

Named-Entity Sentiment Analysis

Matthew Moore
Cooper Gagnon
Ziyu Qiu

Preface

- The media, and news corporations often show bias and opinion in news headlines and articles.
- This influences how the public views these topics.
- It is important to understand the influence the news has.



Our Idea

- To create a simple interface which can display how the news is influencing certain topics.
- Use NLP techniques and tools to extract underlying sentiment.
- Gather large amounts of news headline data.
- Extract sentiment and named-entities, and associated sentiment with entities of same headline.
- Analyse data on a 'per day' basis, to show how it evolves over time.



Data Collection

- Scraped 5000+ news websites for their RSS feed
- Found 2287 unique RSS feeds
- Scraped the titles of all the RSS feeds daily



```
-<item>
  <title>RSS Solutions for Restaurants</title>
  -<description>
    <b>FeedForAll </b>helps Restaurant's communicate with customers. Let your customers know the latest specials or events.
    <br> Entertainment <br> Calendar of Events </i></font>
  </description>
  <link>http://www.feedforall.com/restaurant.htm</link>
  -<category domain="www.dmoz.com">
    Computers/Software/Internet/Site Management/Content Management
  </category>
  <comments>http://www.feedforall.com/forum</comments>
  <pubDate>Tue, 19 Oct 2004 11:09:11 -0400</pubDate>
</item>
```

Source: <https://feedforall.com/sample.xml>

Data

- Data was stored in mySQL server.
- Two tables:
- One table for storing news headline data.
- Another table for storing named entity and sentiment information.
- As of right now the count for named entities is: **190306**
- We utilized techniques from python NLTK.

id	date	update	title	description
3	2012-10-05 01:37:03	2021-02-26 22:26:31	Daryl Hannah Arrested in Texas	Daryl Hannah was arrested Thursday. She was...
4	2012-06-30 05:03:25	2021-02-26 22:26:31	OMG Adele is Having a Baby	The UK soul screaming superstar had whose hit...
5	2012-04-14 11:15:59	2021-02-26 22:26:31	Bobby Petrino Fired	Bobby Petrino, head coach for the University of...
6	2012-04-12 06:43:56	2021-02-26 22:26:31	Lollapalooza 2012 Lineup	This year, Lollapalooza has a very exciting lineu...
7	2012-04-12 02:35:47	2021-02-26 22:26:31	Shabazz Muhammad Signs With UCLA	With an NCAA-record 11 national championship...
8	2012-04-12 12:12:28	2021-02-26 22:26:31	Tony Romo's New Baby Boy	Tony Romo is the proud parent of a new baby b...
9	2012-03-14 07:10:55	2021-02-26 22:26:31	Nikon Coolpix S9100 Wins Best Reviewed Com...	Nikon is legendary in the photography space. B...
10	2012-03-10 02:03:04	2021-02-26 22:26:31	It's Daylight Savings Time 2012	It's that time of year again. Kind of a love hate r...
11	2020-06-17 03:00:16	2021-02-26 22:26:32	UEFA xác nhận giữ trọn 12 thành phố đăng cai...	Liên đoàn bóng đá châu Âu xác nhận giữ trọn 1...
12	2020-05-24 03:01:25	2021-02-26 22:26:32	Ramos, Modric và những tên tuổi lớn có thể lờ...	LĐBĐ châu Âu (UEFA) đã buộc phải đưa ra quy...
13	2020-05-19 03:06:44	2021-02-26 22:26:32	'EURO 2020 có thể rút xuống với 8 hoặc 9 thàn...	UEFA sẽ hoãn phiên họp chốt danh sách thành...
14	2020-04-23 03:00:23	2021-02-26 22:26:32	UEFA giữ nguyên danh xưng Euro 2020 dù hoả...	Giải vô địch châu Âu vào mùa hè tới vẫn sẽ đư...
15	2020-04-22 03:10:20	2021-02-26 22:26:32	Play-off EURO 2020 sẽ tổ chức vào tháng 10 v...	Theo ông Gary Owens, Giám đốc điều hành củ...
16	2020-02-24 09:02:12	2021-02-26 22:26:32	Đại tướng Hà Lan và hàng nghìn ngư 'khủng' tạ...	Hilary Ronald Koeman sẽ phải đau đầu để lựa ch...

