

Tweets telling stories; Tweets as data

Adam Marcus
MIT Computer Science and
Artificial Intelligence Lab (CSAIL)

(with Michael Bernstein, Osama Badar,
David Karger, Sam Madden, Rob Miller)

Twitter data for fun and stories

Twitter data for fun and stories

How can we tell stories with the tweet stream?

Twitter data for fun and stories

How can we tell stories with the tweet stream?

How can we extract data from the tweet stream?

Twitter data for fun and stories

How can we tell stories with the tweet stream?

TwitInfo

How can we extract data from the tweet stream?

Twitter data for fun and stories

How can we tell stories with the tweet stream?

TwitInfo

How can we extract data from the tweet stream?

TweeQL

TwitInfo

How can we tell stories with the tweet stream?

TwitInfo

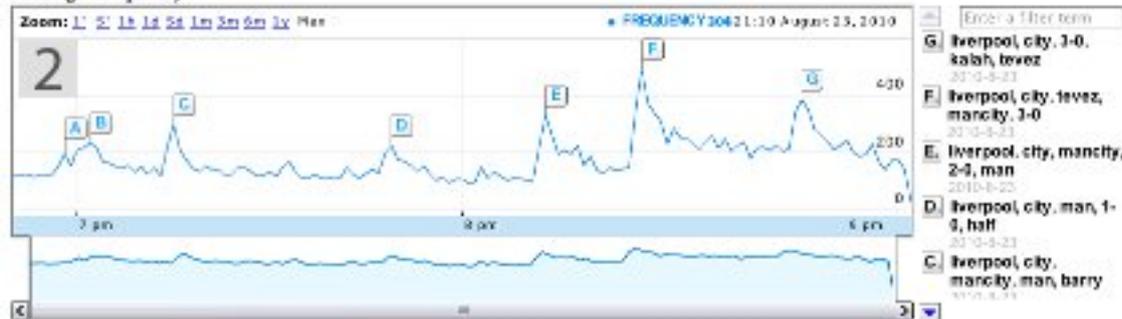
How can we tell stories with the tweet stream?

twitInfo

august 23 manchester city vs. liverpool **1**

Keywords: football, soccer, nfl, premier, league, premierleague, manchester city, mancity, liverpool.
Event dates: Aug. 23, 2010, 6:50 p.m. - Aug. 25, 2010, 9:10 p.m.

Message Frequency



Tweet Map **3**



Relevant Tweets **4**

I'm getting ready for the liverpool game. I'm so excited

@football.Love Sheikh Mansour bin Zayed ManCity - Liverpool maçına izlemek için tribünde. İlk kez maça geliyorum!

Javier Mascherano refuses to face Manchester City as Barcelona make Liverpool £12m offer - report <http://tinyurl.com/267fhhq>

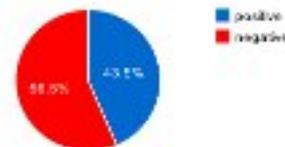
online football gambling sites? <http://isort.us/ze4t7r>

peluit ni diambil ga? RT @rwisnuwardhana: Watching man.city vs liverpool (@

Popular Links **5**

<http://bit.ly/PROVa> (cited by 4)
<http://tinyurl.com/2d4s46d> (cited by 4)

Overall Sentiment **6**



TwitInfo Uses

Automatically identify events
e.g., goals, earthquakes

TwitInfo Uses

Automatically identify events
e.g., goals, earthquakes

Backgrounding
e.g., Obama's last two weeks

TwitInfo Uses

Automatically identify events
e.g., goals, earthquakes

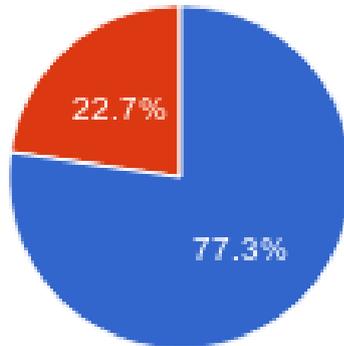
Backgrounding
e.g., Obama's last two weeks

Identifying sources on the ground
e.g., interviewing earthquake survivors

Cautionary Tweet Tales

Sentiment is tricky

e.g., “My warmest prayers go out to the people of Christchurch.”



