

IDEAL Tweet Collection Categorization

May 3, 2015

CS 4624 Multimedia, Hypertext and Information Access

Virginia Tech

Blacksburg, VA 24061

Team member: Stephen Won

Client: Sunshin Lee



























Background, Goal and Approach

- 1.1 billion tweets archived by April 2016
 - Goal
 - Aid IDEAL team to manage > 1000 collections
 - Approach
 - Interface for collection management
 - Tagging/Categorizing collections

GUI - table

- Create interactive table
 - filter the collections from the search box
 - will enable users to use column ordering.

Show entries Search:

ARCHIVE ID	KEYWORD / HASHTAG	DESCRIPTION	TAGS	SCREEN NAME	COUNT	CREATE TIME	
1	#egypt	Tweets for Egyptian revolution		dirf	690	Sun, 07 Oct 2012 18:28:33 -0400	 
6	#yemen			dirf	87	Sun, 07 Oct 2012 18:29:26 -0400	 
7	#datatable	datatable created	ytk, datatable	step63n1	0	Sat, 02 Apr 2016 12:14:11 -0400	 
8	keyword	description	tags	screen_name	0	Wed, 31 Dec 1969 19:00:00 -0500	 
9	OccupyWallStreet			dirf	68	Sun, 07 Oct 2012 18:31:26 -0400	 
10	#nv	new river valley (blacksburg) related tweets		dirf	313	Sun, 07 Oct 2012 18:32:04 -0400	 
11	virginia tech			dirf	99	Sun, 07 Oct 2012 18:32:26 -0400	 
12	iran earthquake			dirf	0	Sun, 07 Oct 2012 18:33:15 -0400	 
13	diabetes	health category		dirf	0	Sun, 07 Oct 2012 18:33:37 -0400	 
14	heart attack	health category		dirf	0	Sun, 07 Oct 2012 18:33:56 -0400	 
15	foursquare			dirf	0	Sun, 07 Oct 2012 18:34:18 -0400	 
16	#isaac	hurricane Isaac in Aug. 2012		dirf	0	Sun, 07 Oct 2012 18:34:48 -0400	 
17	turkey syria	violence between Turkey and Syria in Oct. 2012		dirf	0	Sun, 07 Oct 2012 18:35:27 -0400	 

GUI - Categories Added

Man-Made Disaster

Natural Disaster

Man-Made Disaster

Shooting

school shooting

mall shooting

Human Failure

Social

Revolution

Protest


Bombing



IDEAL Project: Tweet Archive DB

Total number of archived tweets: **1,077,198,895**

Archive ID	Keyword / Hashtag	Description	Tags	Screen Name	Count	Create Time
1	#egypt	Tweets for Egyptian revolution		sslee777	13,117,562	07/10/2012
2	#libya			dlrl	2,739,565	07/10/2012
3	#blacksburg			dlrl	195,746	07/10/2012
4	#jan25			dlrl	1,127,028	07/10/2012
5	#bahrain			dlrl	21,947,874	07/10/2012
6	#yemen			dlrl	3,123,394	07/10/2012
7	japan earthquake			dlrl	1,220,333	07/10/2012
8	#syria			dlrl	16,813,409	07/10/2012
9	OccupyWallStreet			dlrl	957,178	07/10/2012
10	#nrv	new river valley (blacksburg) related tweets		dlrl	174,768	07/10/2012
11	virginia tech			dlrl	1,654,085	07/10/2012
12	iran earthquake			dlrl	292,759	07/10/2012
13	diabetes	health category		dlrl	10,241,376	07/10/2012
14	heart attack	health category		dlrl	17,645,612	07/10/2012
15	foursquare			dlrl	43,482,879	07/10/2012







 Sign in with Twitter

Your TwapperKeeper

archive your own tweets

Show entries

Search:

