Visual Analysis of Financial News Articles

By CUI XIWEN

HUA1,Advised by Prof. Huamin

Introduction

Big data and big data analytics is a new and upcoming field in technology. The key to dealing with so much data is to arrange it visually to extract what really matters and where the things are going. The project aims to develop a visualization system, which can filter out the remarkable portion of data from summary and extract the trend words and illustrate them in a words cloud. In addition, the colors and font size reflect frequency of occurrence. We choose the first frequency 3 words (key words), which indicate the trends during a time period and generate a line plot that shows their frequencies and at the same time analyze the sentiments from where the key words showed up.

Objectives

The main objective of this project is to design a visualization system that translates financial news articles into a graphical presentation that contains the following functions:

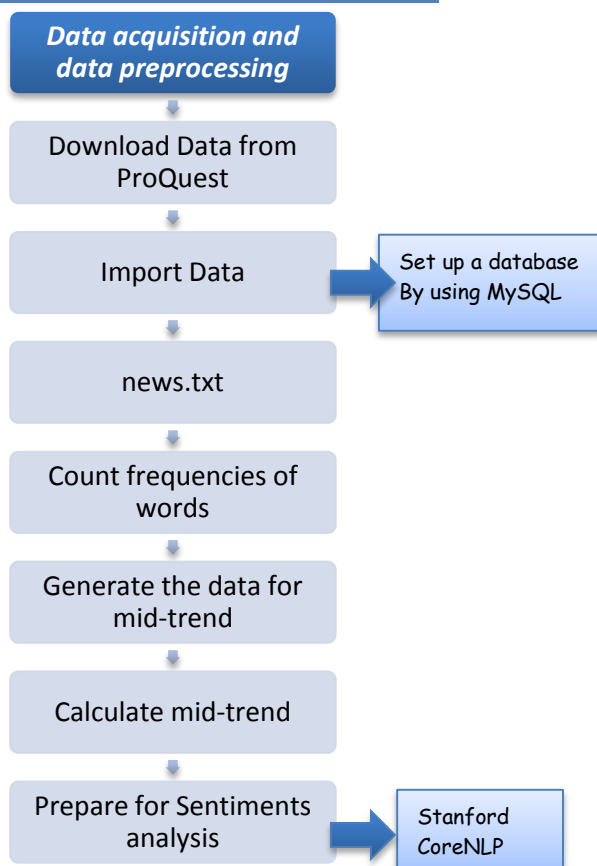| Trend Word Visualization	• Seek Trends • Generate a words cloud
Mid-Trend Line	• How China,economic & growth's frequencies change • The trend is flat increasing or decreasing
Sentiments-Trend Chart(Scatterplot)	• Stanford CoreNLP to do localized sentence level • Generate a 2-D scatterplot of sentiments

Methodology

This project is mainly divided into two parts:
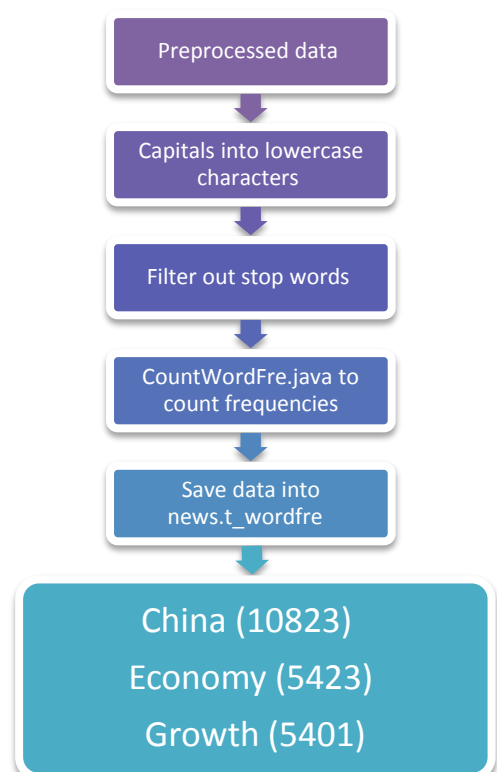***Data acquisition and data preprocessing***
Data Visualization.

