

The Role of Social Media and Artificial Intelligence for Disaster Response

Muhammad Imran

Qatar Computing Research Institute

Hamad Bin Khalifa University

Doha, Qatar

 Follow @mimran15

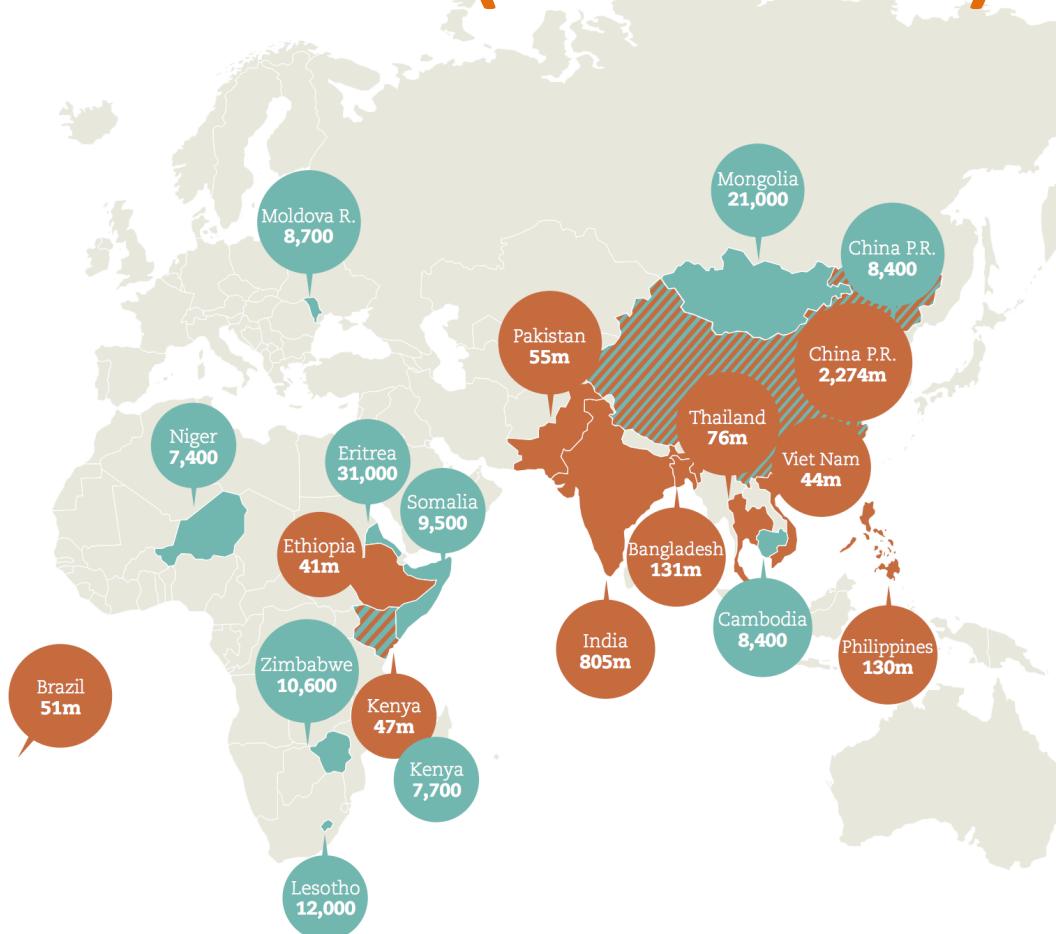
<http://mimran.me/>



This Talk is About...

- **The Role of Information in Time-critical Situations**
 - Natural disasters and their destructions
 - Man-made disasters and mass convergence events
- **The Role of Social Media for Disaster Response**
 - Particular focus on micro-blogging platforms
 - Availability of various types of information and opportunities
- **The Role of Artificial Intelligence for Disaster Response**
 - How AI is useful in disaster response
 - Various AI techniques, approaches, and tools
 - Work of crisis computing group at QCRI
 - Ongoing research
 - Future directions

Most Affected Countries by Natural Disasters (1995-2015)



- Top 10 countries with highest proportion of affected people over the total population (per 100,000 inhabitants)
- Top 10 countries with the highest absolute number of affected people (in million)



Source: UNISDR

Humans Suffering and Economic Damage by Disasters



Humans suffering from the impacts of disasters, crises, and armed conflict

Millions of people affected each year by disasters;

At an annual cost to the global economy that exceeds \$300 billion

Plan and Prepare



**Disasters are unavoidable
but planning can lessen
their effects**



Humans suffering from the impacts of disasters, crises, and armed conflict

Millions of people affected each year by disasters;
At an annual cost to the global economy that exceeds \$300 billion

Plan and Prepare



Provide helping hand...

Humans suffering from the impacts of disasters, crises, and armed conflict

Millions of people affected each year by disasters;

At an annual cost to the global economy that exceeds \$300 billion

The Urgency to Act and Plan



THE PRESIDENT
OF THE
GENERAL ASSEMBLY

30 March, 2016

Excellency,

The world is witnessing levels of human suffering unseen in generations. More than 120 million women, men and children worldwide are in need of humanitarian assistance in 2016.

Human suffering from the impacts of armed conflicts and natural disasters has reached staggering levels. Nearly 60 million people, half of them children, have been forced from their homes due to conflict and violence. Between 2008 and 2014, a total of 184 million people were displaced by natural disasters, an average of 26.4 million each year.

Millions of refugees in Africa are in urgent need of humanitarian assistance and regrettably they have not received the required attention and support from the international community. The severe effect of El Nino has resulted in drought conditions across eastern, southern and the Horn of Africa and putting further pressure on refugee hosting countries.

I will organize an informal meeting of the General Assembly titled "*Humanitarian Response in Africa: The Urgency to Act*" to be held in the Trusteeship Council on 8 April 2016 between 3-6 pm. The meeting will provide an opportunity to timely engage on and effectively assist with addressing the humanitarian needs in Africa, particularly the needs of refugees.

The informal meeting will consist of interventions from a number of panelists, followed by an interactive discussion between Member States, the UN system and other stakeholders. The concept note and draft programme of the meeting are attached.

Please note that there will be no list of speakers for the informal meeting and Member States are encouraged to engage in an interactive discussion with the panelists.

Please accept, Excellency, the assurances of my highest consideration.



Mogens Lykketoft

President of the
United Nations General Assembly



Emblem of the United Nations



Incumbent
 Mogens Lykketoft

To All Permanent Representatives and
Permanent Observers to the United Nations
New York

Information: A Lifeline During Disasters

The **opaqueness** induced by disasters is **overwhelming**

People need information as much as water, food, medicine or shelter

Lack of information can make people **victims of disaster** and **targets of aid**



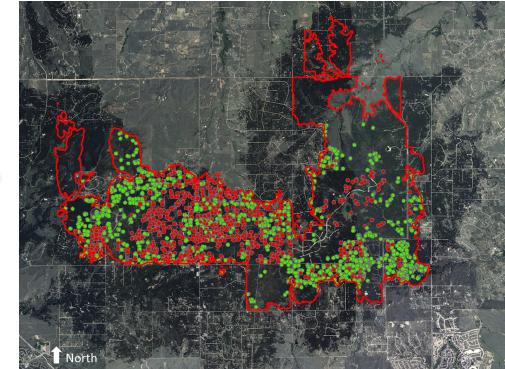
Data or Dialogue? The Role of Information in Disasters

Iain Logan, former head of disaster operations, International Federation:

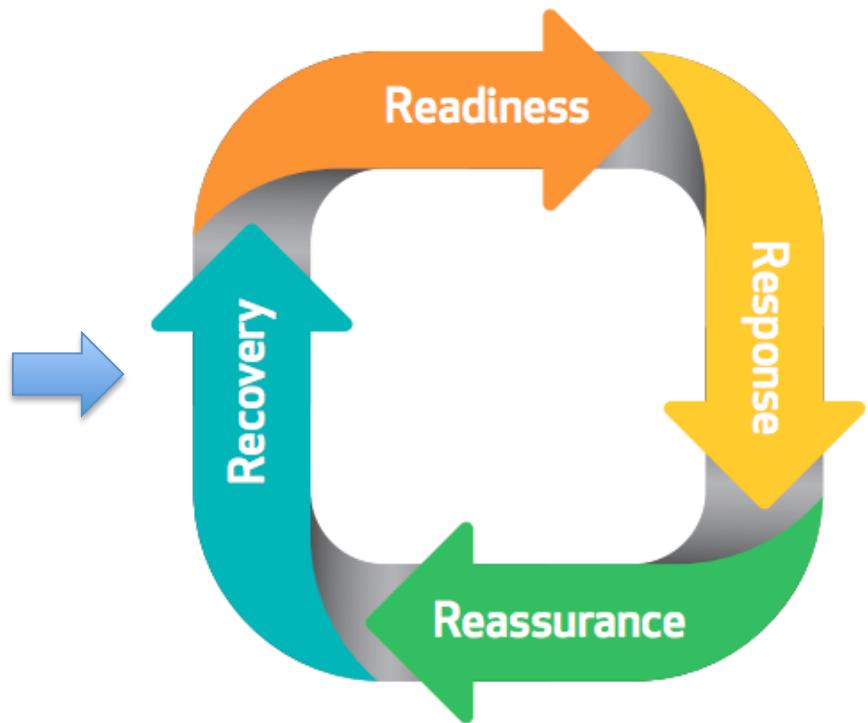
"The very first thing you need to do is climb into a helicopter. You can't get to see how

- Information Bestows Power -

Salvador, 2001), the logistics involved, the population centers – also where not to send people. Then you must get on the ground to get the quality."



The Role of Social Media





PINTEREST

**SOCIAL SITE
THAT IS ALL ABOUT
DISCOVERY**

**LARGEST
OPPORTUNITIES**



USERS ARE:

♂ 20% MALE

♀ 80% FEMALE

70 MILLION
USERS



TWITTER

**MICRO BLOGGING
SOCIAL SITE
THAT LIMITS EACH
POST TO 140
CHARACTERS**

**LARGEST
PENETRATION**



BUT SPREADING
SLOWLY AND STEADILY

9,100 TWEETS
HAPPEN
EVERY
SECOND



289 MILLION
ACTIVE USERS



FACEBOOK

**MOBILE IS
FACEBOOK'S
CASH COW**



MOBILE AD REVENUE
MAKES UP 76%
OF ALL AD REVENUE
(\$2.9 BILLION IN Q2 OF 2015)

USERS
SHARE
1 MILLION LINKS
EVERY 20 MINUTES

f 1.5
BILLION
MONTHLY
ACTIVE USERS



INSTAGRAM

**SOCIAL SHARING
SITE ALL AROUND
PICTURES
AND NOW 15 SECOND
VIDEOS**

MANY BRANDS
ARE PARTICIPATING
THROUGH THE USE OF

HASHTAGS

AND POSTING
PICTURES
CONSUMERS
CAN RELATE TO

MOST FOLLOWED
BRAND IS
NIKE

300 MILLION
ACTIVE USERS



GOOGLE+

**SOCIAL NETWORK
BUILT BY GOOGLE
THAT ALLOWS FOR
BRANDS
AND USERS
TO BUILD CIRCLES**

NOT AS MANY
BRANDS ACTIVE,
BUT THE ONES THAT ARE
TEND TO BE A
GOOD FIT WITH A
GREAT FOLLOWING

HANGOUTS
AND PHOTOS
HAVE BEEN SEPARATED
FROM GOOGLE+
BUT POSTS WILL REMAIN
AS "STREAMS"

g+ 300
MILLION
ACTIVE USERS



LINKEDIN

**BUSINESS
ORIENTED
SOCIAL NETWORKING SITE**

BRANDS THAT ARE
PARTICIPATING
ARE CORPORATE
BRANDS
GIVING POTENTIAL AND
CURRENT ASSOCIATES
A PLACE TO NETWORK
& CONNECT



79% OF USERS
ARE 35
OR OLDER

in 380
MILLION
USERS WORLDWIDE

Social Media Use During Christchurch Earthquake



A self-organized workforce of **10,000 volunteers** gathered on Facebook

UC Student Volunteer Army

Muhammad Home 20+

Serve for New Zealand Anzac Day 25 April 2016

STUDENT VOLUNTEER ARMY

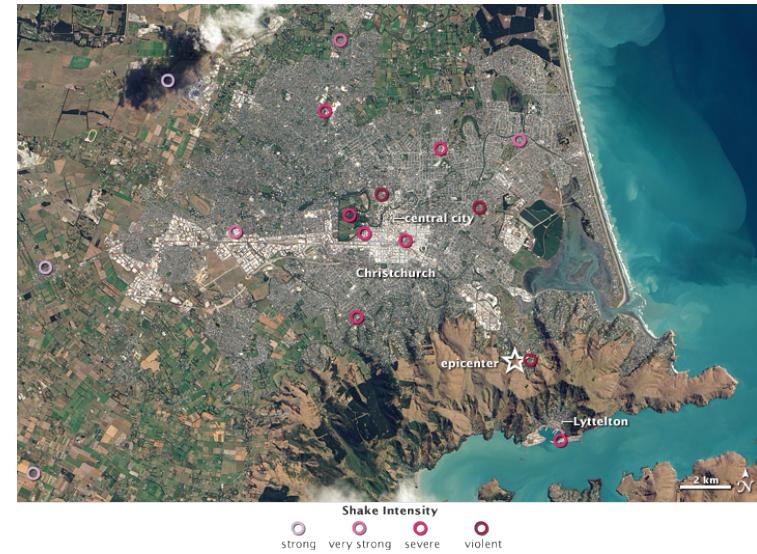
UNIVERSITY OF CANTERBURY

100 YEARS GALLIPOLI

UC Student Volunteer Army Disaster Relief · Youth Organization · Community Organization

Sign Up Like Message

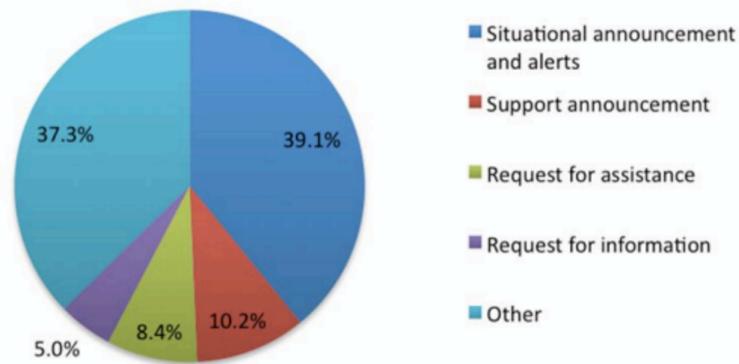
Timeline About Events Donate More



The Role of Twitter During Thailand Floods



Thailand Twitter Message Year 2011



[Alisa Kongthon et al. 2011]

Twitter Breaks Events Faster

Hudson Plane Crash



First report



Janis Krums
@jkrums



 Follow

<http://twitpic.com/135xa> - There's a plane in the Hudson. I'm on the ferry going to pick up the people. Crazy.

RETWEETS
244

LIKES
878



9:36 PM - 15 Jan 2009



Westgate Mall Attack

Breaks the story 33 minutes before local TV



I come from Somalia.
@geekinthejungle



 Follow

gun shots in westlands? wtf??

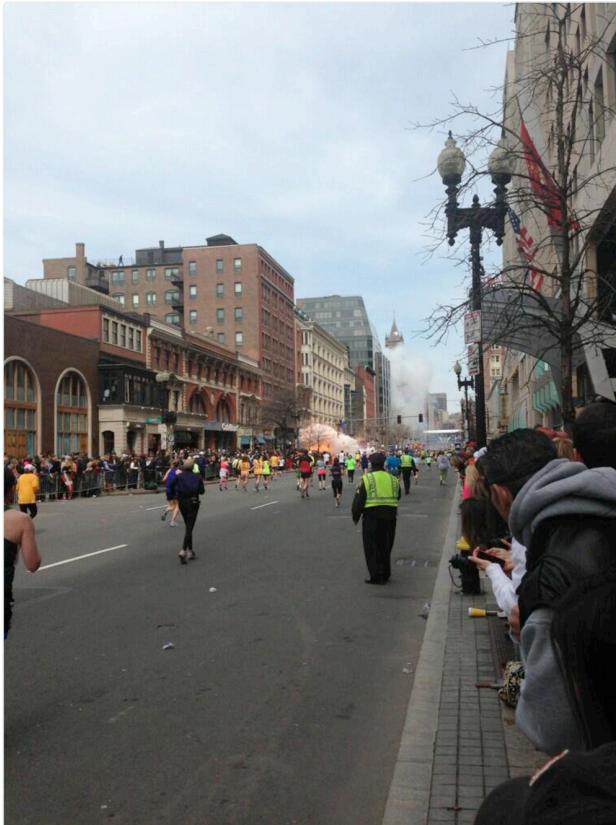
 Reply  Retweet  Favorite  Buffer  Storify  More 

12:38 PM - 21 Sep 13

Twitter Breaks Events Faster

Boston Bombing

First report on Twitter



Dan Lampariello
@WBBJ7Dan

Explosion at copy

9:50 PM - 15 Apr 2013

↳ 1,943 ⚡ 187

[Follow](#)

After 1 minute

Pure Applesauce
@stackii

[Follow](#)

An explosion just went off in downtown Boston. Spectators fleeing the #bostonmarathon course.

