SoLoGlo
Collecting, Archiving, and Linking Social, Local, and Global News

Martin Klein @mart1nkleon
Peter Broadwell @peterbroadwell
Todd Grappone @liber8er
Sharon Farb @farbthink
Social  Local  Global
A Brief History of Timeliness

• Twitter archive at the Library of Congress [1]
  • Last public update from January 4\textsuperscript{th} 2013
  • \textasciitilde{}170 billion tweets, \textasciitilde{}130 TB compressed (late 2012)
  • Single search against 2006-2010 data may take up to 24 hours

• Twitter data access at Massachusetts Institute of Technology, Laboratory for Social Machines [2]
  • Public announcement from October 1\textsuperscript{st} 2014

A Brief History of Timeliness

In case you missed it:

- Twitter makes full archive of tweets available, indexed
- Great, problem solved?
  - No API-level access!
  - How about deleted tweets?
  - Embedded resources change!

[Link to article](https://blog.twitter.com/2014/building-a-complete-tweet-index)
We Can Remember It for You Wholesale

I. Real-time capture of tweets plus rapid analysis

II. Pro-active archiving of tweets, referenced web resources, and media

III. Collection-agnostic linking
Part 1a: Real-Time Capture of Tweets

- Utilize GWU’s Social Feed Manager
  - Filter tweets by keywords, user handles, location
  - Store raw tweets
  - Extract and archive embedded URIs, media
    - Utilize pro-active archiving solution: archive.today
Part 1b: Rapid, Adaptive Analysis
Remembrance of Tweets/Links Past

- UCLA’s dataset about Egyptian revolution in 2011
  - More than 400k tweets
  - Approx. 50k unique users
  - Tweets originated from within 200 miles around Cairo

Tweet by Mohamed Effat:

As i said Police presence will provoke people, and as soon as clashes erupt expect bloodsheds and massacres on the streets of Cairo #jan25

2:51 PM - 30 Jan 2011

13 RETWEETS 1 FAVORITE
Remembrance of Tweets/Links Past

• UCLA’ s dataset about Egyptian revolution
  • 25% of tweets contain references to external resources (web pages, images, videos, etc)
After days of looting, once-vibrant Cairo smolders

BY HANNAH ALLAM
McClatchy Newspapers  January 30, 2011

CAIRO — Egypt's vast popular revolt has transformed this cosmopolitan capital into an urban battlefield, where the smell of smoke hangs in the air and residents walk though ransacked shopping districts in search of bread.

On Sunday, the roar of fighter jets overhead drowned out the chants of thousands of protesters, who once again defied curfew to gather downtown in hopes of toppling President Hosni Mubarak, the U.S.-allied authoritarian whose regime many Egyptians already refer to in past tense.

As Mubarak clings to his office, and the military attempt to restore order, Egyptians are taking stock of the toll so far: At least 150 dead nationwide, hundreds injured in riots, millions of dollars in property damage and a security vacuum filled by ragtag neighborhood watch groups.

RELATED STORIES:
- Egypt's Mubarak on the brink; ElBaradei demands 'new era'
- Why has Egypt's army not confronted protests?
- Clinton calls for 'peaceful, orderly transition' in Egypt
- Egyptian upheaval a 'complete surprise' to Israel
- Will Mubarak heed Obama's call to listen to protesters?

Remembrance of Tweets/Links Past

- UCLA’s dataset about Egyptian revolution
  - 20% of references are dead, after less than 4 years (!!!)
Remembrance of Tweets/Links Past

http://yfrog.com/h02gvclj

→ HTTP GET
← 200 OK

→ HTTP HEAD
← 204 No Content
Remembrance of Tweets/Links Past

• UCLA’s dataset about Egyptian revolution
  • 20% of references are dead AND
  • 60% of these are not archived
This one is!

discovered via #memento
## Part 2: Pro-Active Archiving

### Archived tweets, URIs, and embedded media for #pasig2015

<table>
<thead>
<tr>
<th>Date</th>
<th>Tweet ID</th>
<th>Embedded URI</th>
<th>Embedded Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-03-11T02:01:39</td>
<td>575476647693979649</td>
<td><a href="https://libraries.ucsd">https://libraries.ucsd</a>...</td>
<td><a href="http://pbs.twimg.com/media/">http://pbs.twimg.com/media/</a>...</td>
</tr>
</tbody>
</table>

Part 3: Collection-Agnostic Linking

AND YOU START TO GET YOUR CELL PHONE AND AS YOU WAIT FOR YOUR FRIENDS, NOW FIRST THING ON ANYBODY'S MIND IS GETTING SOMETHING TO EAT. SO OUR GROUP OF RUNNERS, WE HAD STARTED WALKING DOWN THE STREET TO GRAB A QUICK BITE. WE SAT OUTSIDE, HEARD WHAT SOUNDED LIKE A CANON, AND ONE OF THE GUYS SAID, THAT SOUNDED LIKE AN EXPLOSION. AND I SAID, I FIND THAT HARD TO BELIEVE. I COULDN'T IMAGINE -- IT IS PATRIOT'S DAY, MAYBE THEY JUST HAVE A CELEBRATION. AND AS VERNON MENTIONED, VERY QUICKLY, THE SIRENS, FIRE TRUCKS, UNMARKED POLICE CARS, POLICE CARS, STATE TROOPERS, AMBULANCES AND QUICKLY WE JUST TOOK TO TWITTER AND TO THE INTERNET TO FIND OUT WHAT HAPPENED, AS WELL AS STARTED RECEIVING TEXT MESSAGES AND CALLS. AND IT BECAME PRETTY CHAOTIC VERY QUICKLY. >> EVERYONE, AN UPDATE COMING ACROSS THE WIRES RIGHT NOW. POLICE SEARCHING FOR A PINSKY TRUCK THAT TRIED TO GAIN ACCESS TO THE PAR THON ROUTE, JUST BEFORE THE PALAST, BUT WAS TURNED AWAY. IN ADDITION TO RYAN McGRATH, WHO HEARD THE EXPLOSION, VERNON LOEB ALSO WITH US, WITH JOHN DIPETRO. WITH ME RIGHT NOW, REPORTER WITH "THE PATRIOT LEDGER," JACK
Part 3: Collection-Agnostic Linking
A Brief History of Timeliness

