ('and', 'CC'), ('the',
 'DT'), ('freedoms', 'NNS'), ('that', 'WDT'), ('make', 'VBP'), ('this', 'DT'), ('country', 'NN'), ('great', 'JJ')]
>>>
>>>
```

2. For this project, we can use `TextBlob.sentiment` method for getting the sentiment polarity and subjective results. For further usage, there is a classification model using Bayes algorithm. It can classify the tweets into different clusters.

```
>>>
>>> text = "Thrilled to tell you @NASA will contribute the CASE instrument to @ESA's @ARIELTelescope. Together we'll watch starlight streaming through
the atmospheres of planets outside our solar system, to understand what they are made of, and how they came to be: https://go.nasa.gov/2Cr9uy8"
>>> text_tb = TextBlob(text, analyzer=NaiveBayesAnalyzer())
>>> text_tb.sentiment
Sentiment(classification='pos', p_pos=0.9821915837049848, p_neg=0.017808416295010582)
>>> text_tb.sentiment.classification
'pos'
>>> text_tb.sentiment.p_pos
```

IV. Reference

Tutorial: Quickstart¶. (n.d.). Retrieved from https://textblob.readthedocs.io/en/latest/quickstart.html#quickstart.

API Reference¶. (n.d.). Retrieved from https://textblob.readthedocs.io/en/dev/api_reference.html.

Tweepy. (n.d.). Retrieved from https://www.tweepy.org/.

Apply for access – Twitter Developers. (n.d.). Retrieved from https://developer.twitter.com/en/apply-for-access.html.

V. Appendix — Team discussion records