9:51 PM - 15 Apr 2013

↳ 79 ⚡ 11

Pure Applesauce
@stackii

[Follow](#)

I think a bomb just went off in Boston. Can't tell. Can smell smoke. Emergency vehicles everywhere.

instagram.com/p/Yltk1gTIUJ/

9:53 PM - 15 Apr 2013

↳ 308 ⚡ 21

After 2 minutes

Fox Sports 1380 KRKO
@KRKO1380

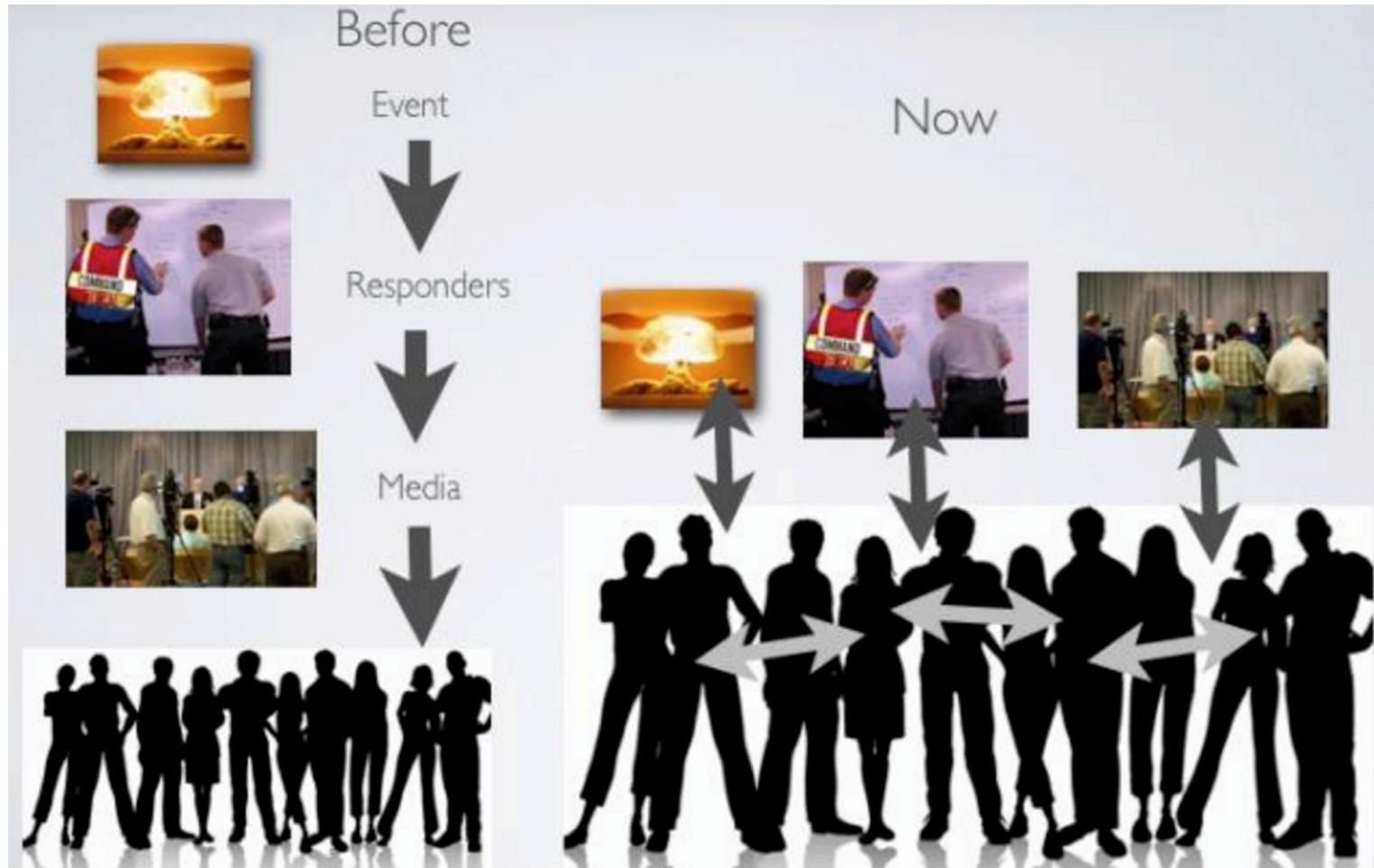
[Follow](#)

BREAKING: Per our man on the ground at the Boston Marathon, @toobladogs, there was an explosion. More to follow.

9:52 PM - 15 Apr 2013

↳ 153 ⚡ 3

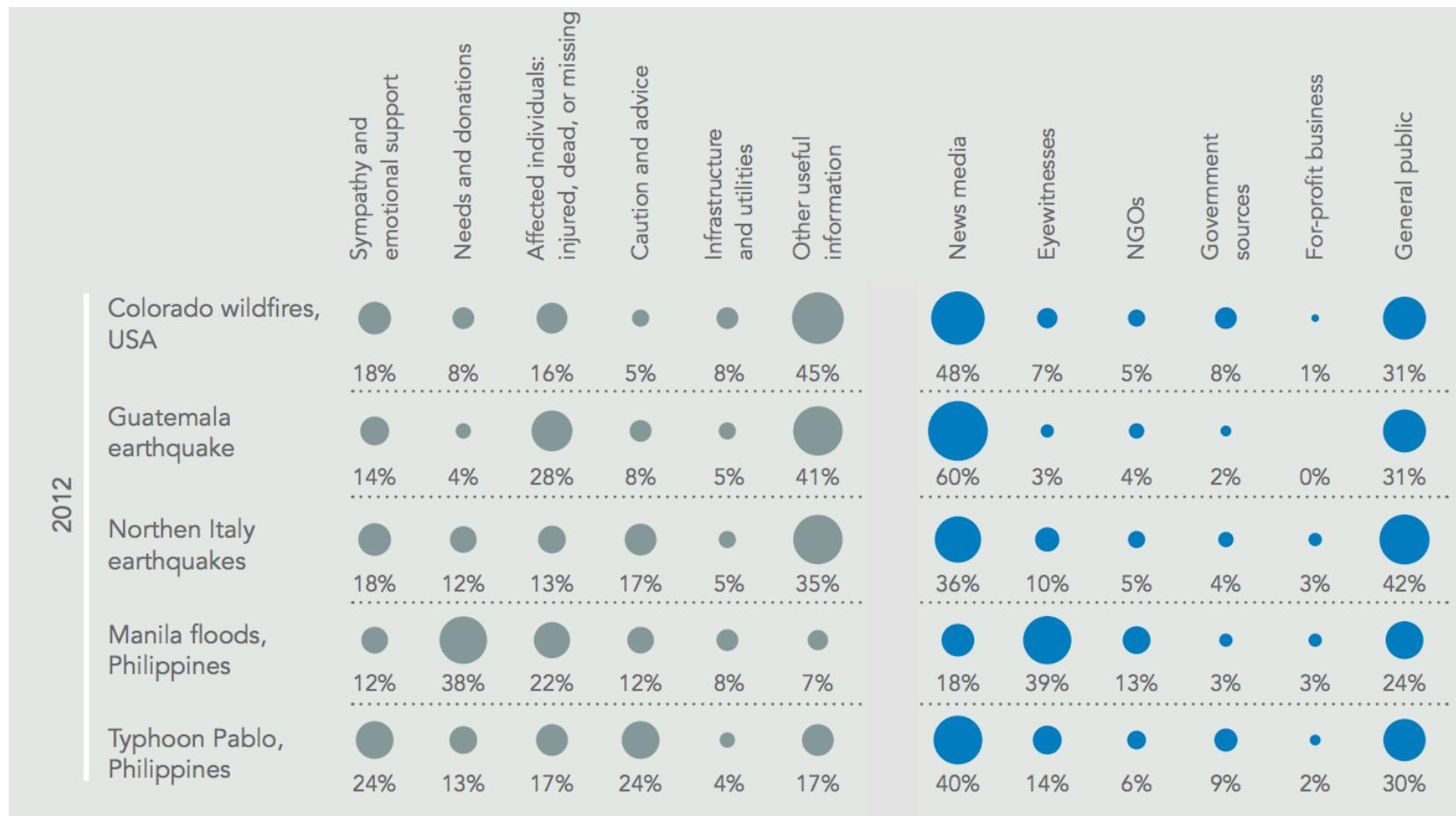
Crisis Communications Before and Now



Analysis of Twitter Crisis-Related Data

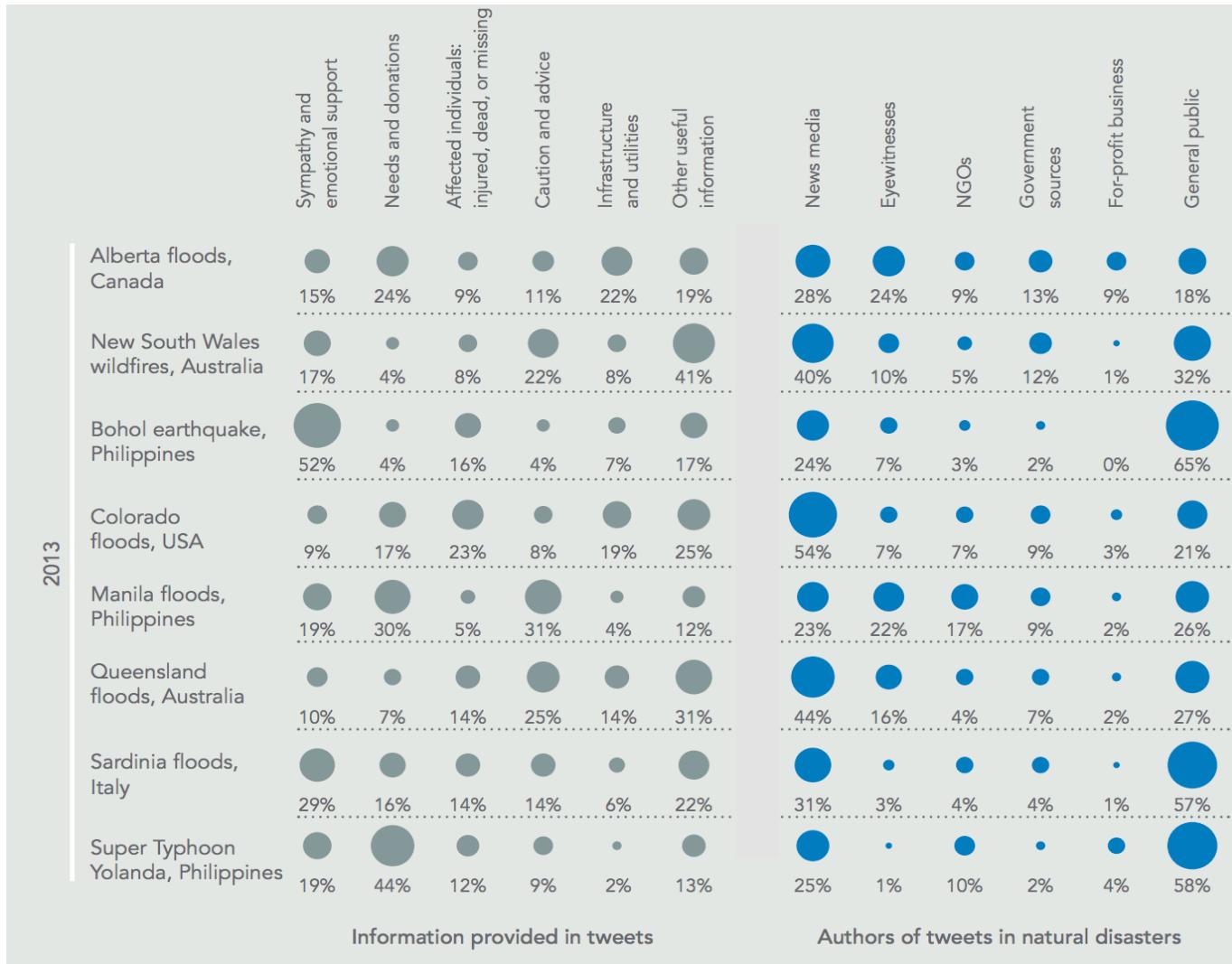
An In-depth Study

Twitter Crisis-Related Data (2012)



Twitter data from **13 crises**; Analyzed over **100,000 tweets**; Information **types and sources**

Twitter Crisis-Related Data (2013)

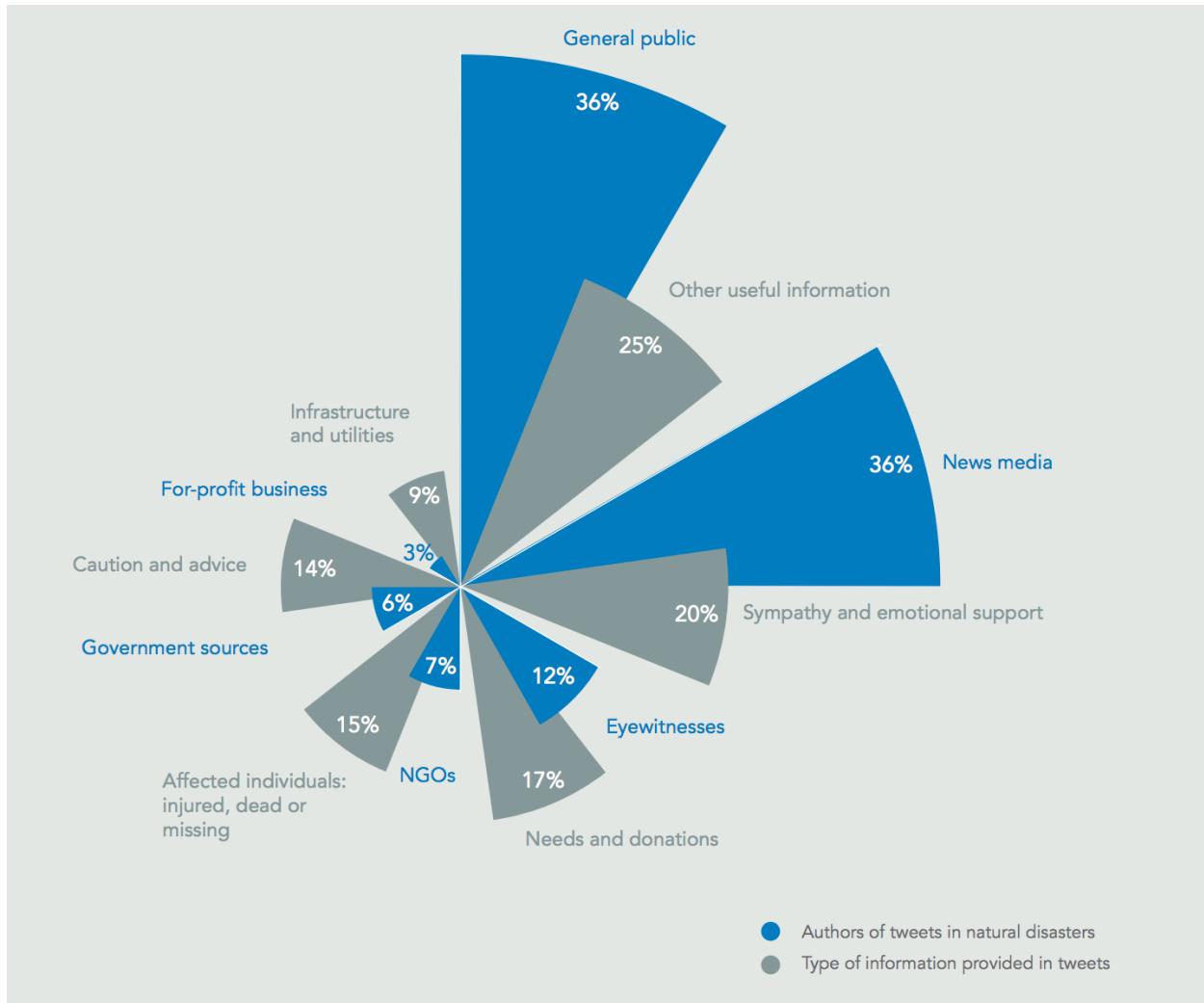


Twitter data from **13 crises**; Analyzed over **100,000 tweets**; Information **types and sources**

Source: Qatar Computing Research Institute - Published in World Humanitarian Data and Trends 2014 (UN OCHA)