Social Feed Manager (Dan Chudnov, GWU)

Django administration

Site administration

Auth
Groups  ➤ Add  ➤ Change
Users  ➤ Add  ➤ Change

Sites
Sites  ➤ Add  ➤ Change

Social Auth
Associations  ➤ Add  ➤ Change
Nonces  ➤ Add  ➤ Change
User social auths  ➤ Add  ➤ Change

UI
Twitter filters  ➤ Add  ➤ Change
Twitter user item urls  ➤ Add  ➤ Change
Twitter user items  ➤ Add  ➤ Change
Twitter user sets  ➤ Add  ➤ Change
Twitter user timeline errors  ➤ Add  ➤ Change
Twitter user timeline jobs  ➤ Add  ➤ Change
Twitter users  ➤ Add  ➤ Change

now with about 399 items from 1 users

a sampling:

tweets per day, past month

recent tweets

<table>
<thead>
<tr>
<th>user</th>
<th>date</th>
<th>rt</th>
<th>text</th>
</tr>
</thead>
<tbody>
<tr>
<td>martinkle1n</td>
<td>14:09:34</td>
<td>71</td>
<td>RT @livesToomas: Oh boy...“Russia plans alternative version of ‘Wikipedia’ <a href="http://t.co/RjxKo8BYg">http://t.co/RjxKo8BYg</a> via @reuters</td>
</tr>
<tr>
<td>martinkle1n</td>
<td>11:35:05</td>
<td>1</td>
<td>RT @leghmonty: There are some fascinating projects on mapping revolutions w/data, more @UCLA Libraries, tx @martinkle1n @PeterBroedel #…</td>
</tr>
<tr>
<td>martinkle1n</td>
<td>09:35:51</td>
<td>0</td>
<td>Snow and 3 degrees in DEN may prove rather destructive for travel back from excellent #savemoney workshop #hopetomonday</td>
</tr>
<tr>
<td>martinkle1n</td>
<td>23:07:34</td>
<td>0</td>
<td>@vphill @phonedebug_min Absolutely! @palewire is equally excited about this. We’ll get this done! #memento</td>
</tr>
<tr>
<td>martinkle1n</td>
<td>16:54:10</td>
<td>0</td>
<td>@Educopia’s Katherine Skinner calls for action in preserving web-based news, proactive <a href="http://t.co/Niffy2bzkB">http://t.co/Niffy2bzkB</a> can be great partner #savemoney</td>
</tr>
</tbody>
</table>

http://social-feed-manager.readthedocs.org/
A Brief History of Timeliness

twarc (Ed Summers, MITH); used for Ferguson data

#saa13 tweets - Archives 2013

created on the command line with twarc from the Twitter JSON

Simms Toomey
simms.toomey
@SNAP_Roundtable why, yes, I am planning to live-tweet during #saa13! :) #naptr
Sun Aug 11 04:09:23 +0000 2013

Terry Baxter
terryx666
At PDX and heading out to #saa13. Better take a nap!
Sun Aug 11 12:06:34 +0000 2013

Sam Crisp
Archivally_Sam
At last, finalized my schedule for #saa13. @ArchivistKira would love to join you @ Sazerac Bar Tuesday night if you're still planning to go.
Sun Aug 11 12:12:38 +0000 2013

Angelique
RandomArchivist
#ATL -> #MSY for #SAA13
Sun Aug 11 12:54:24 +0000 2013

Angelique
RandomArchivist
There is a segment on @CNN on the Levi #archives. Must be fate! #SAA13
Sun Aug 11 13:24:17 +0000 2013

Arian Ravanbakhsh
adravan
NOLA and #saa13 here I come! (@ Baltimore / Washington International Thurgood Marshall Airport (BWI) w/ 44 others) http://t.co/otzGGhxc3VE
Sun Aug 11 13:30:13 +0000 2013

Rachel Donahue
sheeppeh
I really need to take a class on Anime Studio (or Blender) before I decide it would be "fun" to animate something again. z.Zz #saa13 #301
Sun Aug 11 13:40:47 +0000 2013

John D. Bence
jdbence
Here I come, #saa13! Hartsfield-Jackson Atlanta International Airport (ATL) http://t.co/4KZKnpNZTz
Sun Aug 11 13:51:46 +0000 2013

Eira Tansey
eiratansey
please RSVP for #saa13 tweetup by this afternoon so I can send a final head count to our venue http://t.co/NcYc4mHvQg
Sun Aug 11 14:06:18 +0000 2013


http://inkdroid.org/journal/2014/08/30/a-ferguson-twitter-archive/

A Ferguson Twitter Archive

If you are interested in an update about where/how to get the data after reading this see here.

Much has been written about the significance of Twitter as the recent events in Ferguson echoed round the Web, the country, and the world. I happened to be at the Society of American Archivists meeting 5 days after Michael Brown was killed. During our panel discussion someone asked about the role that archivists should play in documenting the event.
Remembrance of Tweets/Links Past

URIs from Ed Summer’s Ferguson dataset

pink == not archived

(Internet Archive) 28%

https://edsu.github.io/ferguson-urls/