Data acquisition and data preprocessing

```
Data acquisition and
data preprocessing
        ↓
Download Data from
ProQuest
        ↓
Import Data  →  Set up a database
                By using MySQL
        ↓
news.txt
        ↓
Count frequencies of
words
        ↓
Generate the data for
mid-trend
        ↓
Calculate mid-trend
        ↓
Prepare for Sentiments  →  Stanford
analysis                   CoreNLP
```

Data Visualization

In data visualization part, we mainly developed three objectives: Trend word visualization-seek trends,Mid-trend Line and Sentiments-Trend chart(ScatterPlot).

Trend word visualization-seek trend

```
Preprocessed data
        ↓
Capitals into lowercase
characters
        ↓
Filter out stop words
        ↓
CountWordFre.java to
count frequencies
        ↓
Save data into
news.t_wordfre
        ↓
China (10823)
Economy (5423)
Growth (5401)
```
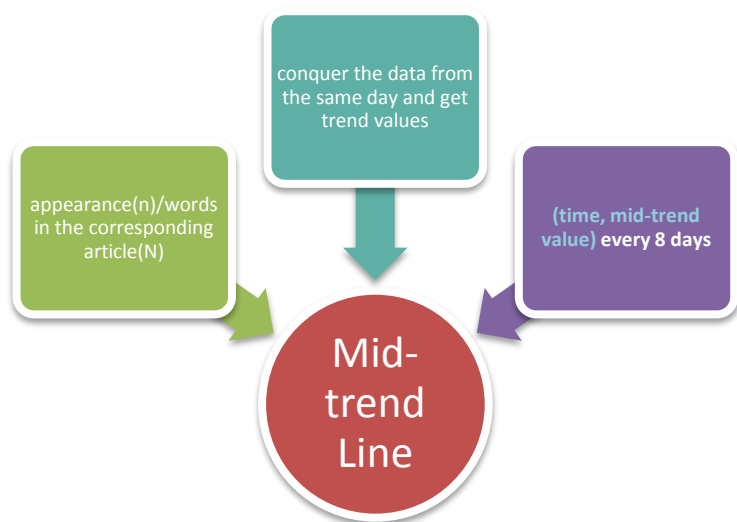
Input the number the user want how many words will be shown in the word cloud and use D3's *d3.layout.cloud.js* to generate a word cloud (The bigger the word is, the frequency it showed up is higher)

Mid-trend Line

In the part of Trend words visualization we settle 3 key words (**China economic & growth**) and make them our centers in this part. After finding the specific topics and themes, we want to know whether the trend is flat, increasing or decreasing and whether it is variable or stable.

The diagram shows the trend of the three key words (Blue line stands for China, green one stands for growth and the orange one stands for economic.)

$$Trend\ value = \frac{n}{N} * 100$$

where n stands for the word's appearance and N stands for words in the corresponding article. This equation is used for calculate Trend Value.

Sentiments-Trend Chart

Different from mid-trend line, we do "localized sentence level" semantic estimation. Since we just focus on the key words, we localized the semantic value of the variable keywords in the sentence and perform the semantic calculation based on *Stanford CoreNLP*. And the trend chart (scatter plot) shows the semantic values of the key words we selected before. When the mouse moves over a node, the selected node will be highlighted and brief information such as the appearance frequency and semantic value will be given.

For doing sentiments analysis we need to create a negative txt and a positive.txt and set up levels for emotional words.