347	wales	0.26060...	2021-02-27 00:00:00	2021-02-27 00:00:00	4
348	england	4.39990...	2021-02-27 00:00:00	2021-02-27 00:00:00	25
349	john	-5.0963...	2021-02-27 00:00:00	2021-02-27 00:00:00	107
350	red	1.43030...	2021-02-27 00:00:00	2021-02-27 00:00:00	36
351	journeyman	0.5106	2021-02-27 00:00:00	2021-02-27 00:00:00	1
352	leeds	6.42130...	2021-02-27 00:00:00	2021-02-27 00:00:00	30
353	aston	3.477	2021-02-27 00:00:00	2021-02-27 00:00:00	12
354	marcelo	2.4242	2021-02-27 00:00:00	2021-02-27 00:00:00	4
355	sue	0.3071	2021-02-27 00:00:00	2021-02-27 00:00:00	1
356	grimsby	0.4497	2021-02-27 00:00:00	2021-02-27 00:00:00	2

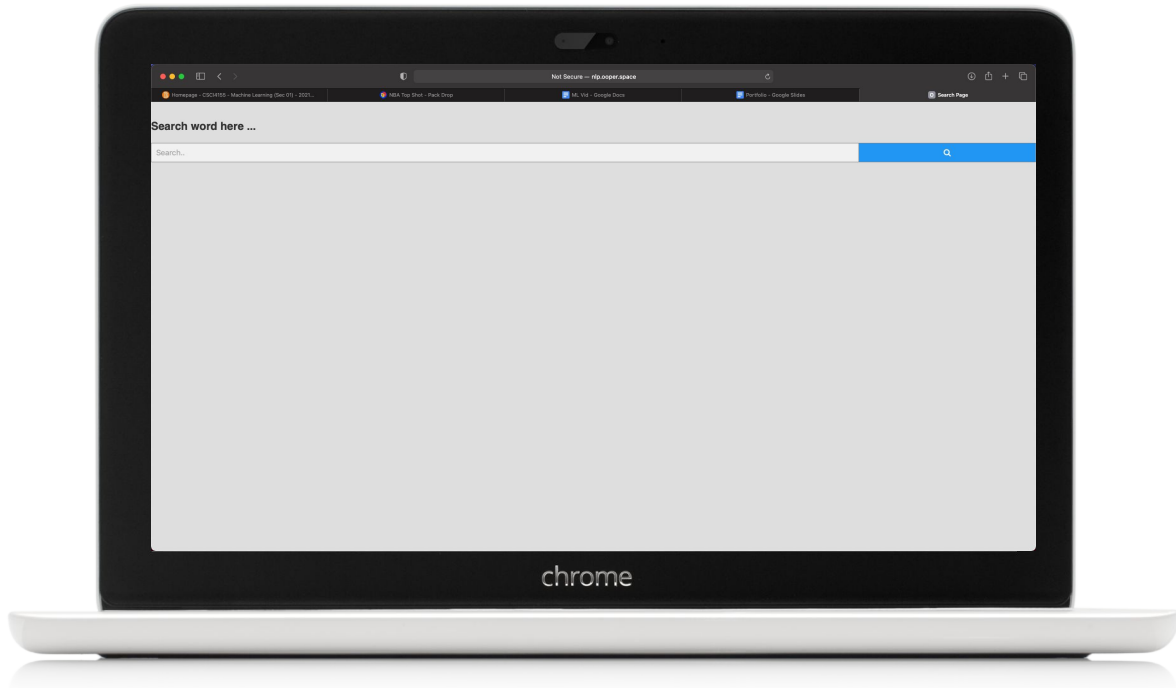
Sentiment Extraction

- Used the Python NLTK.
- Extracted named-entities via chunking.
- VADER Sentiment Intensity Analyzer was ran on entire sentence, and associated with entity in that sentence.
- The sentiment for each sentence was summed in the database, and the number of sentences including the named entity was also recorded, for each day.
- Storing this information allowed for calculating overall average, and graphing day by day average.



User Interface

We created a simple user interface, upon visiting the app one is prompted to search for any word, or any named-entity.



