TwitInfo: Aggregating and Visualizing Microblogs for Event Exploration

Adam Marcus, Michael Bernstein, Osama Badar, David Karger, Sam Madden, Rob Miller

MIT CSAIL
Tell stories using Twitter
Identify events of interest
Identify events of interest
Label events
Identify events of interest
Label events
Add context: sentiment? location?
Identify events of interest
Label events
Add context: sentiment? location?
Streaming
TwitInfo automates the process of telling stories using Twitter

Identify events of interest
Label events
Add context: sentiment? location?
Streaming
august 23 manchester city vs. liverpool
Keywords: football, soccer, epl, premier league, premierleague, manchester city, mancity, liverpool

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 gooollllll

BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66&#039;) http://twurl.nl/knxx4l

Popular Links
- http://twitcam.com/1rmka (cited by 5)

Overall Sentiment
- positive: 33.6%
- negative: 66.4%
earthquakes
Keywords: earthquake
Event dates: Aug. 9, 2010, 7:18 p.m. - Sept. 6, 2010, 6 p.m.

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

Overall Sentiment
- Positive: 22.7%
- Negative: 77.3%
TwitInfo

event detection
event labeling
context
streaming algorithms
TwitInfo

event detection
event labeling
context
streaming algorithms

sentiment correction algorithm
evaluation: news consumers
evaluation: pulitzer prize journalist
[Andre et al. UIST2007] [Diakopoulos et al. VAST 2010] [Diakopoulos+Shamma CHI2010] [Dork et al. InfoVis2010] [Go et al. 2010] [Leskovec et al. KDD2009] [Shamma et al. CSCW2010+2011]
[Andre et al. UIST2007] [Diakopoulos et al. VAST 2010]
[Diakopoulos+Shamma CHI2010] [Dork et al. InfoVis2010]
[Go et al. 2010] [Leskovec et al. KDD2009]
[Shamma et al. CSCW2010+2011]
[Andre et al. UIST2007]  
[Diakopoulos et al. VAST 2010]  
[Diakopoulos+Shamma CHI2010]  
[Dork et al. InfoVis2010]  
[Go et al. 2010]  
[Leskovec et al. KDD2009]  
[Shamma et al. CSCW2010+2011]
TwitInfo is a streaming storytelling layer that complements these visualizations.
ui walkthrough
algorithm design
study/evaluation
discussion
august 23 manchester city vs. liverpool
Keywords: football, soccer, epl, premier league, premierleague, manchester city, mancity, liverpool

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 goooooo!

Popular Links

- http://twitcam.com/1rmka (cited by 5)

Overall Sentiment

66.4% positive
33.6% negative
august 23 manchester city vs. liverpool
Keywords: football, soccer, epl, premier league, premierleague, manchester city, mancity, liverpool

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 goooollllllll
BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66&#039;) http://twurl.nl/knxx4l

Popular Links

http://bit.ly/all_vUw (cited by 5)
http://twitcam.com/1rmka (cited by 5)

Overall Sentiment

66.4% positive
33.6% negative
august 23 manchester city vs. liverpool

Keywords: football, soccer, epl, premier league, premierleague, manchester city, mancity, liverpool

Relevant Tweets

- Tevez... Nets it man! ManCity 3-0 Liverpool... #mogbe
- Man City 3-0 Liverpool (Tevez open)
- penalty for City. And Tevez makes it 3-0 for ManCity
- Mancity vs livpool 3-0 tevez goooollllll

Popular Links

- http://twicam.com/1rmka (cited by 5)

Overall Sentiment
august 23 manchester city vs. liverpool
Keywords: football, soccer, epl, premier league, premierleague, manchester city, mancity, liverpool

Related 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 goooolllll
- BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66#039;) http://twurl.nl/knxx4l

Popular Links
- http://twitcam.com/1rmka (cited by 5)

Overall Sentiment
- Positive: 33.6%
- Negative: 66.4%
august 23 manchester city vs. liverpool
Keywords: football, soccer, epl, premier_league, premierleague, manchester city, mancity, liverpool

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&#039;) http://twurl.nl/knkx4l

Popular Links
- http://twitcam.com/1rmka (cited by 5)

Overall Sentiment
- 33.6% positive
- 66.4% negative
august 23 manchester city vs. liverpool
Keywords: football, soccer, epl, premier league, premierleague, manchester city, mancity, liverpool

Message Frequency

Relevant Tweets
- Tevez... Nets it man! ManCity 3-0 Liverpool... #mogbe
- Man City 3-0 Liverpool (Tevez hpen)
- penalty for City. And Tevez makes it 3-0 for ManCity
- Mancity vs livpool 3-0 tevez gooollllll!
- BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66#039;) http://twurl.nl/lnkxel