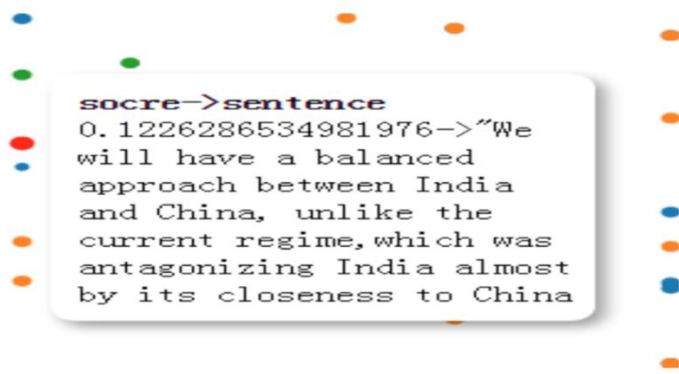| Levels	insufficiently	ish	more	very	over	most
Value	0.5	0.8	1.2	1.25	1.5	2

Here are some example words of different levels with distinct sentiments values.

The interactive 2-Dscatter plot shows the three key words', China (blue), growth (green) and economic (orange), sentiment visualization in the 3000 financial news.


```
socre->sentence
0.1226286534981976->"We
will have a balanced
approach between India
and China, unlike the
current regime,which was
antagonizing India almost
by its closeness to China
```

When the mouse comes to a node the node will shows the specific information about its sentiments value and the sentence where this point comes from the article.

Implementation

Development Tools

MySQL	Used to set up a new database and create tables for searching and preprocessing data
Eclipse	Main programming language-Java
CoreNLP	Used for sentiments analysis
SpringMVC	To create a dynamic web project

	Database
MySQL	Create tables (t_news, t_wordfre, t_toptrend, t_trend, t_mid, t_level, t_positive, t-negitave)

SpringMVC

Objectives	URL & Request
Trend word visualization	http://localhost:8080/SpringMVC/cloud/show http://localhost:8080/SpringMVC/cloud/highFre
Mid-trend line	http://localhost:8080/SpringMVC/trend/show http://localhost:8080/SpringMVC/trend/trenddata
Sentiments-Trend Chart (Scatterplot)	http://localhost:8080/SpringMVC/sentiment/show http://localhost:8080/SpringMVC/sentiment/data

Conclusion

The system developed is aim to do visual analysis of Financial news articles. The main purpose of this project is to design a visualization system that translates financial news into a graphical presentation and the three main objectives are Trend words visualization-seek trend, Mid-trend line and Sentiments-Trend Chart (Scatter plot).
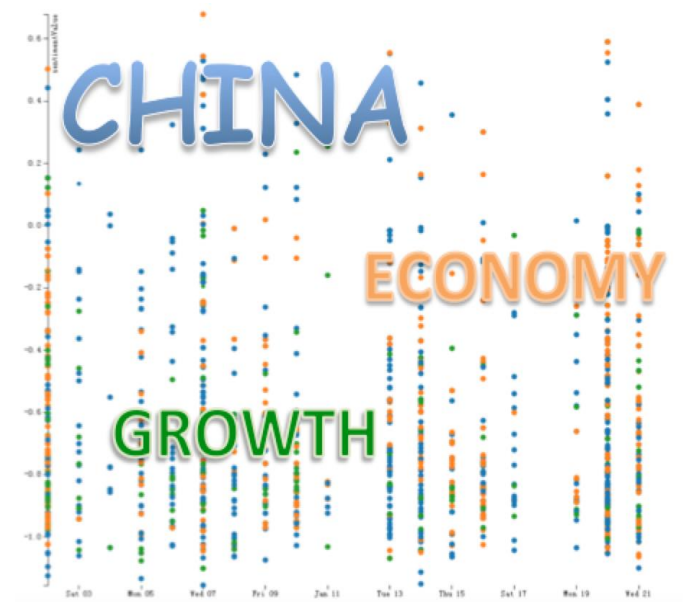
The result of the financial news visualization shows in 2014 the situation of China's economy is not that optimistic since we can see the sentiments points are mostly under the 0 line which means negative. This illustrates the economy situation in China is still under the influence of Financial Crisis and the tendency is not very optimistic. But we can see that different months have different graphs and sentiments scatter plots, which means there are minor differences and this could have some reflection on what happened then that have influences on China's economy.

This project can help people who study this field with do some analysis in economy and make up some business decisions. And maybe help people who want to learn last year's China's economy and forecast the future tendency.