Search Result

usa

jerusalem

susan

crusader

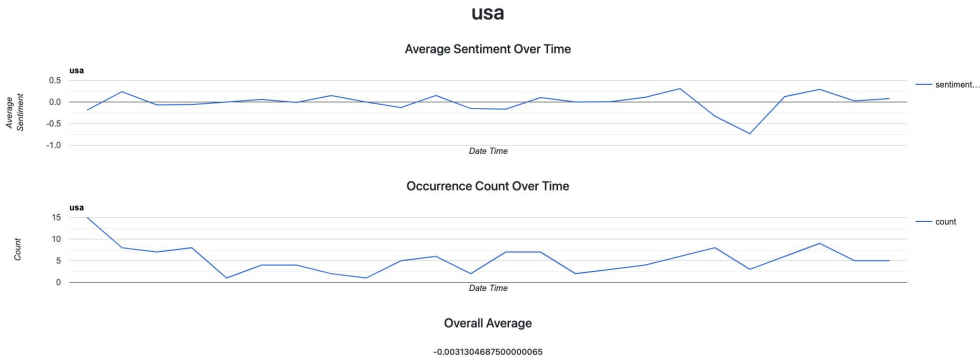
thousands

usaf

thousand

wausau

rusal



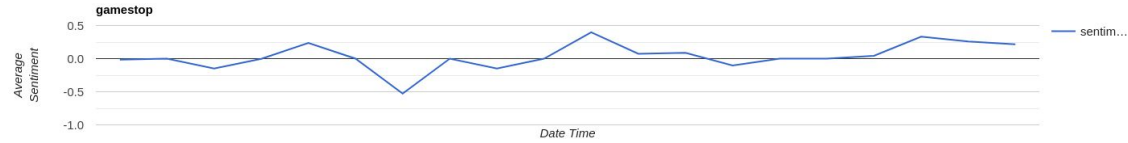
After a search is conducted the user is led to a display of all related words.

Once they select a word they are led to a screen which displays graphs representing the average sentiment over time, the count of the named-entity over time, and an overall average sentiment.

Examples

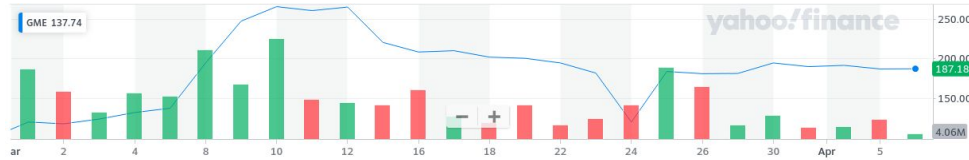
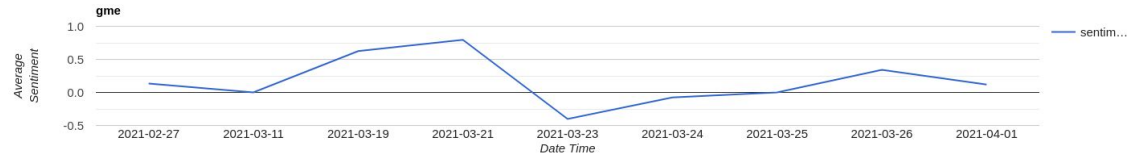
gamestop

Average Sentiment Over Time



gme

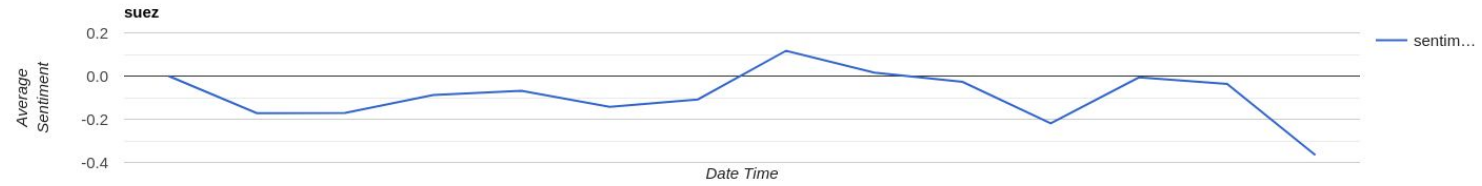
Average Sentiment Over Time



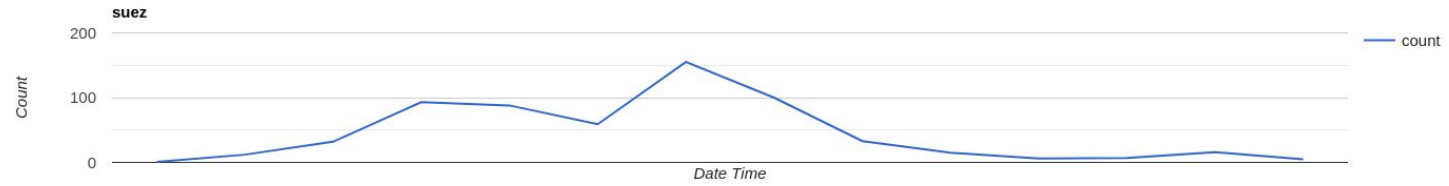
Source: <https://finance.yahoo.com/quote/GME?p=GME>