Popular Links
- http://twitcam.com/1rmka (cited by 5)

Overall Sentiment
ui walkthrough
algorithm design
study/evaluation
discussion
Large
Relatively Large
Relatively Large Streaming
Relatively Large

Streaming mean deviation
Relatively Large

Adaptive Streaming

mean deviation
Relatively Large

Adaptive Streaming

Exponentially weighted moving mean deviation
RFC 2988: Computing TCP’s retransmission timer

Adaptive Streaming

Exponentially weighted moving mean deviation

Message Frequency

Relatively Large
RFC 2988: Computing TCP’s retransmission timer

Adaptive Streaming

Exponentially weighted moving mean deviation

TF-IDF

Relatively Large
2 datasets: earthquakes, soccer games

Precision: 80-100%
Recall: 80-100%
Negative Tweets
(500 tweets)

Classified as Negative
(50 tweets)
Negative Tweets (500 tweets)
Classified as Negative (50 tweets)

Positive Tweets (500 tweets)
Classified as Positive (250 tweets)
Negative Tweets
(500 tweets)

Classified as Negative
(50 tweets)

Positive Tweets
(500 tweets)

Classified as Positive
(250 tweets)

Uncorrected Output:

20%

80%
MAYBE CUT!
ui walkthrough
algorithm design
study/evaluation
discussion
12 participants: four female/8 male
first half: directed tasks
second half: 5 min write article (soccer, Obama)

how do they use the interface?
what can they learn from TwitInfo?
“After the peace talks, Obama traveled to the ASEAN conference and to NATO to work on issues in those parts of the world. He then spent the week dedicated to domestic economic issues. First he proposed a research tax break, then a $50 billion investment in infrastructure, then the issue came up about whether he should keep some tax breaks that Bush had implemented, and he’s asking for some tax breaks from business, and these are generating some controversy because [. . . ]”
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 gooolllll
- BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66&#039;) http://twurl.nl/knpx4l

Popular Links

- http://twitcam.com/1rmka) (cited by 5)
Everyone
Everyone Split

Relevant Tweets
Tevez... Nets it man! ManCity 3-0 Liverpool... #mogbe
Man City 3-0 Liverpool (Tevez

/. And Tevez makes City
pool 3-0 tevez

COM - Update:
--------- ...y - Liverpool 3-0
(Tevez 66&#039;) http://twurl.nl/knx4l

Popular Links
http://bit.ly/allVUw (cited by 5)
http://twitcam.com/lmka) (cited by 5)
Everyone

Split

News Junkies

Relevant Tweets

Tevez... Nets it man! ManCity 3-0 Liverpool... #mogbe

Man City 3-0 Liverpool (Tevez)

/. And Tevez makes City

pool 3-0 tevez

COM - Update:

------------------

y - Liverpool 3-0

(Tevez 66)&#039;) http://twurl.nl/knx4l

Popular Links


earthquakes
“my thoughts and prayers go out to everyone in that country. I hope…”
Journalist Interview
Journalist Interview

backgrounding

barack obama

Keywords: obama

Message Frequency
Journalist Interview

backgrounding  eyewitnesses

barack obama
Keywords: obama

Message Frequency
ui walkthrough
algorithm design
study/evaluation
discussion
appropriate algorithms for the medium
appropriate algorithms for the medium

other uses for streaming algorithms/signal processing
appropriate algorithms for the medium
other uses for streaming algorithms/signal processing

sentiment: caution!
appropriate algorithms for the medium
other uses for streaming algorithms/signal processing
sentiment: caution!
mixed-initiative storytelling
TwitInfo

http://twitinfo.csail.mit.edu

marcua@csail.mit.edu / @marcua

Adam Marcus, 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
How can we tell stories with the tweet stream?
TwitInfo

How can we tell stories with the tweet stream?
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”

data:

- location=Phoenix, temperatureC=89
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
Unstructured data
Unstructured data

+

Structured queries
Unstructured data

+ Structured queries

=> Structured data
Structured queries

Unstructured data + Structured data => Meaningful visualizations
Thanks!

- TweeQL: http://github.com/marcua/tweeql
- 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

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 gooollllll
BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66#039;) http://twurl.nl/knxx41

Popular Links

http://bit.ly/All_VuW (cited by 5)
http://twitcam.com/1rmka (cited by 5)

Overall Sentiment

[Diagram showing sentiment percentages]
barack obama
Keywords: obama

Message Frequency

Tweet Map

Relevant Tweets
- Goolsbee Said to Be Picked to Head White House Economic Panel: President Barack Obama will appoint Austan Goolsbee...
- Top Adviscr to Lead Panel on Economy - NYTimes.com: President Obama on Friday will promote a longtime economic adv...
  http://bit.ly/bklnF1
- WASHINGTON — President Barack Obama has chosen one of...

