
Analytics Project Fall 2014

Analysis of stock performance based on Sentiment Analysis

Correlating stock performance of Apple inc. based on Sentiment Analysis of the internet content and Wikipedia page counts. Data will be aggregated from April to November.

Sharad S Shivmath

Analysis of stock performance based on Sentiment Analysis

Background

- We are basically computing sentiment of content people post on the internet.
- sentiment(polarity, subjectivity)
- sentiment("Textblob is amazingly simple to use. What great fun!") = polarity=0.39166666666666666, subjectivity=0.4357142857142857
- polarity -> [-1.0, 1.0], subjectivity -> [0.0, 1.0]

Motivation

- People are always sharing on the internet.
 - One of the ways companies can understand what their consumers want
 - It is interesting to understand social opinion which was not possible before.
-

Analysis of stock performance based on Sentiment Analysis

Data Sources

- WET file format
 - WARC(Web Archive) file format. Released to public by CommonCrawl project
 - WET contains text . Series of 'records'
 - ```
WARC/1.0
WARC-Type: conversion
WARC-Target-URI: http://advocatehealth.com/condell/emergencyservices3
WARC-Date: 2013-12-04T15:30:35Z
WARC-Record-ID:
WARC-Refers-To:
WARC-Block-Digest: sha1:3SJBHMFPOCUJEHJ7OMGVCRSHQTWLJUUS
Content-Type: text/plain
Content-Length: 5765
```

...Text Content...

---

# Analysis of stock performance based on Sentiment Analysis

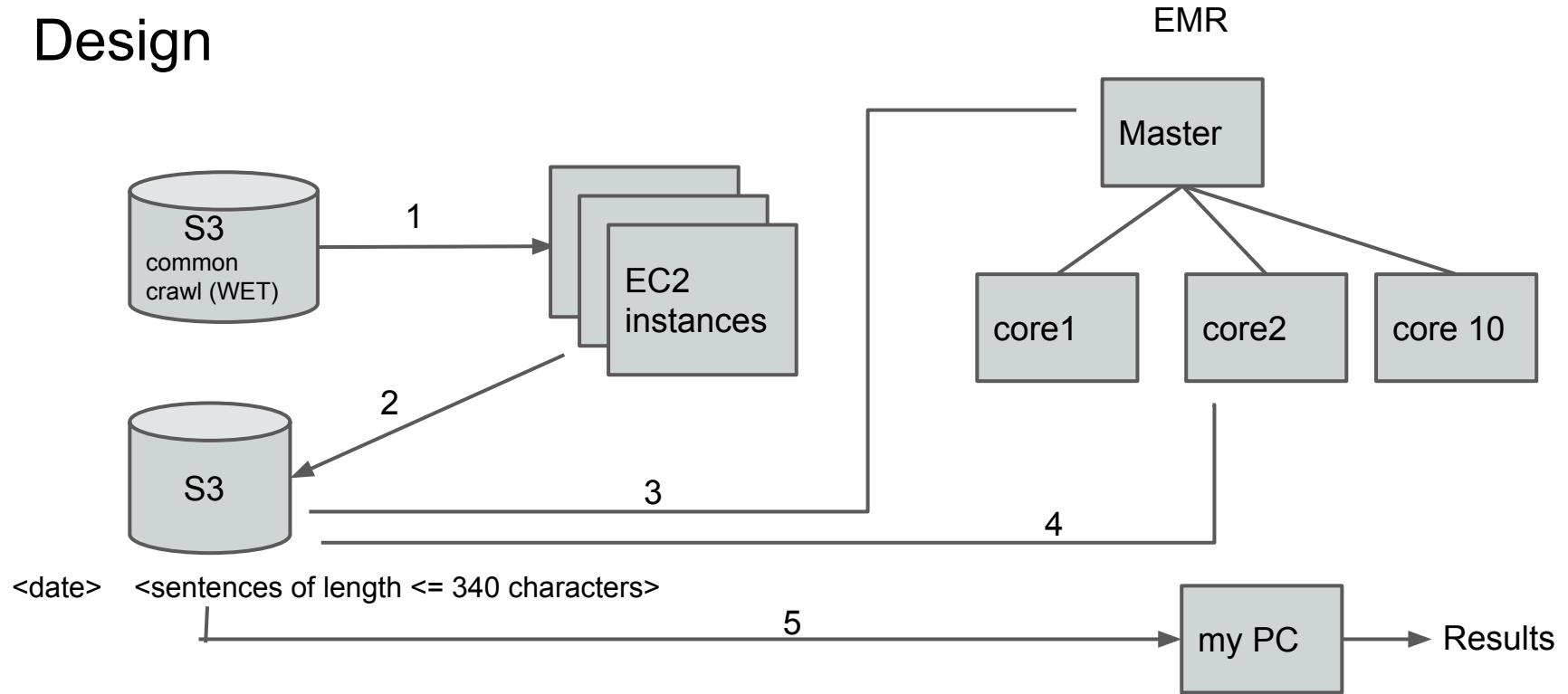
---

- Wikipedia page counts of 'apple inc.'
  - <date>      <page counts>
- Historic daily stock quotes of apple inc.:
  - <date>      <closing value>

# Analysis of stock performance based on Sentiment Analysis

---

## Design



# Analysis of stock performance based on Sentiment Analysis

---

## Results:

- <https://www.google.com/fusiontables/DataSource?docid=13KKVu3TX0F9R2hQOG6FbElf-KrPoFgSmOuxa1mPU>
- Although data is large, Crawler record dates are centered around particular dates, and not distributed. Statistical methods like interpolation, and distribution approximations would have been helpful.
- Opinions are less subjective, with some phases when subjectivity  $> 0.5$ , which is a good thing.
- Spikes in page count could be correlated to fall in stock prices.
- Opinions are generally positive.

<date>   <polarity>   <subjectivity>   <page counts>   <stock closing value>

---

# Analysis of stock performance based on Sentiment Analysis

---

## Obstacles

- Pig is hard.
- Using external libraries in MapReduce program.
- Tried using UDFs in Pig script.

# Analysis of stock performance based on Sentiment Analysis

---

## Conclusion:

- A more consistent data would be good.
- Deeper sentiment analysis definitely hold answers to questions about crowd behavior.
- Learnt about Amazon's cloud offerings and using them.

## Acknowledgments:

- Professor Suzanne McIntosh
  - Amazon Web Services for their voucher
  - Common Crawl Foundation, Wikipedia
-



Analysis of stock performance based on Sentiment Analysis

---

Thank You!

---