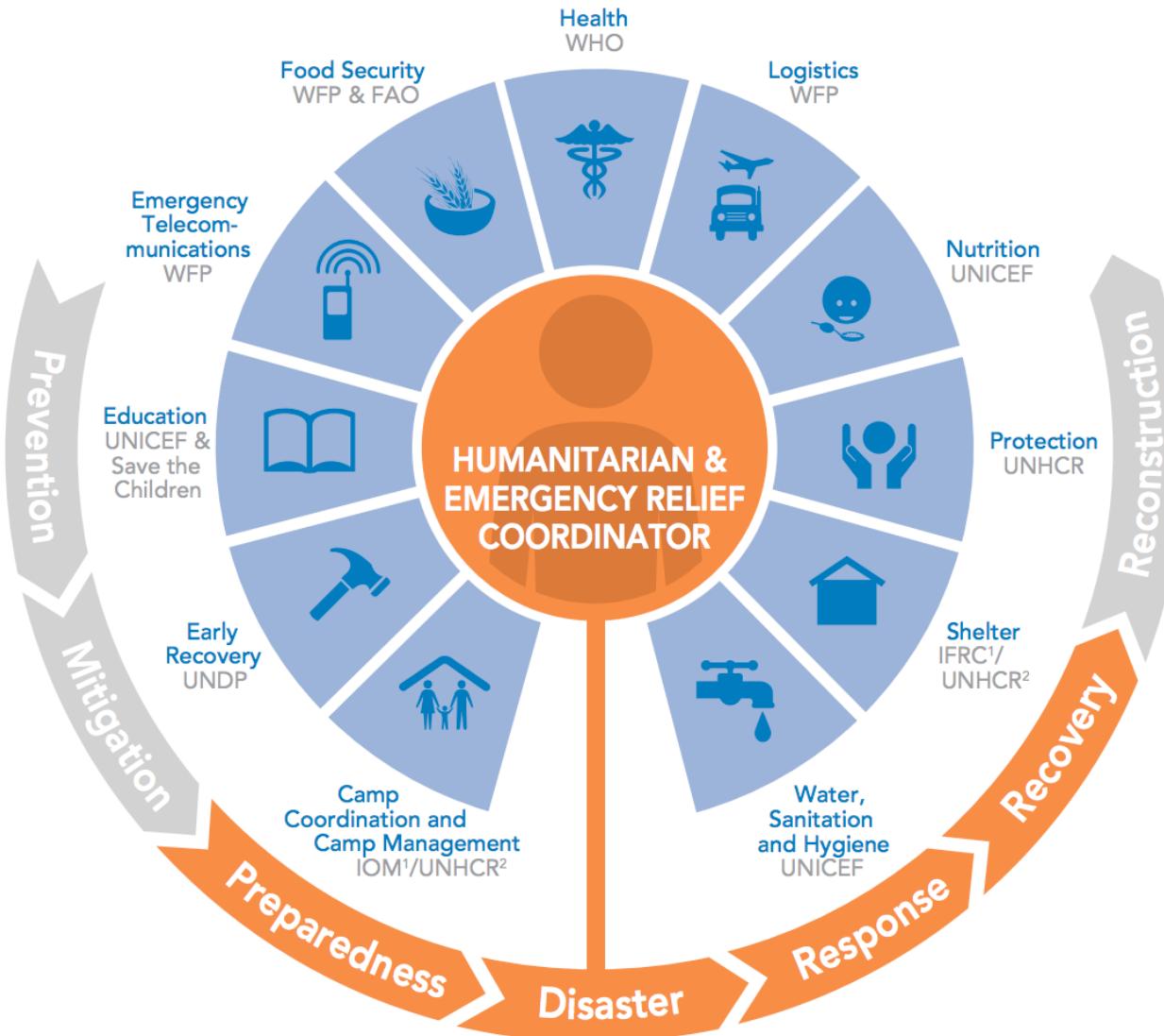
Twitter Crisis-Related Data (All)

- Twitter data from **13 recent crises**
- Over **100,000 tweets**
- **Information types**
- **Types of sources**



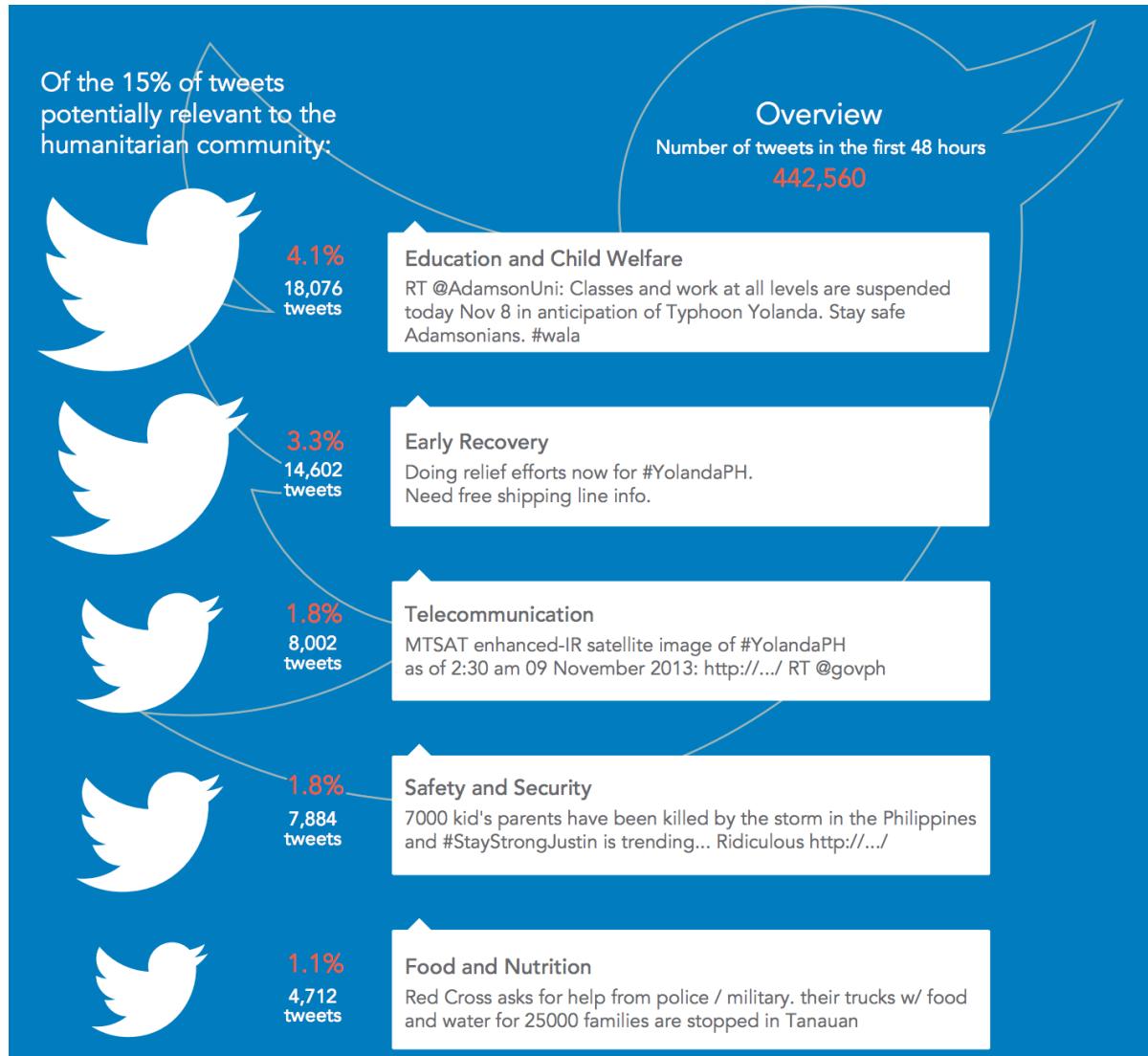
Source: Qatar Computing Research Institute - Published in World Humanitarian Data and Trends 2014 (UN OCHA)

UN OCHA Humanitarian Clusters- Related Information on Twitter



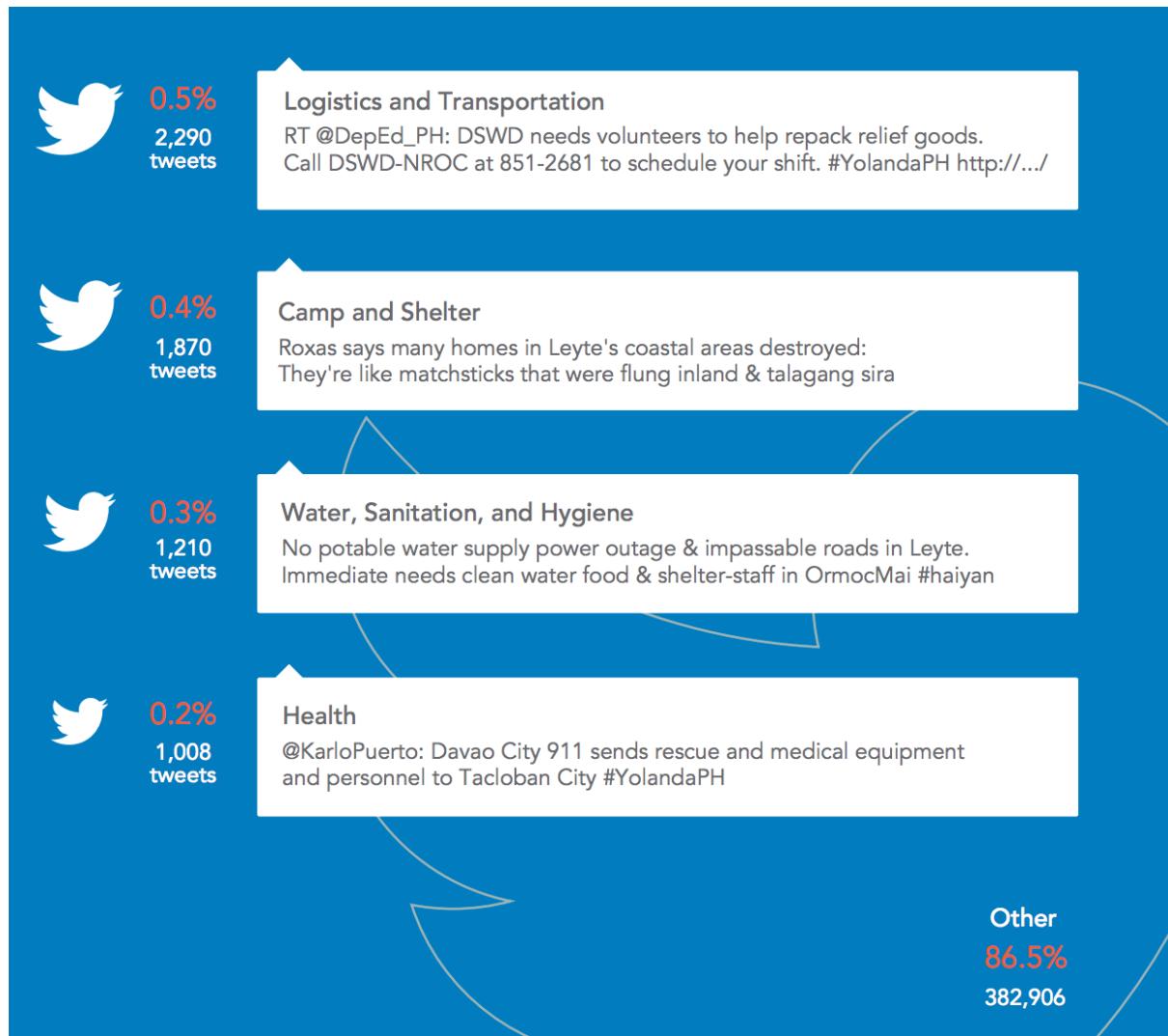
Typhoon Yolanda – OCHA Clusters

- Performed analysis of more than **440,000** tweets during the **first 48 hours**
- **15%** of the tweets found **potentially relevant**



Typhoon Yolanda – OCHA Clusters

- Performed analysis of more than **440,000** tweets during the **first 48 hours**
- **15%** of the tweets found **potentially relevant**



Sandy Hurricane Twitter Data Analysis



freaking out. home alone. will just watch tv #Sandy #NYC.



@NYGovCuomo orders closing of NYC bridges. Only Staten Island bridges unaffected at this time. Bridges must close by 7pm. #Sandy #NYC.



rt @911buff: public help needed: 2 boys 2 & 4 missing nearly 24 hours after they got separated from their mom when car submerged in si. #sandy #911buff



400 Volunteers are needed for areas that #Sandy destroyed.

Sandy Hurricane Twitter Data Analysis



freaking out. home alone. will just watch tv #Sandy #NYC.

Personal



@NYGovCuomo orders closing of NYC bridges. Only Staten Island bridges unaffected at this time. Bridges must close by 7pm. #Sandy #NYC.



rt @911buff: public help needed: 2 boys 2 & 4 missing nearly 24 hours after they got separated from their mom when car submerged in si. #sandy #911buff



400 Volunteers are needed for areas that #Sandy destroyed.

Informative

Sandy Hurricane Twitter Data Analysis



freaking out. home alone. will just watch tv #Sandy #NYC.

Personal



@NYGovCuomo orders closing of NYC bridges. Only Staten Island bridges unaffected at this time. Bridges must close by 7pm. #Sandy #NYC.

Caution and Advice



rt @911buff: public help needed: 2 boys 2 & 4 missing nearly 24 hours after they got separated from their mom when car submerged in si. #sandy #911buff

Casualties and Damage



400 Volunteers are needed for areas that #Sandy destroyed.

Donations

Informative

Sandy Hurricane Twitter Data Analysis



freaking out. home alone. will just watch tv #Sandy #NYC.

Personal



@NYGovCuomo orders closing of NYC bridges. Only Staten Island bridges unaffected at this time. Bridges must close by 7pm. #Sandy #NYC.

Caution and Advice



rt @911buff: public help needed: 2 boys 2 & 4 missing nearly 24 hours after they got separated from their mom when car submerged in si. #sandy #911buff

Casualties and Damage



400 Volunteers are needed for areas that #Sandy destroyed.

Donations

Informative

MERS Outbreak Twitter Data Analysis

Middle East Respiratory Syndrome (MERS)

Twitter data collection from: **2014-04-27** to **2014-07-14** using hashtag **#MERS**
(Total = 215,370)

Data analysis:

Reports of symptoms

Reports of signs or symptoms such as fever, cough or questions

Affected people reports

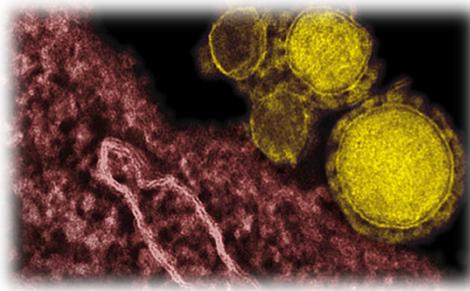
Reports of affected people due to the MERS disease

Death reports

Reports of deaths due to the MERS disease

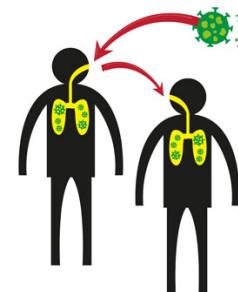
Prevention questions

Questions or suggestions related to the prevention of disease



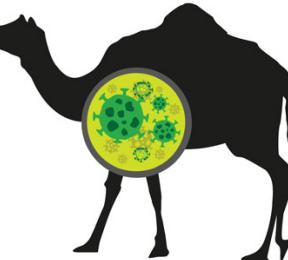
Disease transmission reports

Reports or questions related to the transmission of the disease



Treatment questions

Questions or suggestions regarding the treatment of the disease



Social Media During MERS Outbreak



RT @abettel: Two workers at **FL** hospital exposed to a patient with Middle East Respiratory Syndrome are **showing flu-like symptoms**

Signs and symptoms



Coronavirus symptoms include: **fever, coughing, shortness of breath, congestion in the nose and throat, and in some cases diarrhea.** MERS

Signs and symptoms



#MERS is a relatively new respiratory illness, spread b/w people in close contact.
Symptoms are fever, cough, & shortness of breath.

Signs and symptoms



Saudi Arabia finds another **32 MERS cases** as disease spreads: RIYADH (Reuters) - Saudi Arabia said on Thursday ... <http://t.co/cPhm0uTRCo>

Affected individuals

Social Media During MERS Outbreak



First Case of Deadly Middle Eastern Virus Found in U.S.: The Centers for Disease Control has confirmed that a case of the deadly Midd...

Affected individuals



Third Case of MERS Confirmed in the U.S.: The U.S. Centers for Disease Control and Prevention confirmed on Sat... <http://t.co/Sb8PMyxVUn>

Affected individuals



No clear transmission link btwn camels and humans for MERS. 94% Egyptian camels seropositive but no human cases yet. Hmm #asm2014

Transmission



Saudi health authorities announced on Monday that the **death toll from the MERS coronavirus has reached 115** since the respiratory disease ...