Cautionary Tweet Tales

Sentiment is tricky

e.g., “My warmest prayers go out to the people of Christchurch.”

Location is evasive

Cautionary Tweet Tales

Sentiment is tricky

e.g., “My warmest prayers go out to the people of Christchurch.”

Location is evasive

Spam is everywhere

TwitInfo

Automated event detection makes tweet-based
story-telling possible

TwitInfo

Automated event detection makes tweet-based story-telling possible

Interfaces tell a story, people add context

TwitInfo

Automated event detection makes tweet-based story-telling possible

Interfaces tell a story, people add context



TweeQL

How can we extract data from the tweet stream?

TweeQL

How can we extract data from the tweet stream?

“It's a balmy 89°C in Phoenix”

location=Phoenix, temperatureC=89

TweeQL

How can we extract data from the tweet stream?

“It's a balmy 89°C in Phoenix”

location=Phoenix, temperatureC=89

“I'm starting to dig Obamacare!”

topic=Obama, sentiment=positive

Twitter data hacking is hard

- Learn the API
- Transform data
- Stuff it into a database

Twitter data hacking is hard

- Learn the API
- Transform data
- Stuff it into a database

Ad-hoc data processing

Twitter data hacking is hard

- Learn the API
- Transform data
- Stuff it into a database

Ad-hoc data processing



TweeQL extracts data from tweets as they pass through the stream

TweeQL extracts data from tweets as they pass through the stream

Data extraction

e.g., location, sentiment, temperature, opencalais

TweeQL extracts data from tweets as they pass through the stream

Data extraction

e.g., location, sentiment, temperature, opencalais

SQL-like queries

e.g., `SELECT location, text FROM twitter`

TweeQL extracts data from tweets as they pass through the stream

