Social Media Use During Humanitarian Emergencies and Disasters

Alexandra Olteanu

Social Computing + **Computational Social Science**

IBM Science for Social Good

Applied Data Science IBM TJ Watson Research Center



IBM Science for Social Good Initiative Directed by Aleksandra Mojsilovic & Kush Varshney

IBM Science for Social Good Initiative Exploring a new approach in addressing world's challenges



Responsibility at IBM >

Partners IBM Research scientists with academic fellows and subject matter experts from a diverse range of non-governmental organizations (NGOs) to tackle emerging societal challenges using science and technology.



IBM Science for Social Good Initiative Our projects demonstrate the realm of what's possible





Events & Social Media

High volumes of event-related tweets

- 25+ million tweets for Sandy Hurricane
- 24+ million tweets for 2016 Academy Awards
- 10+ million tweets for Supreme Court ruling on marriage equality
- 4+ million tweets about UK 2015 elections

Endorsed by governmental agencies

Social media are to enhance, not replace



Brad Willis @BradWillis

I am halfway across the world & Twitter is

bringing me news of the #LAX shooting faster than CNN's site. This is...the new age of information



Global Pulse 🕗 @UNGlobalPulse

What's the world saying about #climatechange2014? See our new media monitor: ow.ly/BHL6r cc @UNEP @CFigueres @Momentum_UNFCCC



United Nations 📀

Social media is critical for *#humanitarian* work & now you can see why. Crisis Map of #Libya is now public: http://bit.ly/g8xCtm #UN **#OCHA**



Twitter Philippines @twitter ph

#ReliefPH If you are looking to help by volunteering or dropping off donations, you may find a place near you by checking out @ReliefPHcom



MicroMappers, CrisisLex, CREES, AIDR, ...



CREES Crisis Event Extraction Service **Crisis Event Extraction Service (CREES)** Automatic classification and filtering of crisis-related social media posts for situation awarness. Learn more

CrisisLex

Social media is increasingly being used to broadcast useful information during local crisis situations (e.g. hurricanes, earthquakes, protests, accidents, attacks), as well as to discuss global-scale crises (e.g. climate change, pandemics)

CrisisLex.org is a repository of crisis-related social media data and tools. Currently it includes collections of crisis data and a lexicon of crisis terms. It also includes tools to help you create these collections and lexicons.

CREES API Google Sheets Add-on

Download Collections Create a Collection Download Lexicons Home

CrisisLex.org

Collections of Crisis Data »

Lexicons of Crisis Terms »



WORLD HUMANITARIAN DATA AND TRENDS 2014 An annual OCHA publication that presents global

An annual OCHA publication that presents global and country-level data and trend analysis about humanitarian crises and assistance





What kind of information people share during humanitarian crises?

with Carlos Castillo and Sarah Vieweg





- ✓ Twitter sample API
- Keyword-based searches
- \checkmark 26 crisis events
- \checkmark 1000 annotated tweets per crisis

- ✓ Twitter sample API
 - ✓ 2012 & 2013
 - ~ 1% random sample of Twitter public stream
 - \checkmark ~130+ million tweets per month
- Keyword-based searches
- \checkmark 26 crisis events
- ✓ 1000 annotated tweets per crisis

- ✓ Twitter sample API
- Keyword-based searches
- \checkmark 26 crisis events
- \checkmark 1000 annotated tweets per crisis

- ✓ Twitter sample API
- Keyword-based searches
 - ✓ proper names of affected location ✓ manila floods, boston bombings, #newyork derailment
 - ✓ proper names of meteorological phenomena
 - ✓ sandy hurricane, typhoon yolanda
 - ✓ promoted hashtags
 - ✓ #SafeNow, #RescuePH, #ReliefPH
- \checkmark 26 crisis events
- ✓ 1000 annotated tweets per crisis

- ✓ Twitter sample API
- Keyword-based searches
- \checkmark 26 crisis events
- \checkmark 1000 annotated tweets per crisis