Death reports

ISCRAM Call for Papers

1. Network Theory		8. Social Media Studies	
2. Analytical Modelling and Simulation		9. Community Engagement	
3. Planning, Foresight and Risk Analysis		10. Decision Support System	
4. Ethical, Legal and Social Issues		11. Practitioner Cases And Practitioner-Centered Research	
5. Geospatial Data and Geographical Information Science		12. Researching Crisis: Methodologies	
6. Command & Control Studies		13. Serious Gaming	
7. Human Centred Design and Evaluation		14. Understanding Collaborative Work Practices	

8. Social Media Studies



Aid is Out There!



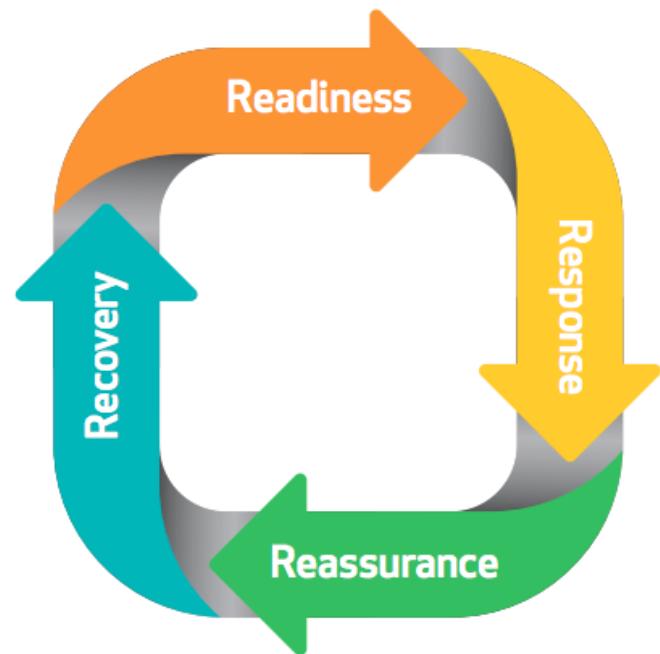
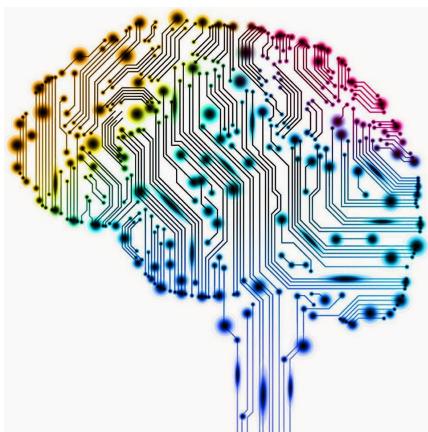
Aider is Out There!



AIDR is Also Out There!



The Role of Artificial Intelligence



Data and Opportunities



2013 Pakistan Earthquake
September 28 at 07:34 UTC



Disease outbreaks



2010 Haiti Earthquake
January 12 at 21:53 UTC

Availability of Immense Data:

 Twitter  @twitter

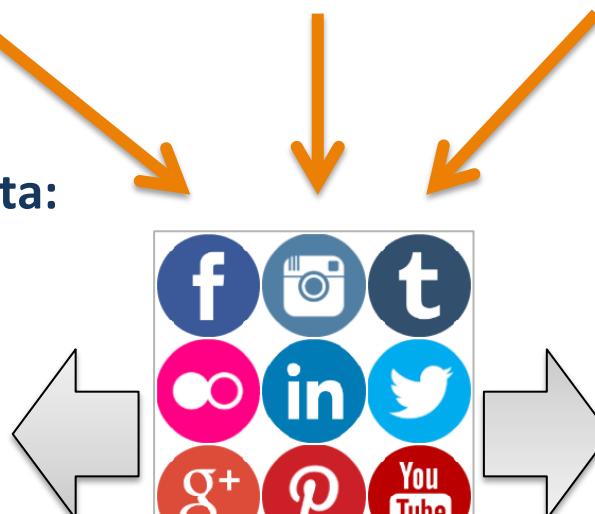
Follow

People sent more than 20 million Tweets about the storm between Oct 27 & Nov 1. Terms tracked: "sandy", "hurricane", #sandy, #hurricane.

6:46 PM - 2 Nov 2012

619 RETWEETS 199 FAVORITES

Around 16 thousands tweets per minute were posted during the hurricane Sandy in the US.



Opportunities:

- Early warning and event detection
- Situational awareness
- Actionable information
- Rapid crisis response
- Post-disaster analysis

Processing Social Media Data

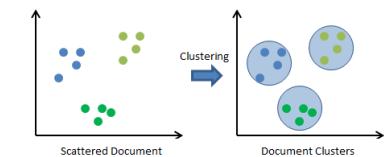
Filter: removing, duplicates, spam and messages from bots



Classify: categorization of items into information types



Cluster: identify trending and emerging topic



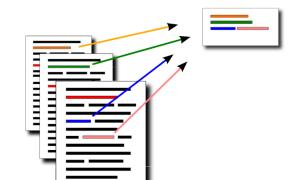
Aggregate: making sense by connecting different pieces



Extract: short snippets of focused information



Summarize: learning a bigger picture of an event



WAIT!

Before applying any technique?

Please!

Have a look at your data first

Data Characteristics and Preparation

- Single-word slangs: pls (please), srsly (seriously)
- Multi-word slangs: imo (in my opinion)
- Misspellings: missin (missing), ovrcme (overcome)
- Phonetic substitution: 2morrow (tomorrow)
- Word without spaces: prayfornepal (pray for nepal)

Can you guess?

“r u ok m8” ?? >> “Are you OK, mate?”



Tools to Process Social Media Data

Systems for Crisis-Relevant Data Processing

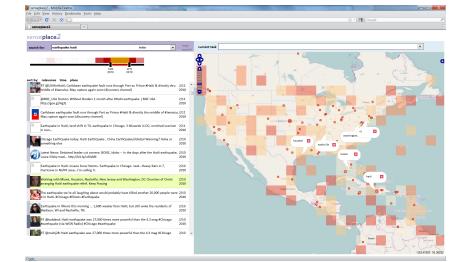
Twitris [Purohit and Sheth 2013]

Twitter; semantic enrichment, classify automatically, geotag



SensePlace2 [MacEachren et al. 2011]

Twitter; geotag, visualize heat-maps based on geotags



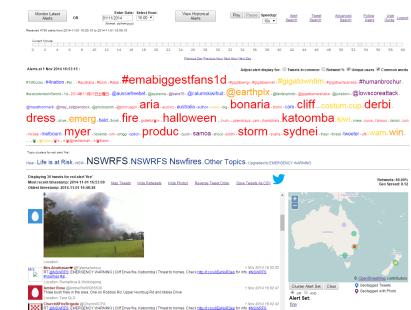
EMERSE Enhanced Messaging for the Emergency Response Sector [Caragea et al. 2011]

Twitter and SMS; machine-translate, classify automatically, alerts



ESA Emergency Situation Awareness [Yin et al. 2012; Power et al. 2014]

Twitter; detect bursts, classify, cluster, geotag



Systems for Crisis-Relevant Data Processing

Twitcident [Abel et al. 2012]

Twitter and TwitPic; semantic enrichment, classify

The screenshot shows a search results page for "Texas Fire". At the top, it says "Started on Sep. 1, 2011 12:00:00 AM" and "Current selection: 22,968 tweets users". Below this is a list of tweets from various users like "objective21", "TellestLambLusk", "Deposito", and "traveler7200". On the right side, there are filters for "Search in Texas Fire tweets", "Filter tweets", and "Remove filters".

CrisisTracker [Rogstadius et al. 2013]

Twitter; cluster, annotate manually

The screenshot shows the Crisis Tracker interface with a map of Syria and Lebanon. The map has several red dots indicating event locations. On the left, there are filters for "WHAT" (Civilian involvement, Military involvement, Crime, Violence, Missing people, Damaged infrastructure), "WHERE" (Natural hazard, Political event, Summary report, Available resource, Request/Need, Warning/Risk/Danger, High impact event), "WHO" (Enter name), and "WHEN" (From, To). Below the map, there is a list of events with details like date, time, title, and tags.

Tweedr [Ashktorab et al. 2014]

Twitter; classify automatically, extract information, geotag

AIDR: Artificial Intelligence for Disaster Response [Imran et al. 2014a]

Twitter; annotate manually, classify automatically

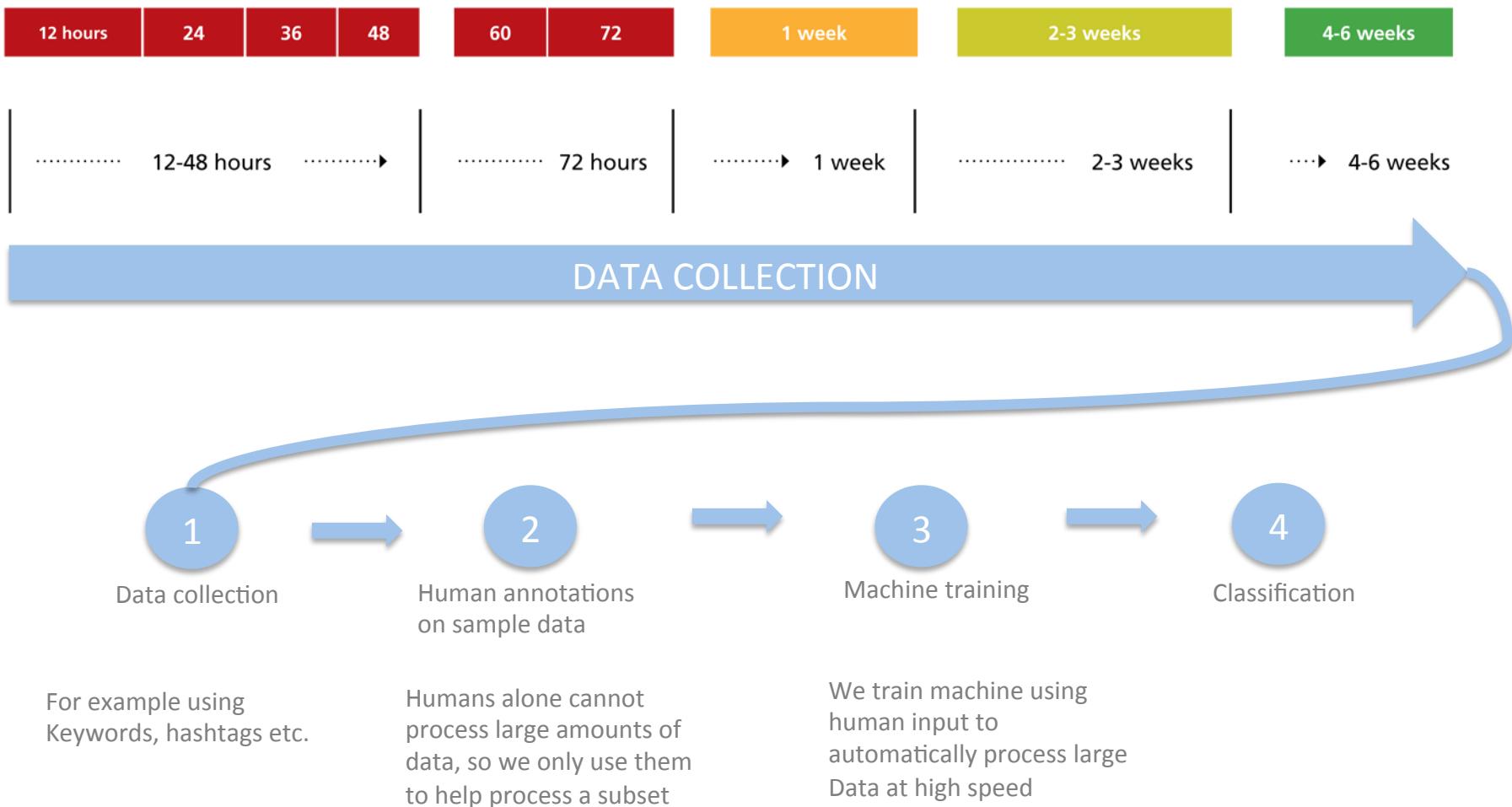
The screenshot shows the AIDR interface with a collection titled "Global HIV-Related Research". It lists tweets from users like "Global HIV-Related Research" and "Zika Virus in Brazil". Each tweet has a "Collect" and "Classify" button next to it. The interface also includes a sidebar for "My Collections" and a "Create Collection" button.

Artificial Intelligence for Disaster Response



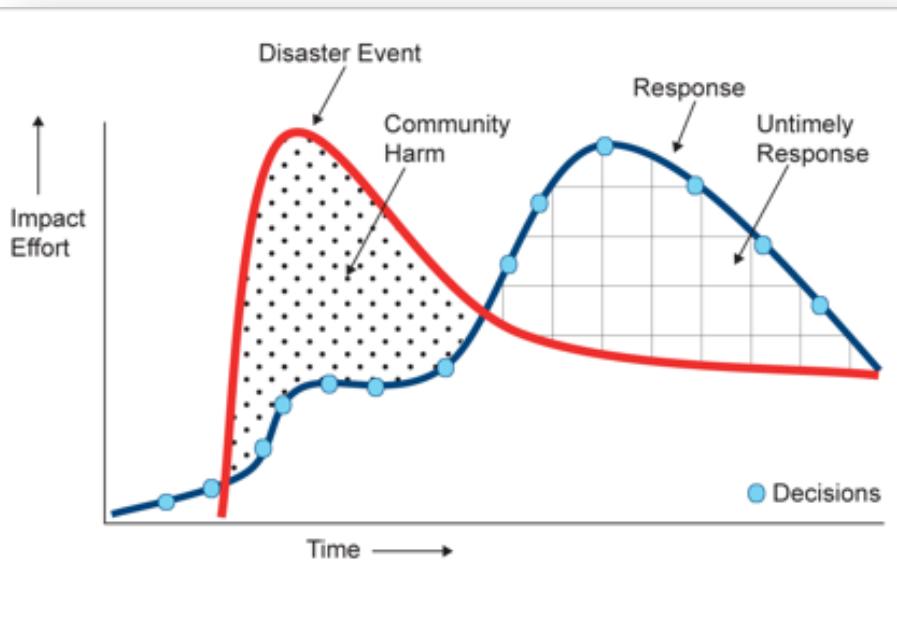
Information Processing

Disaster Timeline:

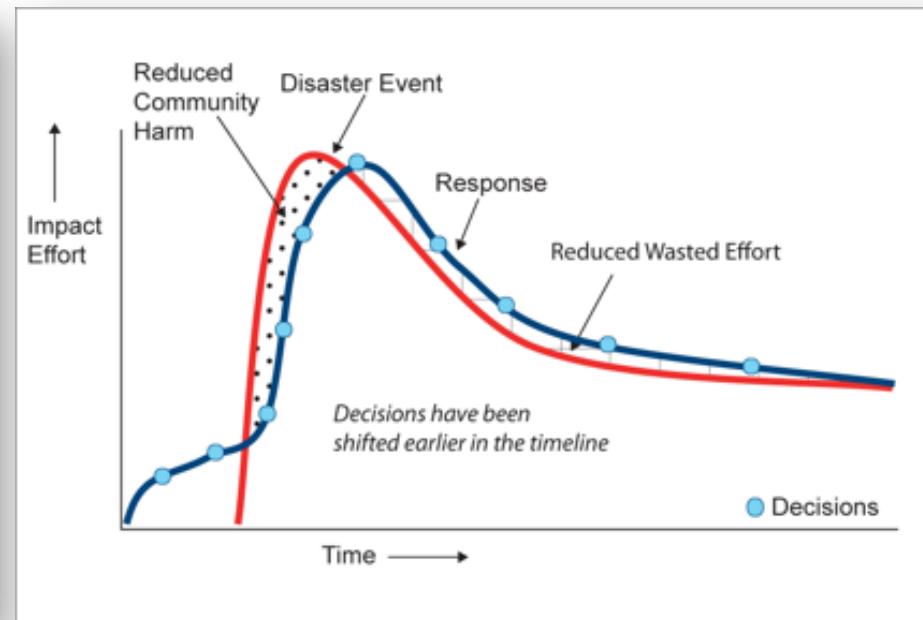


Impact and Response Timeline

Disaster response (today)