Data extraction

e.g., location, sentiment, temperature, opencalais

SQL-like queries

e.g., `SELECT location, text FROM twitter`

stream, not table

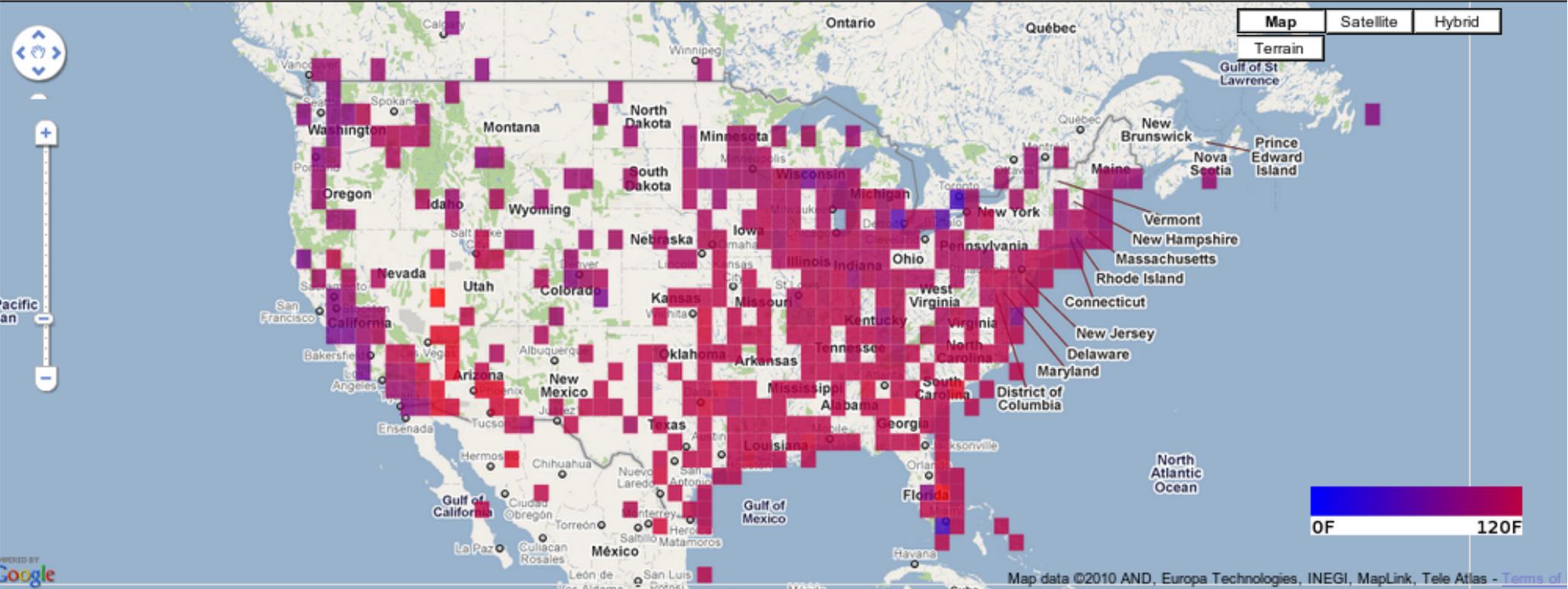


TweeQL demo
nerd alert

TweeQL's other features

- Aggregation
- Outlier detection
- Joins/mashups with other sources

Average temperatures extracted from tweets.

Unstructured data

Unstructured
data

+

Structured
queries

Unstructured
data

+

Structured
queries

=> Structured
data

Unstructured
data

+

Structured
queries

=> Structured
data

=> Meaningful
visualizations

Thanks!

- TweepQL: <http://github.com/marcua/tweepql>
- TwitInfo: <http://twitinfo.csail.mit.edu/>
- Ask me about Mechanical Turk + Journalism!

Adam Marcus

@marcua

marcua@csail.mit.edu

<http://people.csail.mit.edu/marcua>

august 23 manchester city vs. liverpool

Keywords: football, soccer, epl, premier_league, premierleague, manchester city, mancity, liverpool

Event dates: Aug. 23, 2010, 6:50 p.m. - Aug. 23, 2010, 9:10 p.m.

Message Frequency



Tweet Map



Relevant Tweets

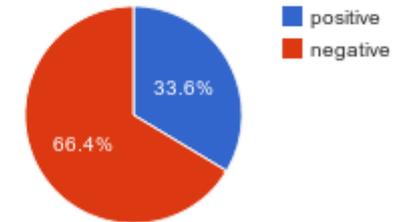
-  Tevez... Nets it man! ManCity 3-0 Liverpool... #mogbe
-  Man City 3-0 Liverpool (Tevez pen)
-  penalty for City. And Tevez makes it 3-0 for ManCity
-  Mancity vs livpool 3-0 tevez gooolllllll
-  BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66') <http://twurl.nl/knkx4l>

- Enter a filter term
- G. liverpool, city, 3-0, kalah, tevez [Label this event](#)
2010-8-23
 - F. liverpool, city, tevez, mancity, 3-0 [Label this event](#)
2010-8-23
 - E. liverpool, city, mancity, 2-0, man [Label this event](#)
2010-8-23

Popular Links

- <http://bit.ly/aLlVUw> (cited by 5)
- <http://bit.ly/cxBF2H> (cited by 5)
- <http://twitcam.com/1rmka> (cited by 5)