- ✓ Twitter sample API
- ✓ Keyword-based searches
- \checkmark 26 crisis events
 - ✓ 14 countries and 8 languages
 - ✓ 12 different hazard types
 - ✓ earthquakes, wildfires, floods, bombings, shootings, etc.
 - ✓ 15 instantaneous crises
- ✓ 1000 annotated tweets per crisis

- ✓ Twitter sample API
- Keyword-based searches
- \checkmark 26 crisis events
- \checkmark 1000 annotated tweets per crisis

- ✓ Twitter sample API
- Keyword-based searches
- \checkmark 26 crisis events
- ✓ 1000 annotated tweets per crisis
 - ✓ Content dimensions
 - ✓ Informativeness
 - ✓ Source of information
 - ✓ Type of information
 - ✓ Crowdsource workers from the affected countries









Data Patterns: Types



Data Patterns: Types







Spain_train_crash'13 Guatemala_earthquake'12 Lac_Megantic_train_crash'13 Savar_building_collapse'13 LA_airport_shootings'13 Costa_Rica_earthquake'12 NY_train_crash'13 West_Texas_explosion'13 Sardinia_floods'13 Venezuela_refinery'12 Bohol_earthquake'13 Glasgow_helicopter_crash'13 Boston_bombings'13 Russia_meteor'13 Brazil_nightclub_fire'13 Typhoon_Yolanda'13 Typhoon_Pablo'12 Queensland_floods'13 Colorado_floods'13 Colorado_wildfires'12 Australia_bushfire'13 Italy_earthquakes'12 Manila_floods'13 Alberta_floods'13 Philipinnes_floods'12 Singapore_haze'13













Instantaneous, focalized, and human-induced

Similar events tend to have a similar distribution of Natural, diffused, and progressive



Sources & Message Types



Infrast. & Utilities	- I		
Caution & Advice			
Donat. & Volun.			
Affected Ind.			
Sympathy	_		
Other Useful Info.			
	Media	Outsiders	



			1
Infrast. & Utilities			
Caution & Advice			
Donat. & Volun.			
Affected Ind.			
Sympathy	_		
Other Useful Info.			
	Media	Outsiders	









Pacific Tsunami Warning Center issues new bulletin, says only Costa Rica, Panama, Nicaragua under tsunami warning bit.ly/REsplq

	•		•	
•		•	•	
•	•	•	•	
•	•	•	•	
•	•	•	•	
•	•		•	
2/1				
JON	48n	 several days		





Infrastructure & Utilities

COSTA RICA: @nacion In Nandayure, per mayor: are no electricity, no water in some parts. There are landslides

•	•		•	•	
•	•		•	•	
•	•		•	•	
•	•		•	•	
36h	48 h	•••	several d	ays	



•••••••••••••••••••••••••••••••••••••••	•	
• • • • •	• • • • • • • • • • • • • • • • • • •	
•	•	 • • • • • • • •
ndividuals		
•	•	
re & Utilities		
	•	 · · · • • • • • • • •

Bangladesh factory building collapse kills over 70, injures hundreds reut.rs/10csR0q

•			-		
•	•		•	•	
•	•		•	•	
•	•		•	•	
•	•		•	•	
•	•		•	•	
36h	4 8h	•••	several day	/S	
			/	•	



	•	• • • • • • • • • • • • • • • • • • • •		• • • • • • • • • • • • • • • • • • • •	•••••••••••••••••••••••••••••••••••••••
		Other Use	ful Infø.	•	
•••••	Affe	ted Individu	als	•••••••••••••••••••••••••••••••••••••••	
			• • • • • • • • • • • • • • • • • • • •	• • • • • • • • • • • •	• • • • • • • • • • • • • • • • • • •
	Infrast	ructure & Ui	ilities	• • • • • • • • • • • • • • • • • • •	• • • • • • • • • • • • • • • • • • •
uppor	No A	larm, Only 1	Exit in Braz	zil Nightclub Fire	
vice				•	
	24h	36h	48h	several da	IYS