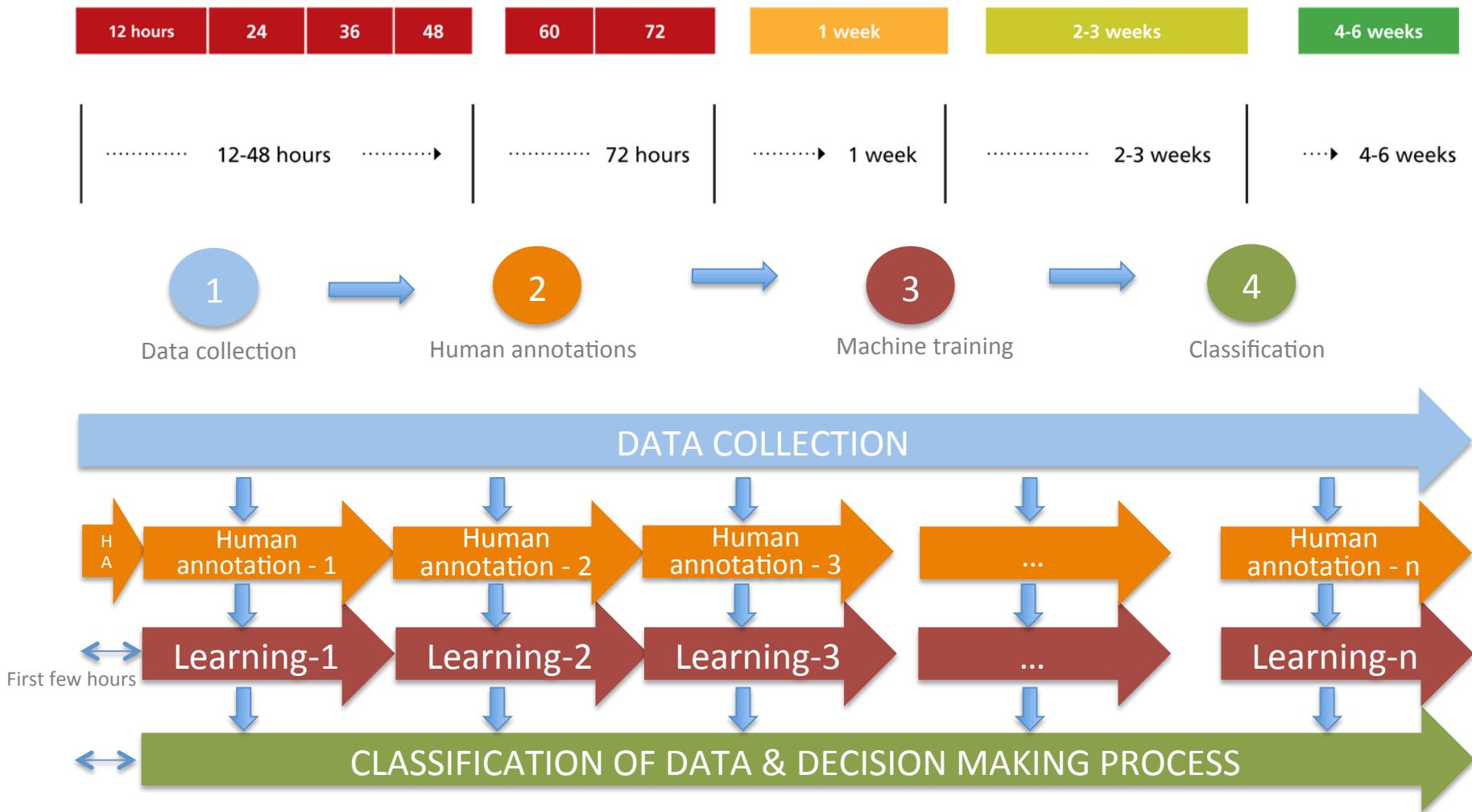
Disaster response (our target)



Requires real-time processing of data

Information Processing (Real-time)

ONLINE APPROACH



Big Challenges – 4Vs

- **Volume**
Scale of data (20m tweets in 5 days Typhoon Oklahoma)
- **Velocity**
Analysis of streaming data (16k/min during Sandy)
- **Variety**
Different forms/types of data (information types)
- **Veracity**
Uncertainty of data

Machine Learning + Crowdsourcing



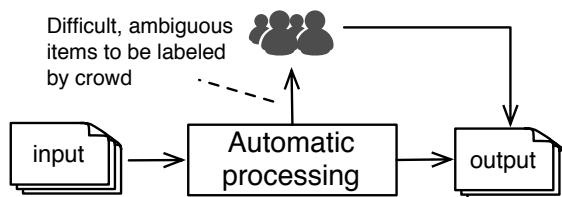
DATA

AIDR = Machine learning + Crowdsourcing

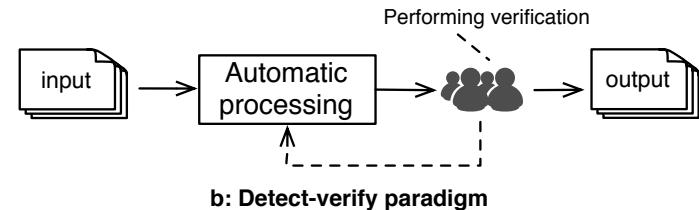


Crowdsourced Stream Processing

Combining human and machine computation



a: Split automatic/manual processing

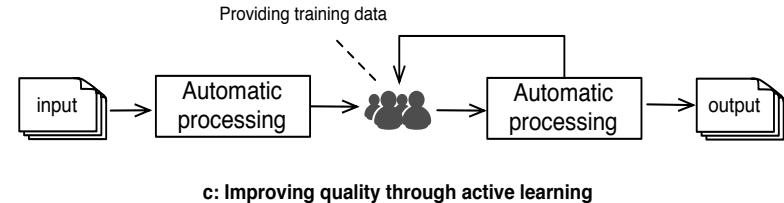


b: Detect-verify paradigm

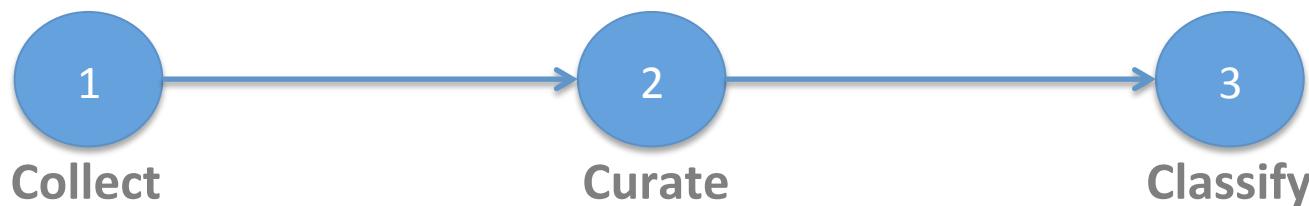
Quality assurance loops: human processing elements do the work, automatic processing elements check for consistency

Process-verify: work is done automatically, humans check low-confidence or borderline cases

Online supervised learning: humans train machines to perform work automatically



c: Improving quality through active learning



AIDR —Artificial Intelligence for Disaster Response— is a free, open-source, and easy-to-use platform to automatically filter and classify relevant tweets posted during humanitarian crises.

<http://aidr.qcri.org/>

Awarded the Grand Prize in the Open Source Software World Challenge 2015

AIDR: From End-users Perspective

2 steps approach

1

Collection

2

Classifier(s)

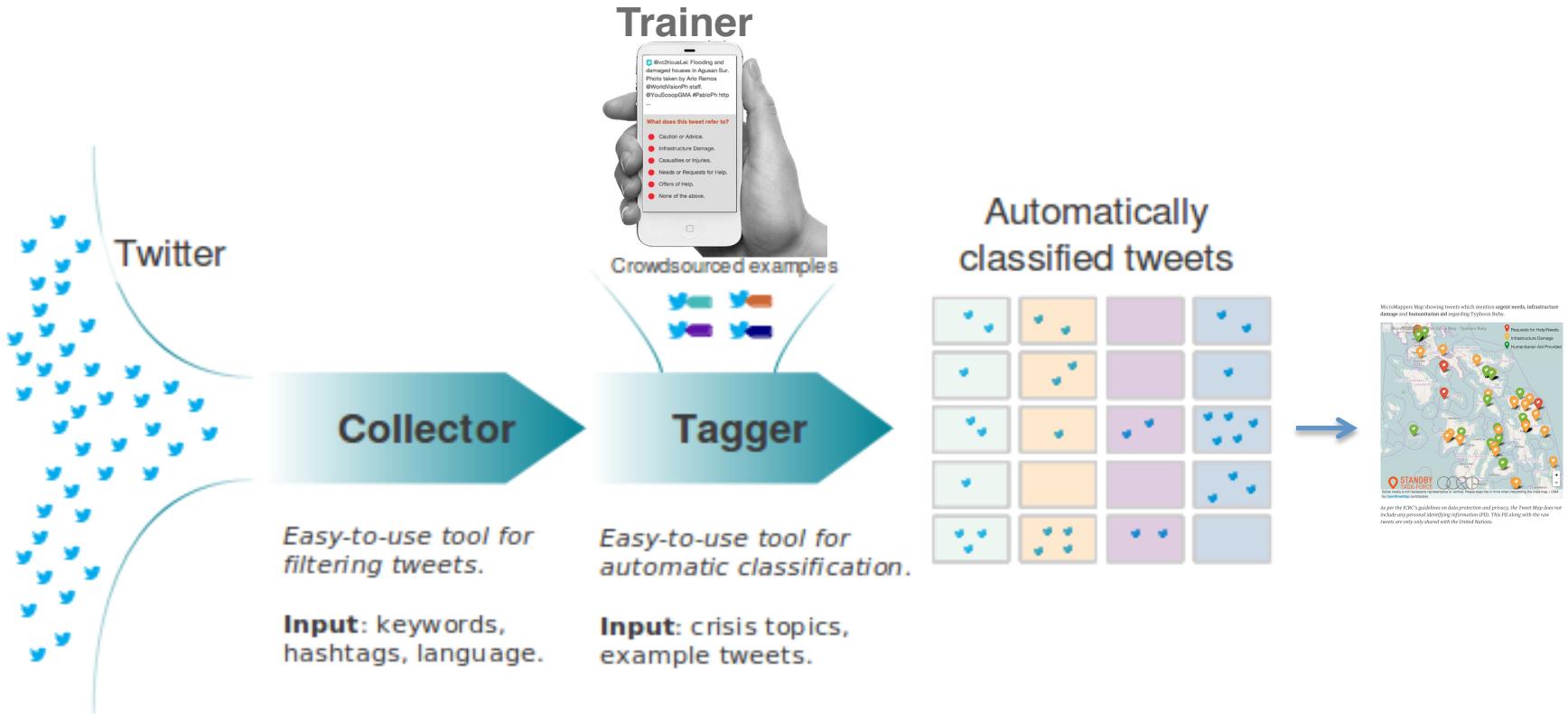
A collection is a set of filters

- Keywords, hashtags
- Geographical bounding box
- Languages
- Follow specific set of users

A classifier is a set of tags

- Donations requests & offers
- Damage & causalities
- Eyewitness accounts
- ...

Real-time Classification in AIDR



AIDR – Collection Setting

Collection definition

My Collections

Flu season

Details History Edit Permissions Download/Export

Name: Flu season
Short name: 2014-01-flu
Keywords: #fluseason, flu
Language(s): English (en)
Automatically stop after: 2 days

* If you need to run your collection for more than 7 days, please contact the AIDR team.

Optional settings

Geographical regions: e.g., 43.43, 22.44, 89.32, 56.43 (max 25)
* boundingbox.klokantech.com ('Copy/paste CSV format of a boundingbox')
Follow specific users: e.g., 47423744, 53324456 (max 5000)

Cancel Save



Language filter



Geographical region filter

Collection detail dashboard

My Collections

Flu season

Details History Edit Permissions Download/Export Start

Short name: 2014-01-flu
STOPPED

Downloaded tweets 15
(since last re-start):

Total downloaded tweets: 99,545
Last downloaded tweets: "@Ashton5SOS: I have the ultimate man flu" MY BABYYYYY

Time period

Created: January 30 2014
Last started: March 2 2014
Last stopped: March 2 2014

Configuration

Keywords: #fluseason, flu
Language(s): English (en)

Collaborators

mimran15, QCRI_SC

<http://aidr.qcri.org/>

AIDR – Classifiers Setting

My Classifiers > Details

Classifier for "Flu season"

Details Edit View/Download

Public link for volunteers (in MicroMappers): <http://clickers.micromappers.org/aidr/2014-01-flu>

Go To Collection

Add a new classifier

Classifiers

Name: **Flu disease related**

Status: Running. 20 more needed to re-train.

Human-tagged items: 0 — [Go to human-tagged items »](#)

Machine-tagged items: 37,201 (since last change of the classifier)

Quality (AUC)*: **0.74**

*If AUC is lower than 0.8–0.9, or AUC is 1.0, you urgently need more human-tagged items.

Add a classifier to "Flu season"

Displaying available classifiers for your collection, except for those you are already using.

Standard classifiers

 Clusters v1.0	Description: Humanitarian Clusters, v1.0	Add
 Eye witness v1.0	Description: Eye witness accounts, v1.0	Add
 Individual needs v1.0	Description: Messages of type needs like water, food, blood.	Add
 Multimedia v1.0	Description: Multimedia resources, v1.0	Add

<http://aidr.qcri.org/>

AIDR – Classifier Setting (cont.)

Classifier "Flu disease related"

Details

History

Edit/Remove

Short name:

flu_related

Name:

Flu disease related

Type:

Custom

Edit

Remove Classifier

Tags



Name: Not related to flu situational awareness
Code: not_related
Description: If the message is of personal type, and does not convey useful information about disaster situational awareness.



Name: Yes: related
Code: yes_related
Description: If the message conveys useful information about flu situational awareness.



Name: N/A: does not apply, or cannot judge
Code: null
Description: If these categories do not apply to this message, or you cannot be sure about which is the correct category

Edit

Human Annotation in AIDR

Internal Tagging Interface

Tagger > Flu season > Flu disease related > New training example

Indicate the label for:

[Back to training data](#)

#California sees high rate of #flu deaths in unusually severe season #H1N1 <http://t.co/AgvzJ8wSJ>

- Not related to flu situational awareness** If the message is of personal type, and does not convey useful information about disaster situational awareness.
- N/A: does not apply, or cannot judge** If these categories do not apply to this message, or you cannot be sure about which is the correct category
- Yes: related** If the message conveys useful information about flu situational awareness.

[Save](#)

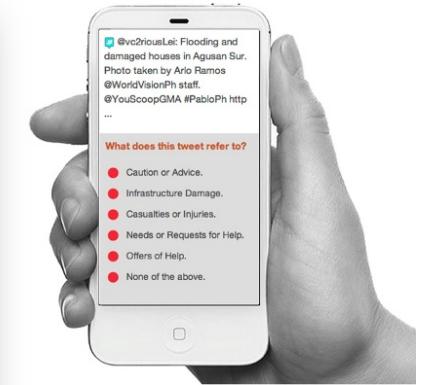
[Skip](#)

Human Annotation Using MicroMappers

MicroMapper Interface (web clicker)

The screenshot shows the MicroMapper web interface. At the top, there's a navigation bar with 'MicroMappers' (highlighted in red), 'Community', 'Applications' (which is the active tab), and 'About'. On the right, there's a 'Sign in' button. Below the navigation, a message from a user (@hakaaase) reads: '@hakaaase I'm pretty sure I've caught that super bad cold/flu/virus thing here so I'll be doing the same haha gonna scare my momma'. To the right of the message, there's a stats box: 'You are working now on task: 22611' and 'You have completed: 25 tweets from 166'. Below the stats is a 'Tip' section with a link to a 'Tutorial'. At the bottom left, there are three radio button options: 'Not related' (selected), 'N/A: does not apply, or cannot judge', and 'Yes related'. A 'Submit' button is at the bottom right.