suez

Average Sentiment Over Time



Occurrence Count Over Time

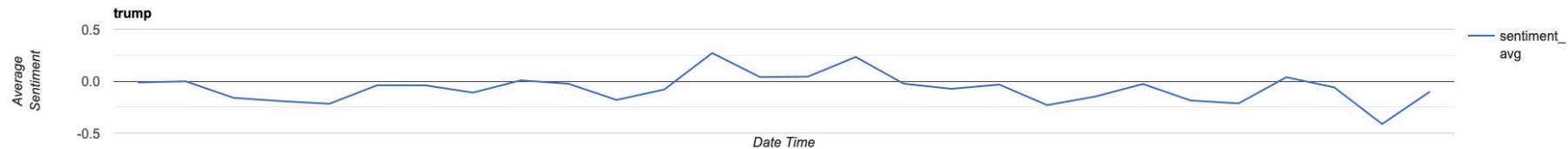


Overall Average

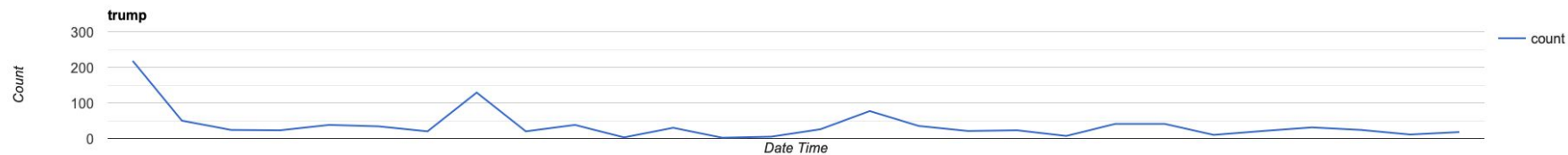
-0.06251848874598073

trump

Average Sentiment Over Time



Occurrence Count Over Time

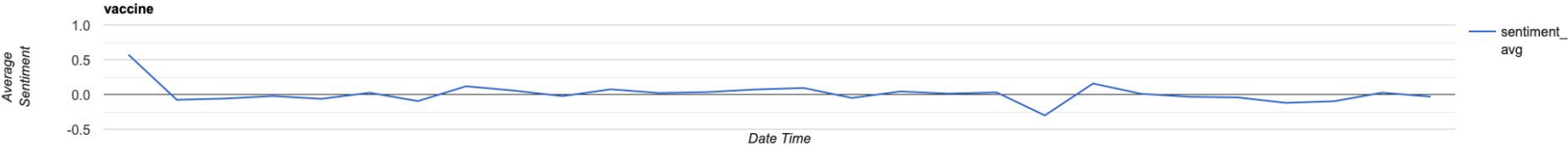


Overall Average

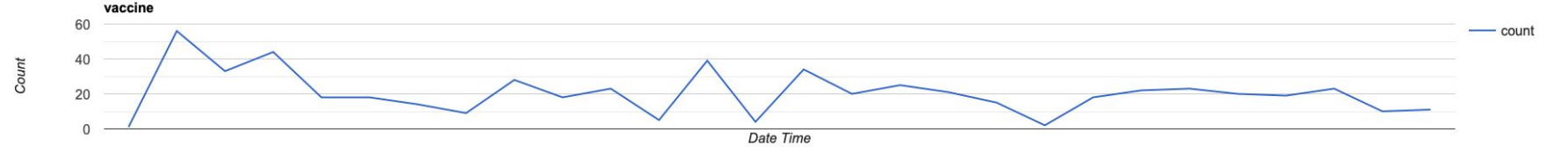
-0.04363254901960784

vaccine

Average Sentiment Over Time



Occurrence Count Over Time



Overall Average

-0.005084467713787086