Overall Sentiment

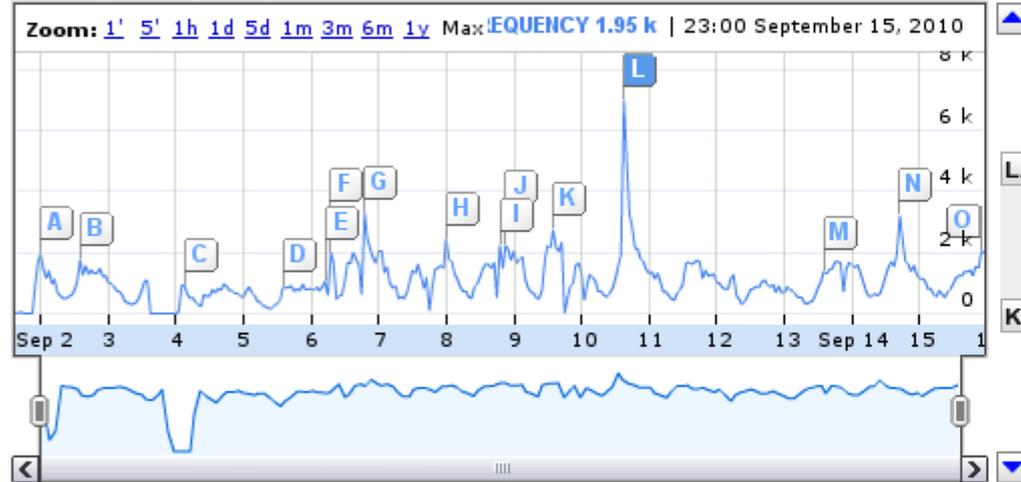


barack obama

Keywords: obama

Event dates: Sept. 1, 2010, 3 p.m. - Sept. 16, 2010, midnight

Message Frequency



Tweet Map



Relevant Tweets

VOTE Goolsbee Said to Be Picked to Head White House Economic Panel: President Barack Obama will appoint Austan Goolsbee... <http://bit.ly/birIPH>

Top Adviser to Lead Panel on Economy - NYTimes.com: President Obama on Friday will promote a longtime economic adv... <http://bit.ly/bklmF1>

WASHINGTON — President Barack Obama has chosen one of

administration
[Label this event](#)

2010-9-13

L. obama, president, economic, barack, economy
[Label this event](#)

2010-9-10

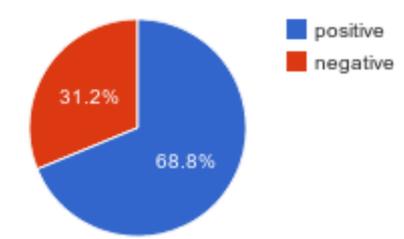
K. obama, burning, koran, quran, burn
[Label this event](#)

2010-9-9

Popular Links

- <http://bit.ly/czJdxP> (cited by 22)
- <http://lat.ms/dcig1f> (cited by 4)
- <http://bit.ly/9XzAyd> (cited by 4)

Overall Sentiment

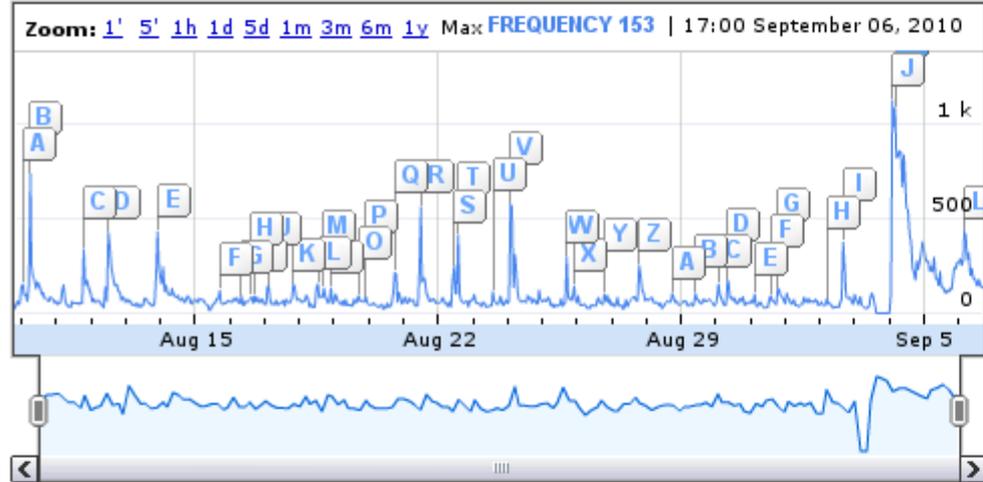


earthquakes

Keywords: earthquake

Event dates: Aug. 9, 2010, 7:18 p.m. - Sept. 6, 2010, 6 p.m.