Archive ID	Keyword / Hashtag	Description	Tags	Screen Name	Count	Create Time
636	Wayne Community College	Shooting		ideal_vt	18	Mon, 13 Apr 2015 10:28:52 -0400 
505	Taliban	Shooting		ideal_vt	98	Tue, 16 Dec 2014 09:23:55 -0500 
680	South Carolina church shooting	Shooting		ideal_vt	66	Thu, 18 Jun 2015 00:11:40 -0400 
502	oregon school shooting	Shooting		ideal_vt	16	Mon, 15 Dec 2014 13:36:37 -0500 
473	Montreal ecole Polytechnique	Shooting		ideal_vt	2	Fri, 05 Dec 2014 21:07:47 -0500 
530	missouri shooting	shooting		ideal_vt	94	Thu, 25 Dec 2014 22:18:42 -0500 

Sign in with Twitter


Your TwapperKeeper

archive your own tweets

Show 25 entries

Search:

ARCHIVE ID	KEYWORD / HASHTAG	DESCRIPTION	TAGS	SCREEN NAME	COUNT	CREATE TIME	
1	#egypt	Tweets for Egyptian revolution		dirf	690	Sun, 07 Oct 2012 18:28:33 -0400	
6	#yemen			dirf	87	Sun, 07 Oct 2012 18:29:26 -0400	
7	#datatable	datatable created	ytk, datatable	step63n1	0	Sat, 02 Apr 2016 12:14:11 -0400	
8	keyword	description	tags	screen_name	0	Wed, 31 Dec 1969 19:00:00 -0500	
9	OccupyWallStreet			dirf	68	Sun, 07 Oct 2012 18:31:26 -0400	
10	#nrv	new river valley (blacksburg) related tweets		dirf	313	Sun, 07 Oct 2012 18:32:04 -0400	
11	virginia tech			dirf	99	Sun, 07 Oct 2012 18:32:26 -0400	
12	iran earthquake			dirf	0	Sun, 07 Oct 2012 18:33:15 -0400	
13	diabetes	health category		dirf	0	Sun, 07 Oct 2012 18:33:37 -0400	
14	heart attack	health category		dirf	0	Sun, 07 Oct 2012 18:33:56 -0400	
15	foursquare			dirf	0	Sun, 07 Oct 2012 18:34:18 -0400	
16	#isaac	hurricane Isaac in Aug. 2012		dirf	0	Sun, 07 Oct 2012 18:34:48 -0400	
17	turkey syria	violence between Turkey and Syria in Oct. 2012		dirf	0	Sun, 07 Oct 2012 18:35:27 -0400	

 Sign in with Twitter

Your TwapperKeeper











archive your own tweets

[Man-Made Disaster](#)

[Natural Disaster](#)

Show entries

Search:

ARCHIVE ID	KEYWORD / HASHTAG	DESCRIPTION	TAGS	SCREEN NAME	COUNT	CREATE TIME	
1	#hellegypt	Tweets for Egyptian revolution		dlri	690	Sun, 07 Oct 2012 18:28:33 -0400	
2	#aalasibya			dlri	486	Sun, 07 Oct 2012 18:28:52 -0400	
3	#blacksburg			dlri	500	Sun, 07 Oct 2012 18:29:01 -0400	
4	#jan25			dlri	1465	Sun, 07 Oct 2012 18:29:08 -0400	
5	#bahrain			dlri	67	Sun, 07 Oct 2012 18:29:18 -0400	
6	#yemen			dlri	87	Sun, 07 Oct 2012 18:29:26 -0400	
7	japan earthquake			dlri	94	Sun, 07 Oct 2012 18:29:37 -0400	
8	#syria			dlri	1151	Sun, 07 Oct 2012 18:30:00 -0400	
9	OccupyWallStreet			dlri	68	Sun, 07 Oct 2012 18:31:26 -0400	
10	#nrv	new river valley (blacksburg) related tweets		dlri	313	Sun, 07 Oct 2012 18:32:04 -0400	