Mobile clicker



Powered by  PyBossa

A project by  معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
Member of Qatar Foundation

Tagged Items and Machine Output

Training examples

Training data for "Flu disease related" in collection "Flu season"

Status: Running. Has classified 292 messages. [Details of running classifier »](#)

Add training data

198 training examples. Click on a message to see/edit details

Value	Text	Labeler	Action
Not related to flu situatio	Done with 3 liters of mineral water today. Seriously, flu spoils my day.	mimran15	<button>Delete</button>
Not related to flu situatio	@Jielimnnnn... flu abit only loop haissss	mimran15	<button>Delete</button>
Not related to flu situatio	@JanoAddicted I think I got the flu again..	mimran15	<button>Delete</button>
Not related to flu situatio	With an onset of flu-related symptoms and stress, don't be fucking dumb and do nothing about it	mimran15	<button>Delete</button>
Not related to flu situatio	Fuck this flu	mimran15	<button>Delete</button>
Not related to flu situatio	just gonna like eat cold and flu tablets lool	mimran15	<button>Delete</button>
Not related to flu situatio	@imbrucemaxwell I'm sick, Bruce. Maybe flu-ish. Let me whine today. #WAAAMBULANCE	mimran15	<button>Delete</button>
Not related to flu situatio	@Shelby_Jadee hahahaha don't!! Nope had woman flu and had work today couldn't be arsed wh:	mimran15	<button>Delete</button>
Not related to flu situatio	@blindnascarfan she was sick sick. Had the stomach flu.	mimran15	<button>Delete</button>
Yes: related	@thesynergyco here's a great natural cold and flu remedy using #manukahoney and #oiloforegar	mimran15	<button>Delete</button>
Yes: related	Save 30% on Flu immunizations at Walmart	mimran15	<button>Delete</button>
Yes: related	Make your own flu shot that's chemical free! #doterra #essentialoils @doterralarisha	http://t.co/V mimran15	<button>Delete</button>

Classifiers' output

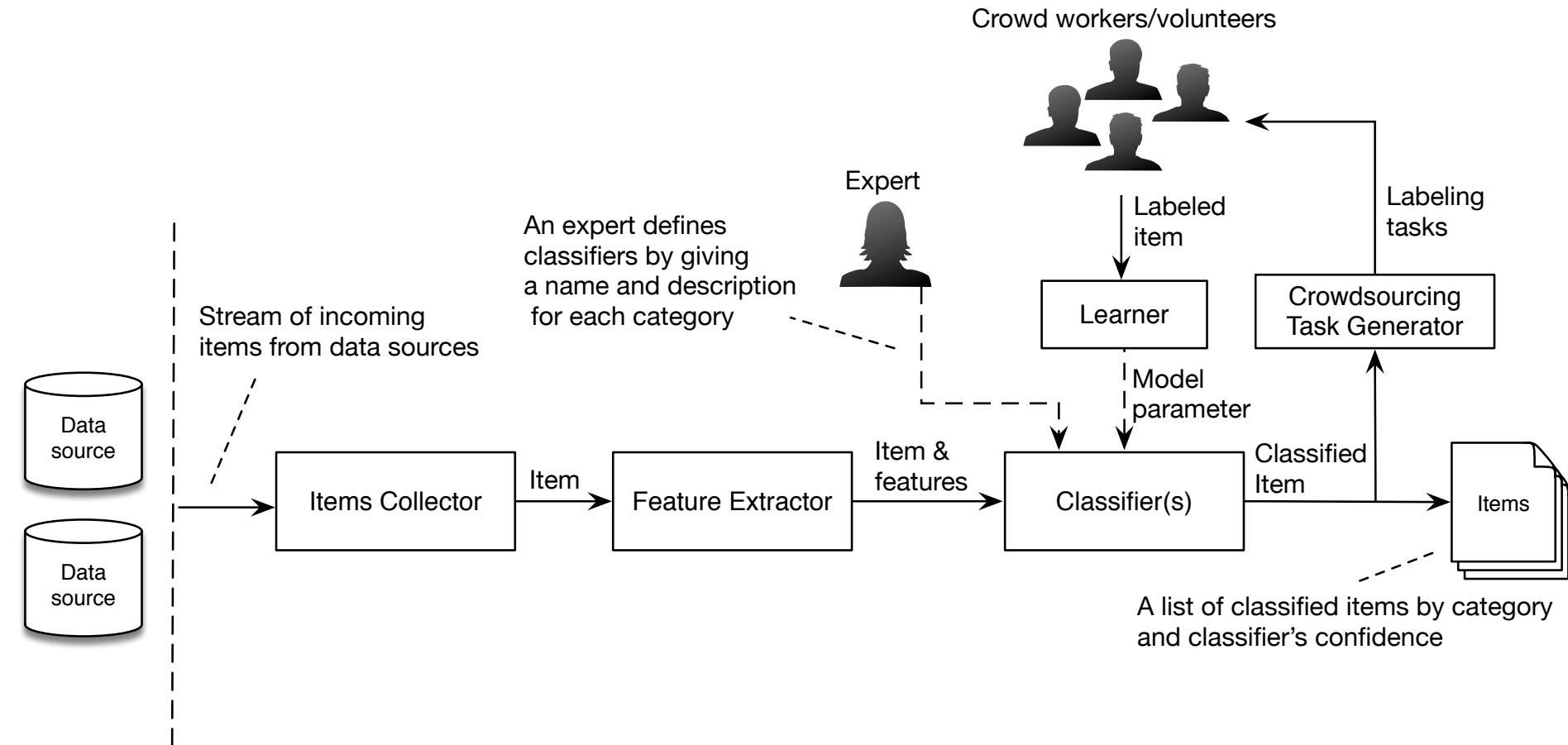
Flu season

Show 10 entries	Search:		
Tweet	Attribute	Label	Confidence
INew strain of 'deadly' bird flu http://t.co/qoBPu6o5zG	Flu disease related	Yes: related	1.00
#Innovation in flu #vaccines for season 2013-14; the four-strain flu vaccine explained #infographic http://t.co/j7GzFn9Q6e	Flu disease related	Yes: related	0.60
#SmalzysSSOSPlease follow me I will be forever great flu because I've been working to get followed OA99OC http://t.co/W1SezvvM0s	Flu disease related	Yes: related	0.60
@HolbornLaw @John_Cooper_QC Carbolic Smoke Ball name was sold to fish& chip shop, &Elizabeth Carlill died 50 years later, aged 96, of flu	Flu disease related	Not related to flu situational awareness	0.60
140210 Sooyoung UFO [Fan] The weather is very cold~ Be careful of flu^^ Things you can do is nothing than (cont) http://t.co/bp3bKAKV	Flu disease related	Not related to flu situational awareness	0.90
140210 Sooyoung UFO http://t.co/ODAGd397yf [Fan] The weather is very cold~ Be careful of flu^^ http://t.co/DKcvzN19jQ via @jamesic_jung	Flu disease related	Not related to flu situational awareness	0.80
140210 Sooyoung UFO http://t.co/VhLzbREHfg [Fan] The weather is very cold~ Be careful of flu^^ Things you (cont) http://t.co/s8jQwvchzZ	Flu disease related	Not related to flu situational awareness	0.90
90% sure I have the flu. Cool.	Flu disease related	Not related to flu situational awareness	0.89
:(' my poor lil bum,, just got a flu injection.	Flu disease related	Not related to flu situational awareness	0.90
@AtomPushers yeez, flu or what? :p	Flu disease related	Not related to flu situational awareness	1.00

Showing 1 to 10 of 157 entries

[First](#) [Previous](#) [1](#) [2](#) [3](#) [4](#) [5](#) [Next](#) [Last](#)

High-level Architecture



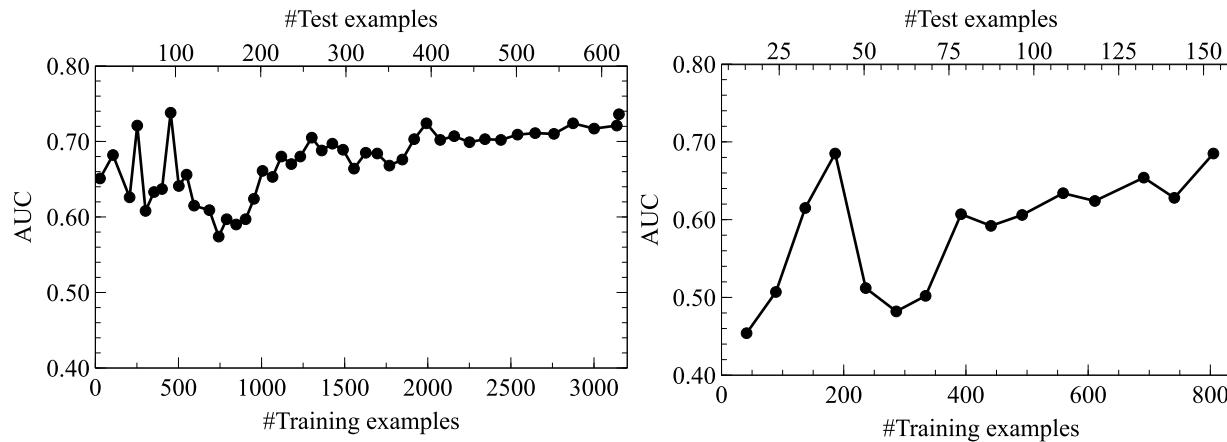
Quality, Cost, and Performance of AIDR

Quality vs. Cost in AIDR

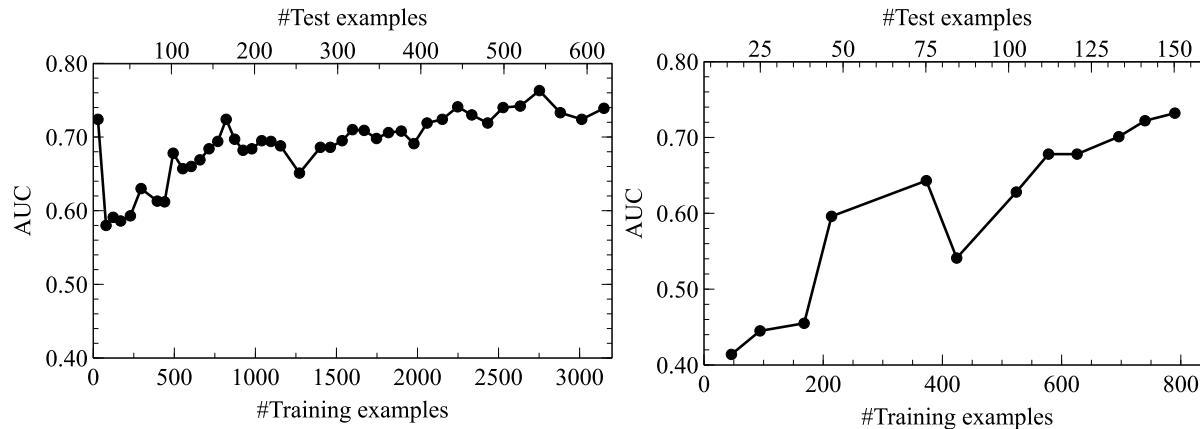
Goal: Maximizing quality while minimizing cost

- **Quality**
 - classification accuracy
 - Precision
- **Cost (human labels)**
 - monetary in case of paid-workers
 - time in case of volunteers

Quality vs. Cost in AIDR



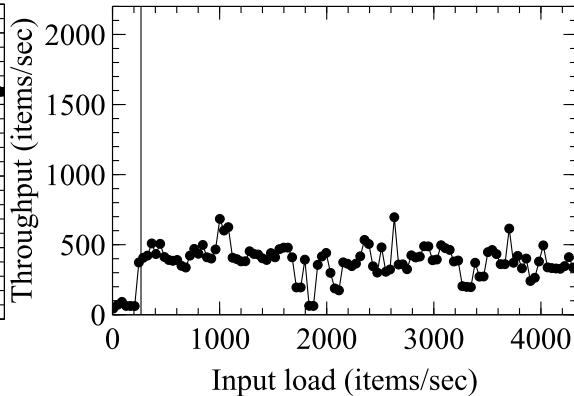
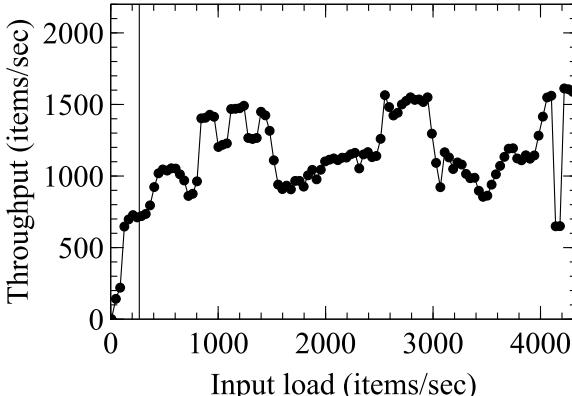
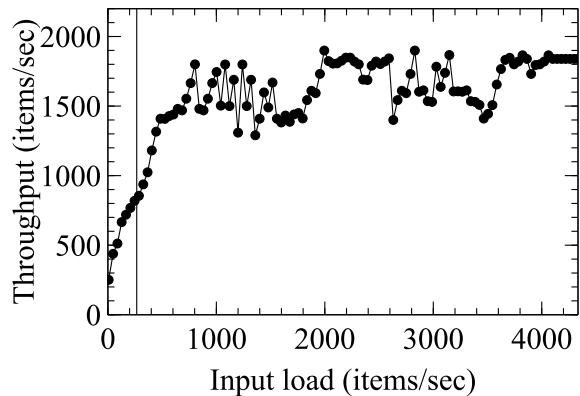
Quality vs. cost using passive learning and de-duplication



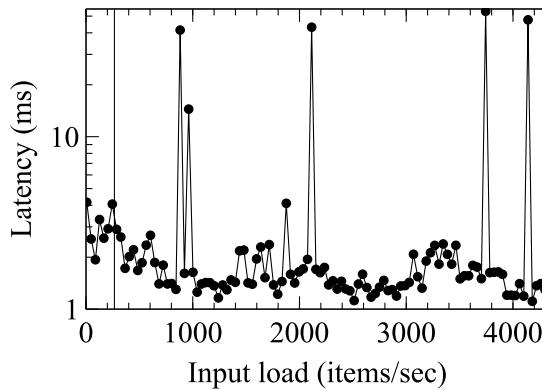
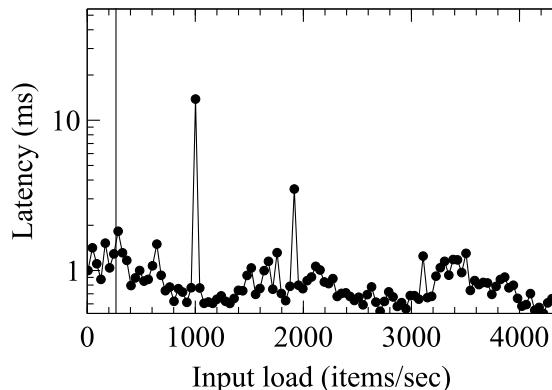
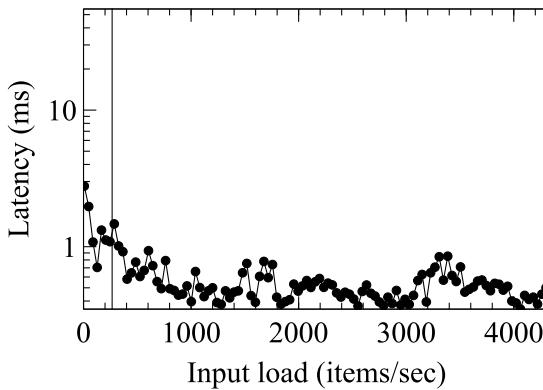
Quality vs. cost using active learning and de-duplication

Performance

In terms of throughput and latency



Throughput of feature extractor, classifier, and the system



Latency of feature extractor, classifier, and the system

Typhoon HAGUPIT (2014)

the guardian
Winner of the Pulitzer prize

UK world sport football opinion culture economy lifestyle fashion environment tech money travel

home > world europe US americas asia australia africa middle east

≡ browse all sections

Typhoon Hagupit

Typhoon Hagupit: UN using crowdsourcing platform to help assess damage

The United Nations is working with crowdsourcing platform, MicroMappers, to assess how much destruction has been caused by the latest typhoon to hit the Philippines

Caroline Bannock

[@carlanine](#)

Tuesday 9 December 2014 12.22 GMT

<Shares 309



Fishermen dock their boats at Manila Bay as Typhoon Hagupit (locally known as Ruby) approaches on 7 December 2014, Pasay City, Philippines. Photograph: Mark Cristino / Barcroft Media

At least 27 people have been killed and a million people evacuated after Typhoon Hagupit (locally known as Ruby) made landfall in the Philippines on Saturday night. Though Hagupit appears to be weakening as it approaches the Philippines capital, Manila, it's thought to have destroyed around 80% of all the homes along some coastal areas.

Like Typhoon Haiyan, last year, technology is again being used to help identify damage and needs assessment on the ground. The [United Nations Office for the Coordination of Humanitarian Affairs \(OCHA\)](#) in Manila has requested support

Advertisement

 AWS

Linux or Windows Server **FREE** for 1 Year

[Try it now](#)

replay

Most popular



India vs West Indies: Cricket World Cup - live | Dan Lucas and Simon Burnton

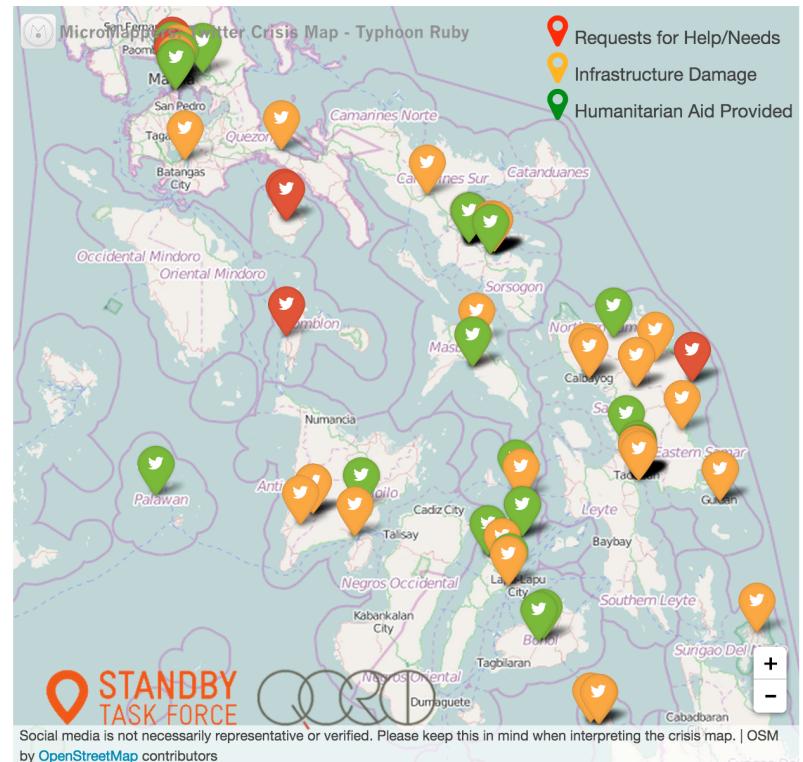


Harrison Ford hospitalised after crashing his plane on golf course



Nasa finds evidence of a vast ancient ocean on Mars

MicroMappers Map showing tweets which mention urgent needs, infrastructure damage and humanitarian aid regarding Typhoon Ruby.