Message Frequency



- Enter a filter term
- K. earthquake, christchurch, zealand, thoughts, prayers
[Label this event](#)
2010-9-4
 - J. earthquake, christchurch, 7.0, zealand, nz
[Label this event](#)
2010-9-4
 - I. earthquake, 3.3, felt, pablo, san
[Label this event](#)
2010-9-2

Relevant Tweets

RT @mariacelina: This morning, my thoughts & prayers go out to those affected by the earthquake that recently hit Christchurch, New Zealand.

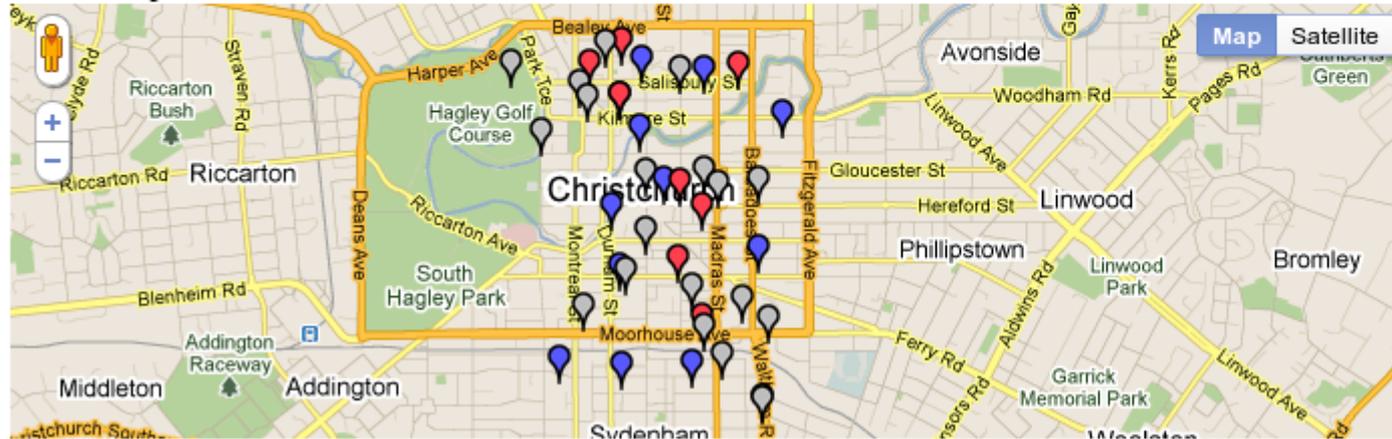
RT @KennyHamilton: Just heard about the earthquake in New Zealand...my thoughts and prayers go out to everyone in that country. I hope ...

RT @EvermoreBand: Thoughts and prayers with the people of Christchurch. Massive earthquake there, and no power or water for the whole city.

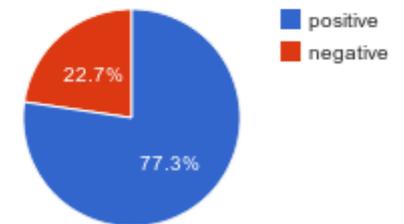
Popular Links

- <http://on.cnn.com/8YDPyG> (cited by 11)
- <http://bit.ly/ddWgtK> (cited by 6)
- <http://bit.ly/aFh7RM> (cited by 5)

Tweet Map



Overall Sentiment



TweeQL Lessons Learned

- Geographic limitations
- Requires grungy regular expressions
- Data is not necessarily relational