Showing 1 to 10 of 667 entries

Previous 2 3 4 5 ... 67 Next

Man-Made Disaster

Shooting

school shooting

mall shooting

Human Failure

Social

Revolution

Protest

War

Terrorism

Fire

Bombing

Civil War

Epidemic

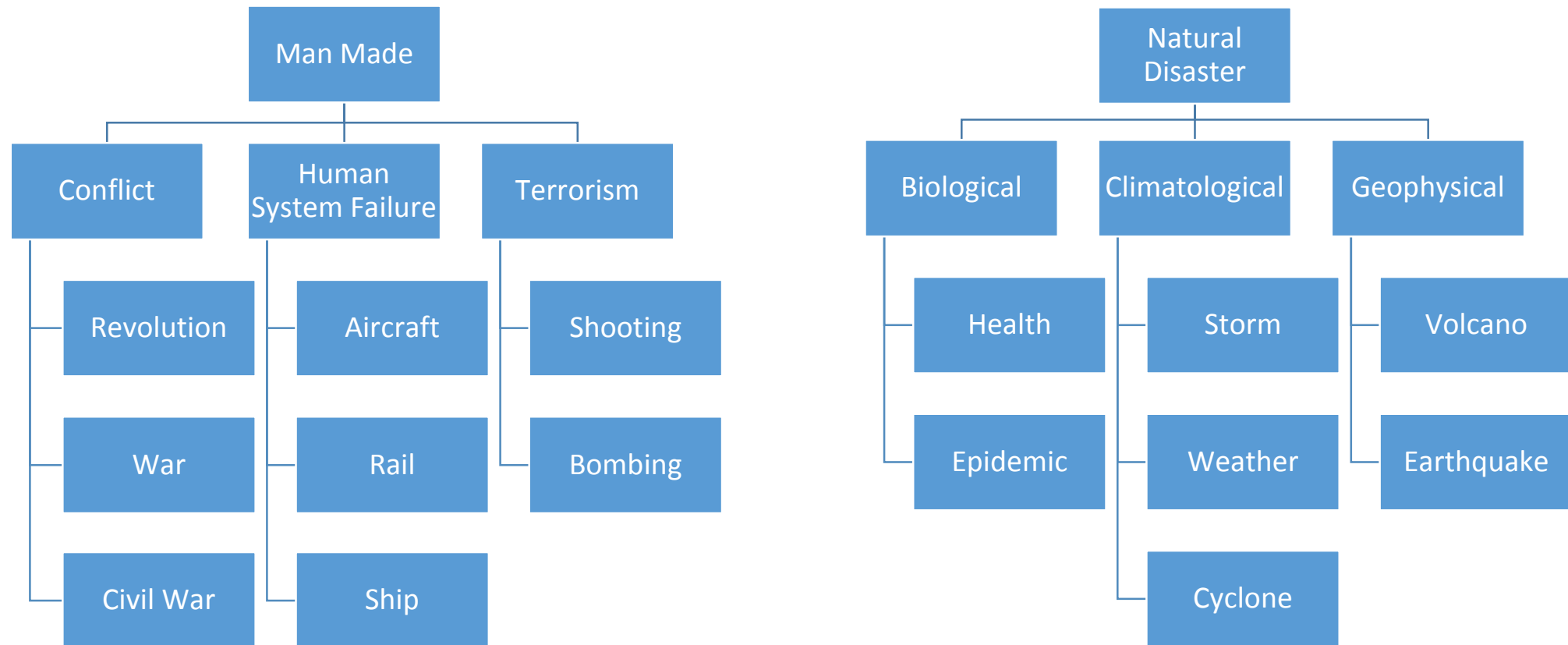
Suicide

Disorder

Prison Break

Categorization - taxonomy

- Top levels of taxonomy of 2 sub-trees for disasters/crises



Tagging

- Use tag system for event type, place and date.

Event Type	Types of event
Place	Place where the event happened, or related places of the event
Date	Date the event has occurred

Search on Place Tag



Search tag
& keyword

Place | USA

Keyword	Event type	Place	Date
Hurricane sandy	Hurricane	Carolinas, Virginia, Washington D.C, USA ...	October, November, 2012
Hurricane issac	Hurricane	Arkansas, Mississippi, Alabama, Louisiana, USA ...	August, September, 2012
Oregon shooting	Shooting	Clackamas, Oregon, USA	December, 2012
Hurricane	Hurricane		
#Tucson shooting	Shooting	Tucson, Arizona, USA	January, 2011
...			