As per the ICRC's guidelines on data protection and privacy, the Tweet Map does not include any personal identifying information (PII). This PII along with the raw tweets are only shared with the United Nations.

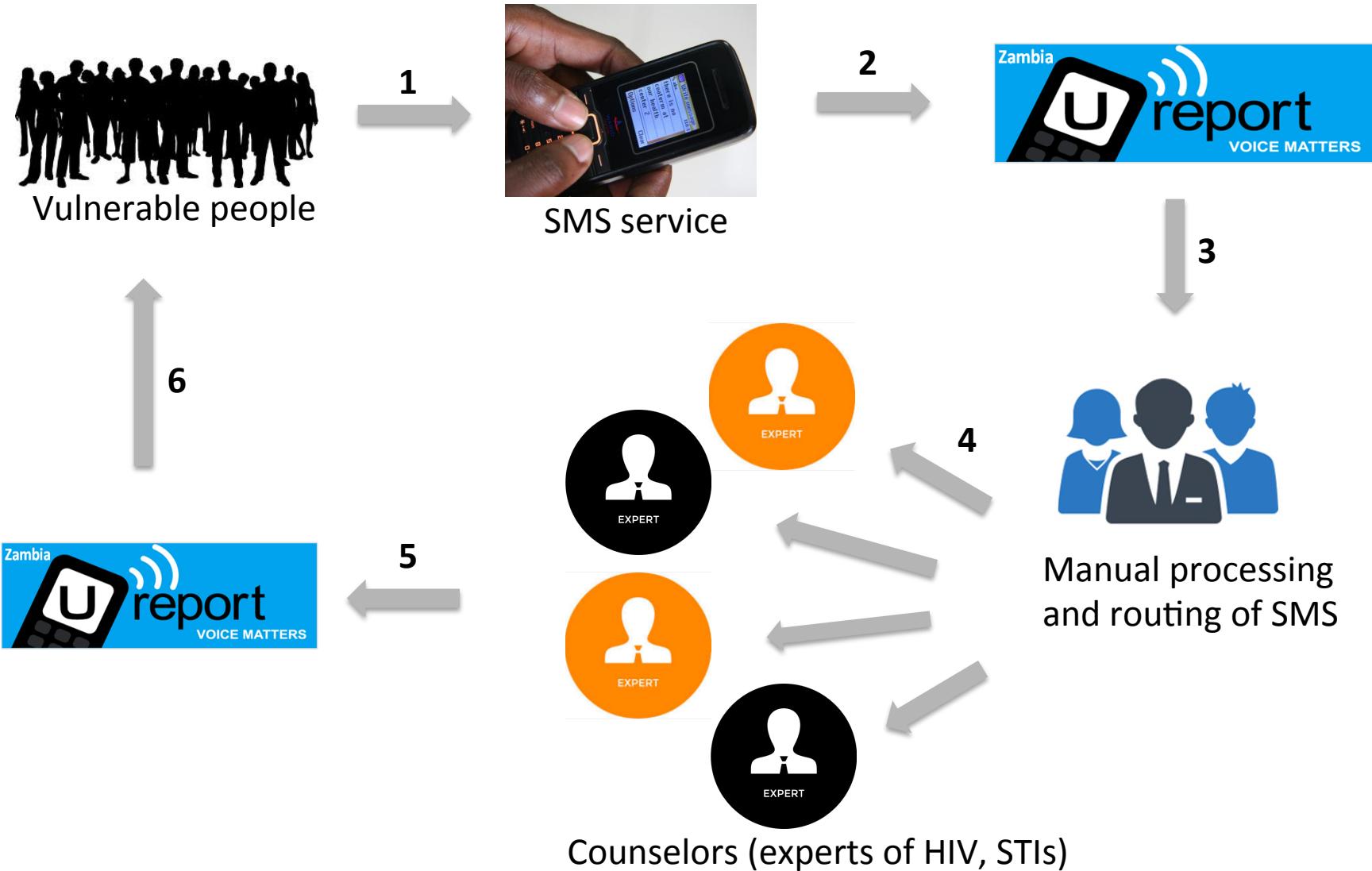
AI to Answer Health Queries via SMS

- Every hour Zambian youth get infected with HIV/AIDS
- UNICEF launched U-Report project in Zambia
- Usage of U-Report platform has recently increased 300%

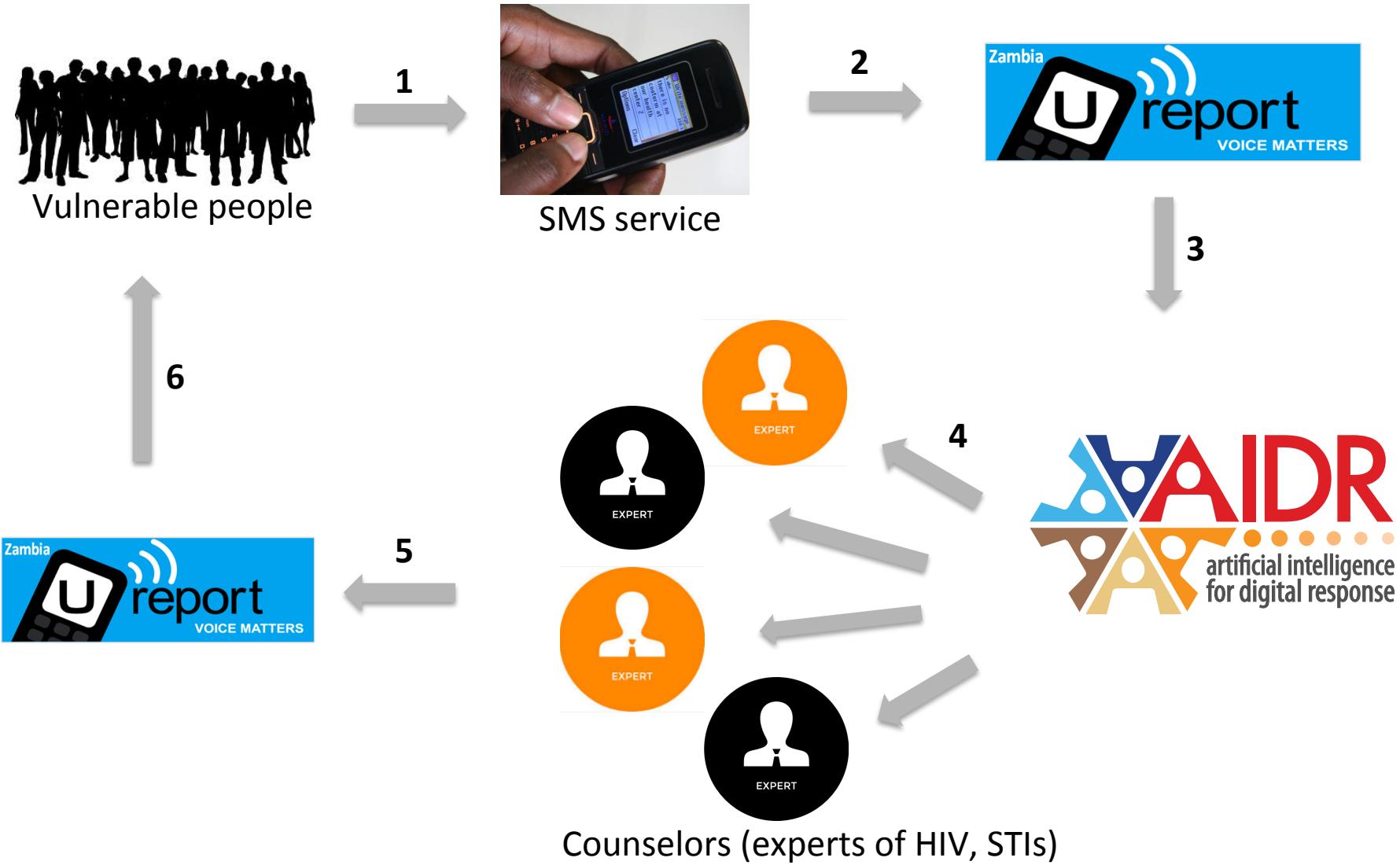
UNICEF U-Report and AIDR-SMS



UNICEF U-Report in Zambia



UNICEF U-Report in Zambia + AIDR



New Scientist Featured This Work



DAILY NEWS 4 April 2016

AI helps answer thousands of health queries in Zambia via SMS



SMS has proven an exceptionally useful technology in parts of Africa
GIANLUIGI GUERCIA/AFP/Getty Images

For many people in Zambia with health queries, sending a text message is the best way to get it answered. U-report, a free SMS-based service set up by UNICEF and run by volunteers, receives many thousands of questions a month, many specifically about [HIV and AIDS](#).

Also popular in Uganda, U-report has seen usage triple in the last three years, and about a thousand new users register every day. The volume of messages is growing so fast that the volunteers can't keep up, so UNICEF is testing software that reads and responds to many of the messages automatically.

In Zambia, there are roughly 27,000 new HIV infections a year, according to UNICEF, and 40 per cent of these are in those aged 15 to 24. With people constantly texting U-report for all kinds of HIV information and advice, the automated version [uses machine learning algorithms](#) to sort messages into eight categories: symptoms, HIV testing, treatment, pregnancy, transmission, prevention, definition, and male circumcision.

To train the system, [Patrick Meier](#), then at the Qatar Computing Research Institute in Doha, and colleagues fed in at least 50 messages for each category that had been selected by hand, and asked it to identify patterns that it could then use to do the sorting itself. As well as how to handle typos, the system learned to cope with textspeak such as "HOW 2 AVOID SPREADING HIV/AIDS 2 OTHERS?" and "I feelin bad becoz im th only one wh hs hiv wht shld i do?"

Media Coverage

nature

Forbes

Mashable

WIRED.CO.UK

New
Scientist

GULF TIMES



THE WALL STREET JOURNAL.
WSJ

Ongoing Work

Human Annotation

Selection and scheduling for supervised classification system

Human Annotation - Challenges

Crowdsourcing is a big research topic. We address two challenges here:

1- Labeling task selection

- Which tasks to pick for labeling?
- No duplicate tasks should be labeled
- Prioritize tasks that are likely to increase classifier's accuracy

Twitter Crises Datasets

1. Joplin-2011

- Consists of 206,764 tweets collected using (#joplin)

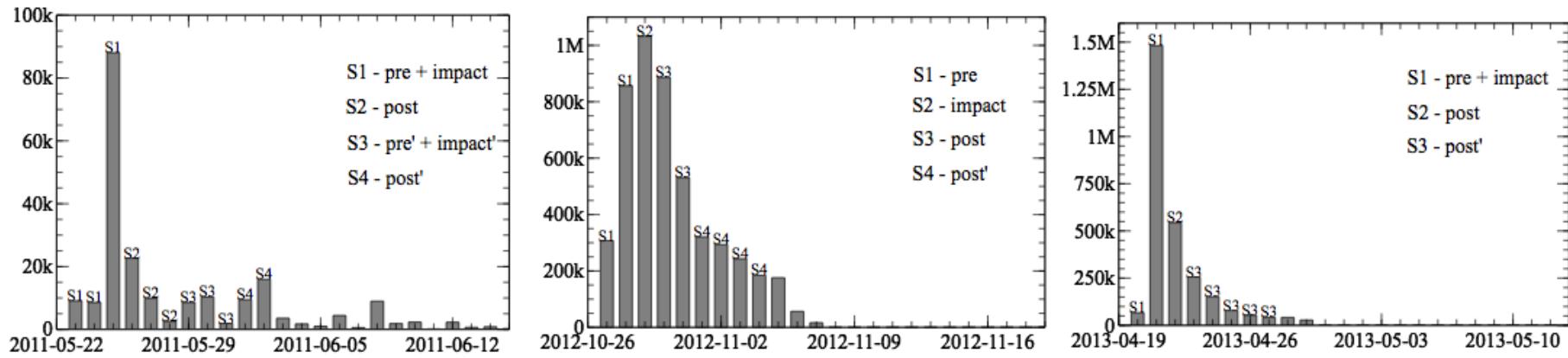
2. Sandy-2012

- Consists of 4,906,521 tweets collected using (#sandy, hurricane sandy, ...)

3. Oklahoma-2013

- Consists of 2,742,588 tweets collected using (Oklahoma, tornado, ...)

Distribution of Tweets into Phases



Joplin (left), Sandy (center), and Oklahoma (right). Number of tweets per day in all datasets.

Pre: preparedness phase

Impact: phase corresponds to the period in which the main **effects** are felt

Post: corresponds to **response and recovery phase**

Labeling Task Selection

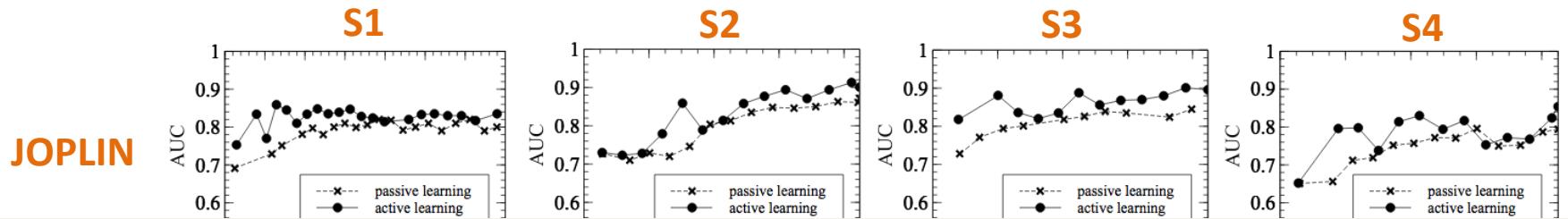
Experiment: Is de-duplication necessary?