•	•		•	•
36h	4 8h	•••	several days	









12h 24h 36

•	•		•	•	
• • • • • • • • • • • • • • • • • • •	• • • • • • • • • • • • • • • • • • •		• • • • • • • • • • • • • •	,	
•	•		•	•	
• • •	• • •		• • •	• •	
• • • • • • • • • • • • • • • • • • • •	• • • • • • • • • • •	• • • • • • • • • •	• • •	• • • • • • • • • • • •	
• •	•		• •	• •	
•	•		•	•	
• • • • • • • • • • • • • • • • •	•	• • • • • • • • • •	•	• • • • • • • • • • • •	
• • •	• •		• •	• • •	
• • •	• •		• • •	• •	
• • • • • • • • • • • • • • • • • •	• • • • • • • • • • •		• • • • • • • • • • • • • •	• • • • • • • • • • • • •	
• •	• •		• • •	• •	
• •	• •		• •	• •	
•	•		•	•	
lorado w	ildfires	s wors	en 320		
				000	
- goo.gl	/news	/shzz	via	•	
				•	
e.com					
•	•		•		
•	•		•	•	
• • •	• •		• • •	• • •	
son	4 0 N	•••	several do	ays	



12h

24h

36h

•••



12h

24h

36h 48h ... several days



24h

How do we collect social media data during humanitarian crises?

with Carlos Castillo, Fernando Diaz, and Sarah Vieweg





Data Collection: How Is It Done?

Tweets are queried by

Content

#prayforwest
#abflood

Location

longitude: [-97.5, -96.5] & latitude: [31.5, 32]

Data Collection: How Is It Done?

Tweets are queried by

Content

#prayforwest
#abflood

Location

longitude: [-97.5, -96.5] & latitude: [31.5, 32] Maximum 1% of all tweets

Low recall: 33% Not everyone uses the keywords. Maximum 400 terms.

Low precision: 12% Not everyone on the ground talks about the event. Maximum 25 geo-rectangles.

Key Insight: Distill a Crisis Lexicon



••• •••

Gov. McDonnell: Virginia 'Spared' in Hurricane Sandy Damage patch.com/A-zgoL

Deeply touched by the blast at the Boston Marathon. Our thoughts with the affected people and their families. EH

NOTE: Oklahoma University is providing shelter in their dorm facilities for people displaced by the tornado in OK today - @KFOR

If you are able to donate blood- Providence Hospital will have a blood drive in Waco from 11 a.m. to 5 p.m. **#prayforwest**

Best way to help tornado victims is to donate to the Red Cross at redcross.org or text REDCROSS to 90999. #okwx

This flooding is crazy! Hoping my fellow Albertans and Calgarians stay safe! **#abflood #yycflood**



Precision vs. Recall

Precision is straightforward to measure. Recall requires a complete data collection. We use geo data as proxy.

Precision vs. Recall Precision is straightforward to measure. Recall requires a complete data collection. We use geo data as proxy.



Precision vs. Recall Precision is straightforward to measure. Recall requires a complete data collection. We use geo data as proxy.



Precision vs. Recall Precision is straightforward to measure. Recall requires a complete data collection. We use geo data as proxy.



Keyword-based samples appear to overrepresent media reports and underrepresent eyewitness accounts.

Lexicon-based samples better preserve the original distribution of messages types and sources.







Geo-sampling

Keyword-based samples appear to overrepresent media reports and underrepresent eyewitness accounts.

Lexicon-based samples better preserve the original distribution of messages types and sources.



Keyword-sampling







Geo-sampling

Keyword-based samples appear to overrepresent media reports and underrepresent eyewitness accounts.

Lexicon-based samples better preserve the original distribution of messages types and sources.



Keyword-sampling







Geo-sampling

Keyword-based samples appear to overrepresent media reports and underrepresent eyewitness accounts.

Lexicon-based samples better preserve the original distribution of messages types and sources.









KEEP CALM AND QUESTION EVERYTHING

Email: alexandra@aolteanu.com

Web: aolteanu.com

Twitter: @o_saja



KEEP CALM AND QUESTION EVERYTHING