Popular Links
- http://lat.ms/dcg1f (cited by 4)

Overall Sentiment

positive 31.2%
negative 68.8%
earthquakes
Keywords: earthquake
Event dates: Aug. 9, 2010, 7:18 p.m. - Sept. 6, 2010, 6 p.m.

Message Frequency

Relevant Tweets

- @mariacelina: This morning, my thoughts & prayers go out to those affected by the earthquake that recently hit Christchurch, New Zealand.
- @KennyHamilton: Just heard about the earthquake in New Zealand...my thoughts and prayers go out to everyone in that country. I hope ...
- @EvermoreBand: Thoughts and prayers with the people of Christchurch. Massive earthquake there, and no power or water for the whole city...

Popular Links


Overall Sentiment

- Positive: 22.7%
- Negative: 77.3%
TweeQL Lessons Learned

- Geographic limitations
- Requires grungy regular expressions
- Data is not necessarily relational
Behind the Hunt for Bin Laden

President Obama and members of his national security team receiving an update on Sunday. A classified document in front of Hillary Rodham Clinton was blurred before this photo was released. More Photos »
Behind the Hunt for Bin Laden

President Obama and members of his national security team receiving an update on Sunday. A classified document in front of Hillary Rodham Clinton was blurred before this photo was released. More Photos »
Behind the Hunt for Bin Laden

President Obama and members of his national security team receiving an update on Sunday. A classified document in front of Hillary Rodham Clinton was blurred before this photo was released. More Photos »

Reporting was contributed by Elisabeth Bumiller, Charlie Savage and Steven Lee Myers from Washington, Adam Ellick from New York, and Ismail Khan from Peshawar, Pakistan.
Behind the Hunt for Bin Laden

President Obama and members of his national security team receiving an update on Sunday. A classified document in front of Hillary Rodham Clinton was blurred before this photo was released. More Photos »

Reporting was contributed by Elisabeth Bumiller, Charlie Savage and Steven Lee Myers from Washington, Adam Ellick from New York, and Ismail Khan from Peshawar, Pakistan.
Behind the Hunt for Bin Laden

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

3-0! Mancity vs liverpool

gool de carlos tevez para el manchester city 3-0 liverpool

Tevez scores and City now has a very commanding 3-0 lead over Liverpool
Behind the Hunt for Bin Laden

Man City 3-0 Liverpool (Tevez pen)

penalty for City. And Tevez makes it 3-0 for ManCity

Mancity vs livpool 3-0 tevez gooooollll!!!

BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66') http://twurl.nl/knkx4l

3-0! Mancity vs liverpool

gool de carlos tevez para el manchester city 3-0 liverpool

Tevez scores and City now has a very commanding 3-0 lead over Liverpool
markng Mark Ng
great way to start my morning. Kittens : Check ; #yes2av : Check - http://www.youtube.com/watch?v=HlHiDD_oTk
2 minutes ago

newsyc20 Hacker News 20
Browser tabs should indicate which pages are playing audio
http://news.ycombinator.com/item?id=2509624 #active
3 minutes ago

PaulLewis Paul Lewis  by markng
Ian Tomlinson’s reaction: verdict was "statement of the blindingly obvious" gu.com/p/2zz50tw
10 minutes ago

derekcollison Derek Collison  by antirez
The #cfoundry community has been amazing. Thanks.
http://fj.mp/IP6qea
7 minutes ago

markng Mark Ng
nice to see that the UK justice system seems to be working correctly around the Tomlinson police killing.
7 minutes ago

newsyc20 Hacker News 20
Getting Started with Android Development http://bit.ly/mgS4Eo
(http://bit.ly/mnpBZz) via @razin #guru
8 minutes ago

NiemanLab Nieman Lab
BBC is not more Google friendly; thanks to transitists of more
Tweets Per Second: Evening of May 1, 2011

Event annotation

- White House announces Pres. Obama speech
- Keith Urbahn tweets about Osama rumor
- ABC, NBC & CBS report Bin Laden's death
- Pres. Obama begins remarks
- Pres. Obama confirms Bin Laden's death
- Pres. Obama concludes remarks

Streaming
Again, glad UBL dead, but this jubilation troubles me. The dead are still dead. We ought strive for a world where killing isn't necessary.
Again, glad UBL dead, but this jubilation troubles me. The dead are still dead. We ought strive for a world where killing isn't necessary.

Do we have an idea how many tweets were from within the US vs. elsewhere last night? Otherwise twitter users <> US network compare flawed