Tagging with Taxonomy

Event Type

- General Event
- More Specific Event
- Even More Specific Event

Place

- Country
- State
- City

Date

- Year
- Month
- Day

1	id	keyword	description	event type	place	date	screen_na
26	25	flood		Flood			drl
27	26	terrorism		Terrorism			drl
28	27	hurricane		Hurricane			drl
29	28	hurricane isaac		Hurricane	Puerto Rico, Hispaniola, C	August, Septemb	drl
30	29	@NOAA	for Hurricane Sandy	Hurricane	Greater Antilles, Bahamas	October, Novem	drl
31	30	@FEMA	for Hurricane Sandy	Hurricane	Greater Antilles, Bahamas	October, Novem	drl
32	31	@RedCross					drl
33	32	@SalvationArmy					drl
34	33	@SalvationArmyUS					drl
35	34	@ReadydotGov	Be informed, make a plan, build a kit.				drl
36	35	@craigatFEMA	FEMA administrator				drl
37	36	guatemala earthquake		Earthquake	Guatemala		drl
38	37	Israel	Israel and Hamas conflict over	War	Gaza, Israel, Egypt, Hamas	2006, ongoing	drl
39	38	obesity		Health			drl
40	39	typhoon		Typhoon			drl
41	40	oregon shooting	A shooting spree in a mall at	Shooting	Clackamas, Oregon, USA	December, 2012	drl
42	41	Connecticut shooting	Shooting at an elementary sc	Shooting	Sandy Hook, Connecticut,	December, 2012	drl
43	42	connecticut school shooting		Shooting	Sandy Hook, Connecticut,	December, 2012	drl
44	43	firefighter shooting	four firefighters were sniped	Shooting		Decmeber, 2012	drl
45	44	tucson shooting	Archiving tweets for the 2nd	Shooting	Tucson, Arizona, USA	January, 2011	drl
46	45	kentucky shooting	2 dead, 1 injured in Kentucky	Shooting	Kentucky, USA	2013	drl
47	46	gun control					drl
48	47	brazil nightclub fire	27-Jan-13	Fire	Santa Maria, Brazil, South	January, 2013	drl
49	48	tsunami		Tsunami			drl

keyword	description	Event type1	Event type2	Event type3	Country	State	City	Year	Month	Day
#egypt	Tweets for Egyptian revolution	Man-Made Disaster	Revolution		Egypt			2011		
#libya					Libya					
#blacksburg					USA	Virginia	Blacksburg			
#jan25										
#bahrain					Bahrain, Arabian Peninsula					
#yemen					Yemen					
japan earthquake		Natural Disaster	Earthquake		Japan					
#syria					Syria					
OccupyWallStreet		Man-Made Disaster	Protest		USA	New York	New York City	2011	September	17
#nrv	new river valley (blacksburg) related tweets				USA	Virginia	New River Valley, Blacksburg			
virginia tech					USA	Virginia	Blacksburg			
iran earthquake		Natural Disaster	Earthquake		Iran					
diabetes	health category	Natural Disaster	Health							
heart attack	health category	Natural Disaster	Health							
foursquare			technology		USA	New York	New York City			
#Isaac	hurricane Isaac in Aug. 2012	Natural Disaster	Climate Change	Hurricane	USA	Louisiana		2012	August, Septemb	
turkey syria	violence between Turkey and	Man-Made Disaster	War		Turkey, Syria					
emergency prepar										
emergency respon										
emergency recove										
emergency mitigat										
emergency manag										
hurricane sandy		Natural Disaster	Climate Change	Hurricane	USA			2012	October, Novem	

Lessons Learned

- Feedback
 - understand the user's needs
 - have clearer goals
- Communication
 - May find easier solutions to problems

Acknowledgements

- **Dr. Edward Fox**

- For all the help, guidance, encouragement and suggestions

- **Sunshin Lee**

- for taking time and guiding throughout the project

- **IDEAL team**

- for valuable feedback

- NSF IIS - 1319578