
Storytelling with Streaming Data

HW1-Presidential Election News Analysis

YU WANG UNI: yw2783 - February 10, 2016

Purpose

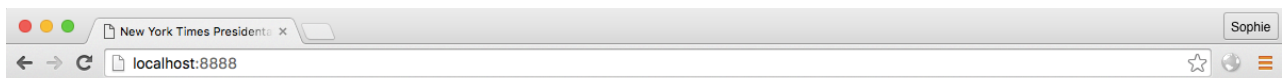
This project consumes streaming data from New York Times API and analyzes news about presidential election and its corresponding sentiment. By parsing the result, news would be further classified to be Republican or Democratic representatives. The amount of news of each cluster would be calculated in real time and sentiment analysis would be conducted simultaneously. Thus we can collect information about how much news are there about two parties and what is the sentiment.

This document will demonstrate how the data will be presented (with demo graph), how the streaming data is processed in each step, what is the structure of source code, and how to run the application. The following sections will elaborate the information respectively.

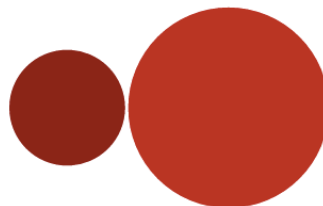
Data Visualization

The project will present news of Presidential Election from two aspects:

1. The number of news for Republican and Democratic in real time. The number is illustrated by the size of bubble in web front page.
2. The sentiment of each news for two parties, whether it is neutral, positive or negative. I virtualize the fact by using D3 graph, the color of each bubble represents the percentage of positive sentiment.

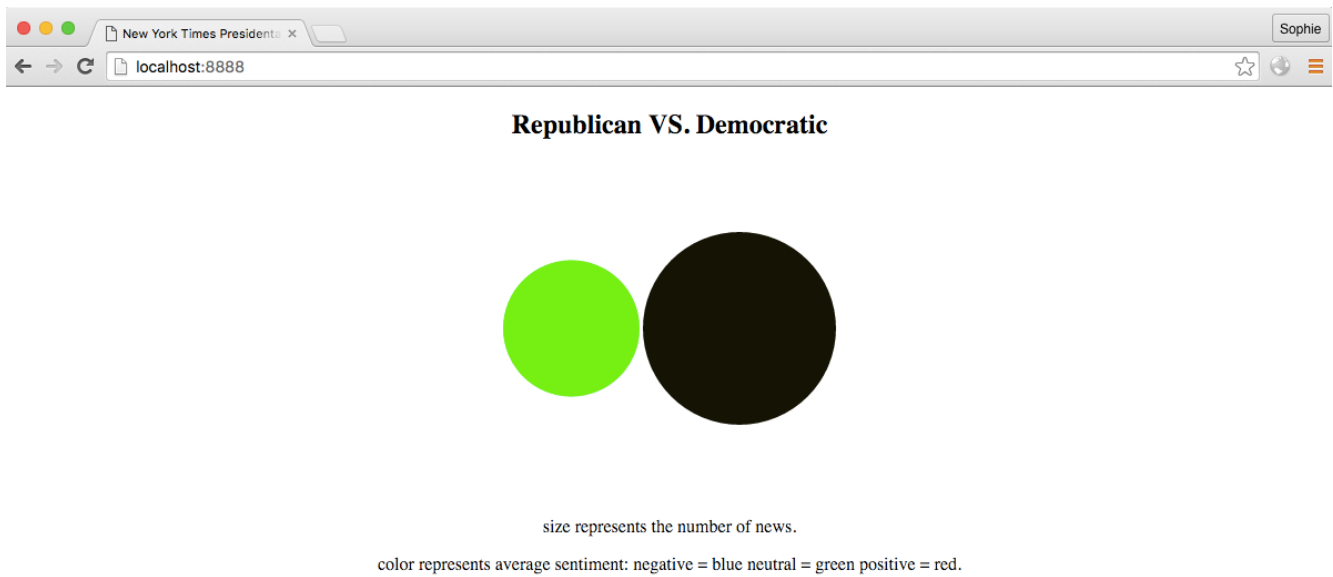


Republican VS. Democratic



size represents the number of news.

color represents average sentiment: negative = blue neutral = green positive = red.

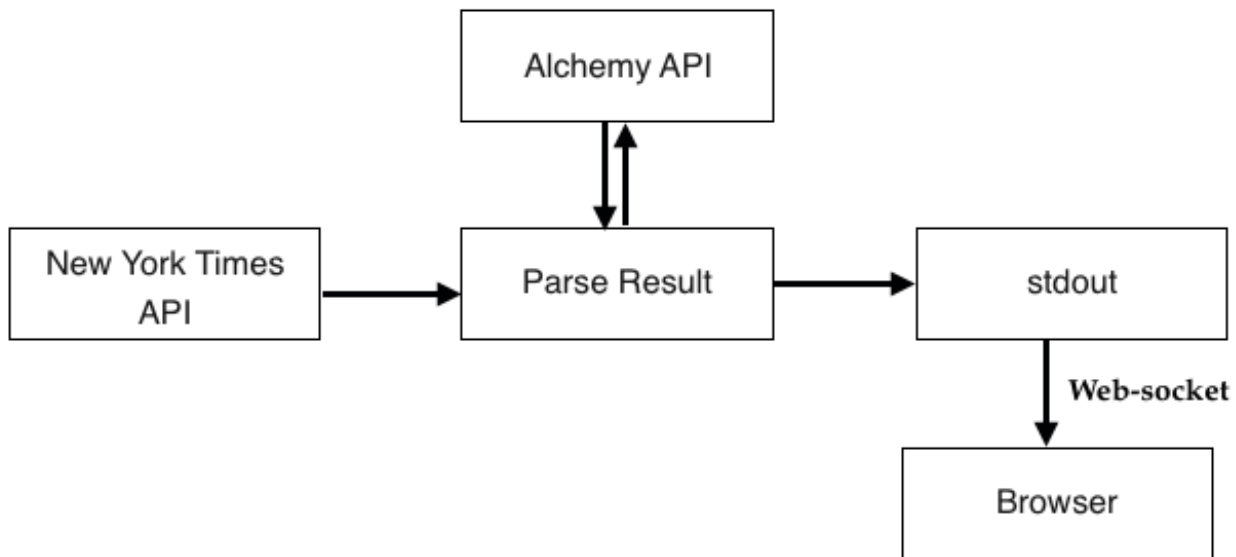


The figure above demonstrates real time news about Republican and Democratic. The diameter of circle represents the amount of news, which shows the popularity of the party. The color of circle means average sentiment score of the news, which shows the attitude of editors. (negative = blue neutral = green positive = red)

How to Run

1. Run the shell script in folder
`# . run.sh`
2. CD to the current project folder and build the python server :
`# python -m SimpleHTTPServer 8888`
3. Open the webpage in browser by link:
`http://localhost:8888/`

Processing Flow



Source Code

- Stream.py
 - Parse the streaming data and retrieve Democratic and Republican news from all category of New York Time News
 - Calculate number of news of each party;
 - Analyze the sentiment of each news.
- script.js
 - Listens the specified port and load information to the front page
- run.sh
 - Build the web socket connection, push stream data into the specific port
- index.html
 - Web front-end page to demonstrate the result and graph