- 29-74% of tweets are re-tweets & 60-75% are near duplicates [Rogstadius et al. 2011]
- Duplication causes an artificial increase in accuracy
- Necessary to reduce classifier bias. Otherwise learning on a fewer concepts
- Necessary to improve workers experience

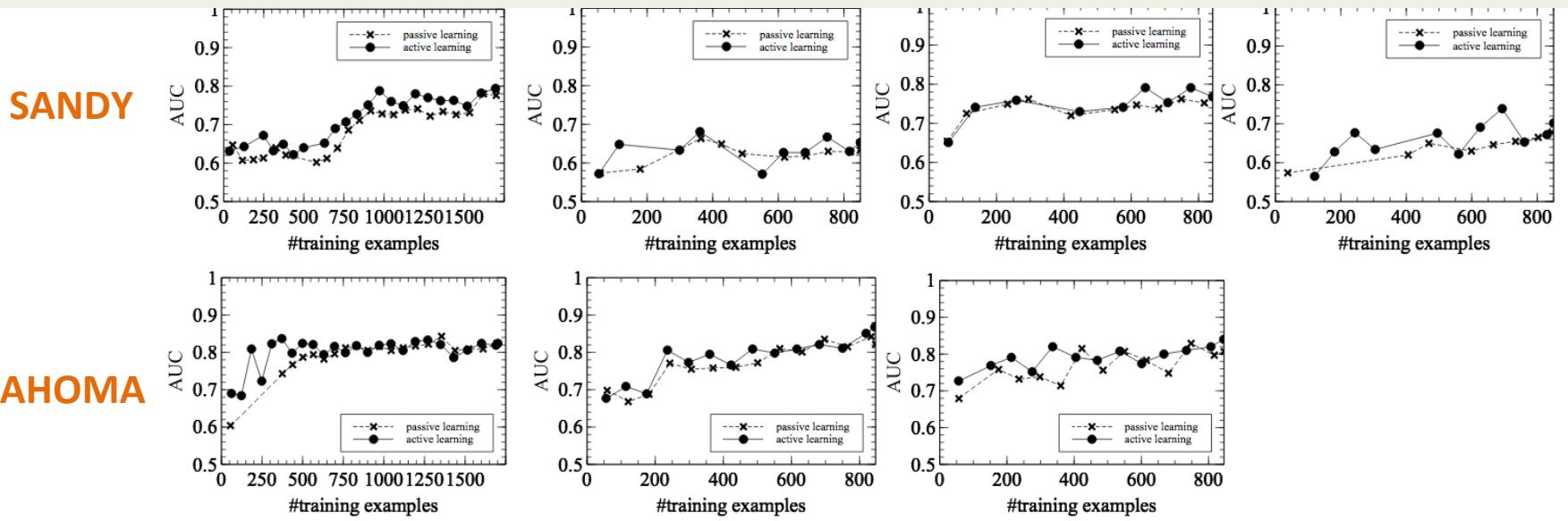
Phase	Train	Phase	Test	AUC (without de-duplication)	AUC (with de-duplication)
S1 (pre)	1,500	S1 (pre)	500	0.78	0.74
S1 (pre)	500	S1 (pre)	500	0.73	0.72
S2 (impact)	500	S2 (impact)	500	0.80	0.72
S3 (post)	500	S3 (post)	500	0.79	0.73
S4 (post')	500	S4 (post')	500	0.70	0.64

Labeling Task Selection

Experiment: Passive learning vs. Active learning



AUC stabilize with fewer training items using active learning



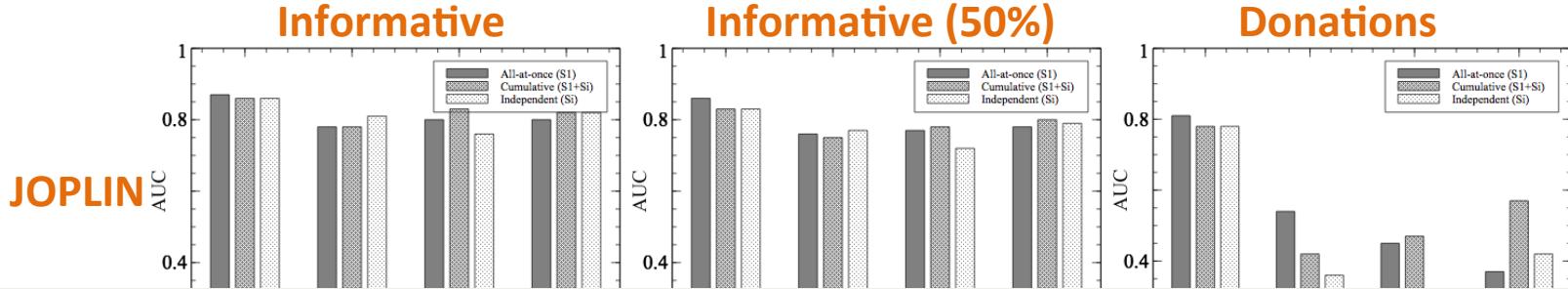
Labeling Task Scheduling

2- Labeling task scheduling

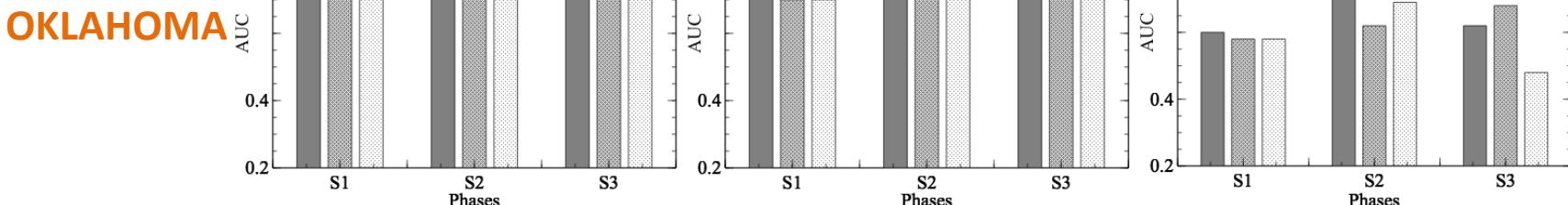
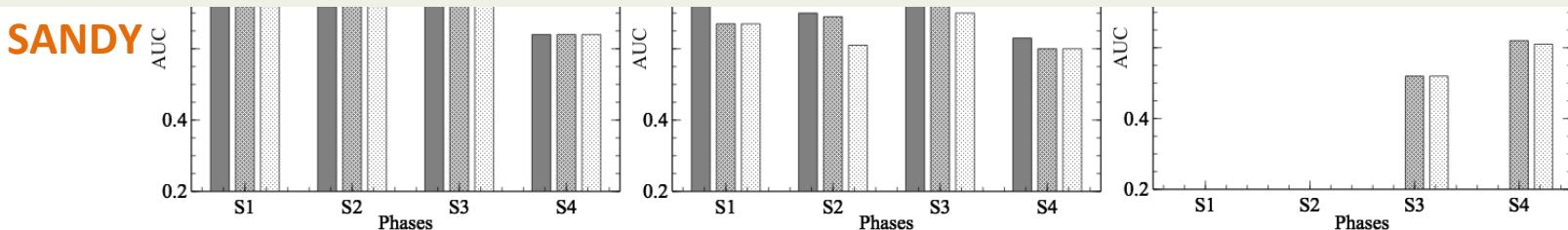
- **All-at-once labeling**
 - Obtain 1,500 labels on S1 and use all for training
- **Cumulative labeling**
 - Obtain 500 labels in each of S1, S2, and S3 and train on labels available up to each phase
- **Independent labeling**
 - Obtain 500 labels in each of S1, S2, and S3 and use the most recent labels for training, discarding old.

Labeling Task Scheduling

Experiment: Which labeling strategy to follow?



All-at-once approach dominates in informative and cumulative strategy seems better for donations



Ongoing Work

Domain Adaptation

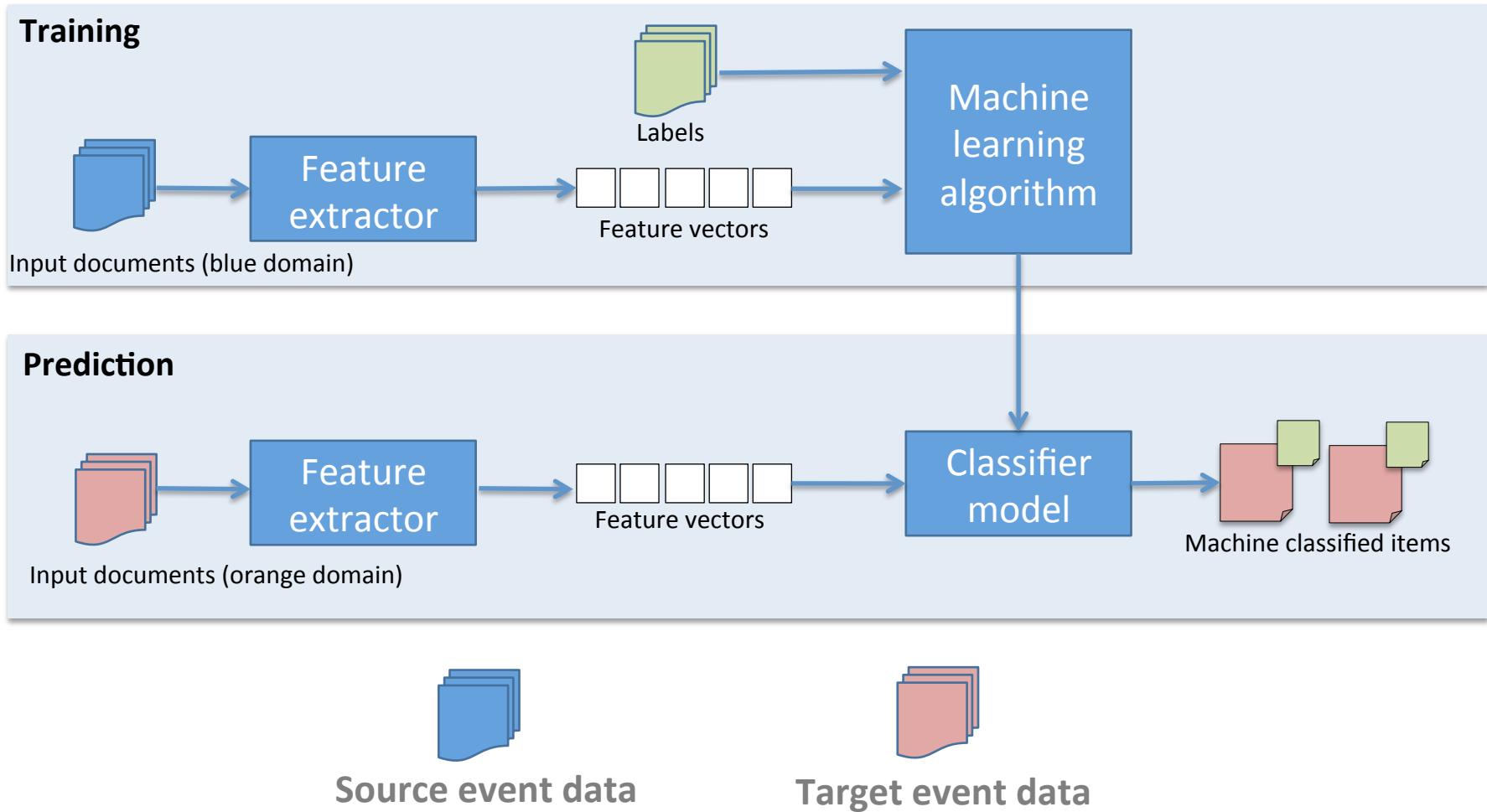
Ability of a system to apply knowledge and skills learned in previous domains to novel domains

Our Goal:

To build a system that can understand natural language

Domain Adaptation

Labeled source, but unlabeled target



Same Domain Learning



Source domain

Apples

Oranges

Target domain

Apples

Oranges



BUT

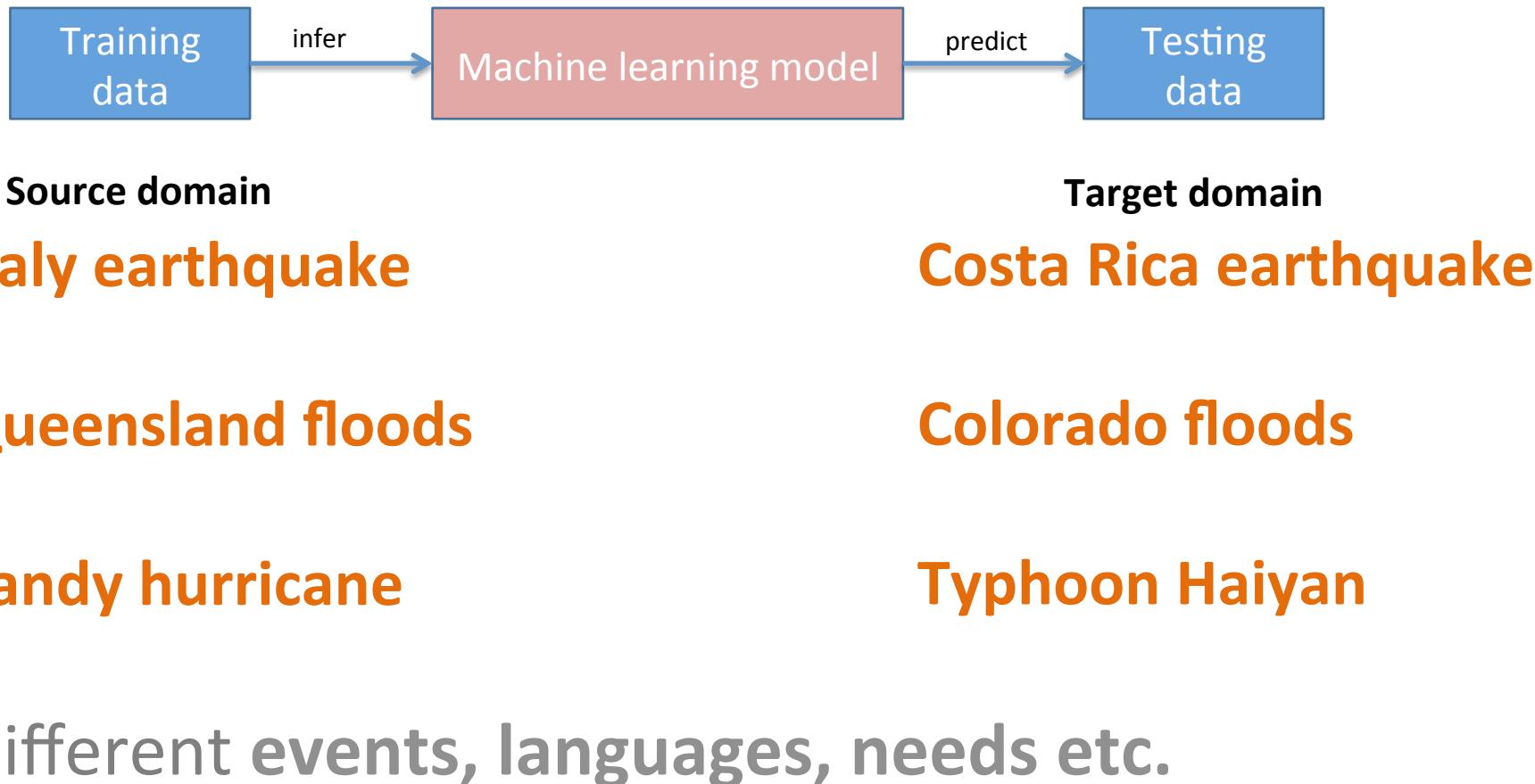
Apples

Oranges



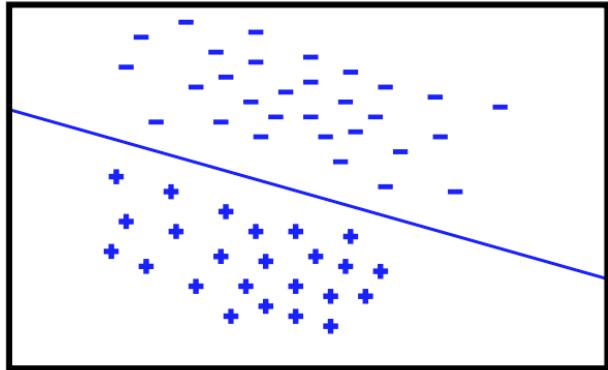
Different shapes, colors, skins, tastes, etc.

Crisis-related Data Classification

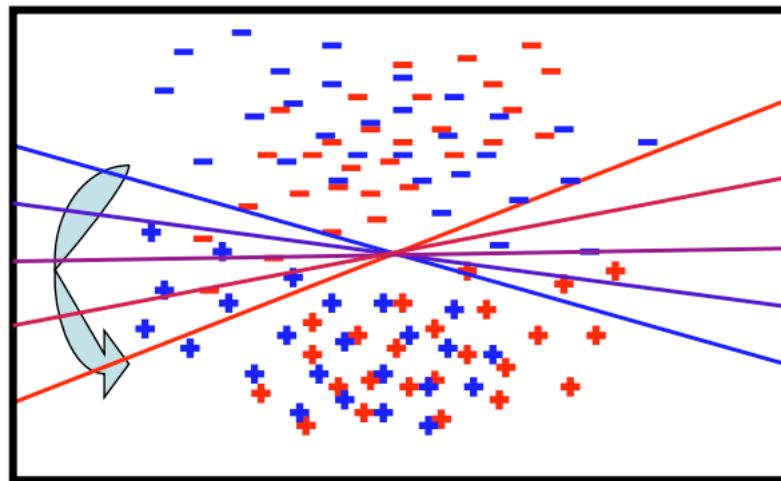
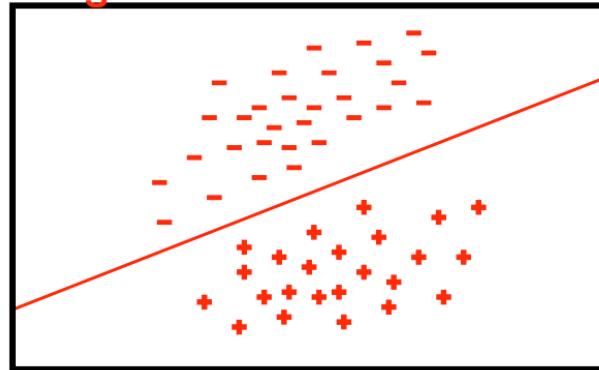


Domain Adaptation

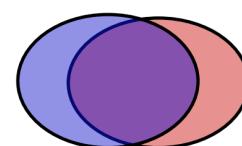
source domain



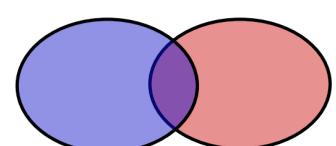
target domain



Easy



Hard



Model Adaptation Experiments

- Model adaptation **using single source**
 - Using both: in-domain and cross-domain
- Model adaptation using **multiple sources**
 - In-domain
 - Multiple source events without the target
 - Multiple source events with the target
- Model adaptation in **special cases**
 - Same languages
 - Similar languages

Observations & Findings

- Data from early hours of a crisis help
- Past events of same type are useful
- Same language data as target event is also useful
- Similar languages are also useful
- Cross-domain training does not show significant improvements

Future Directions

Rapid Crisis Response

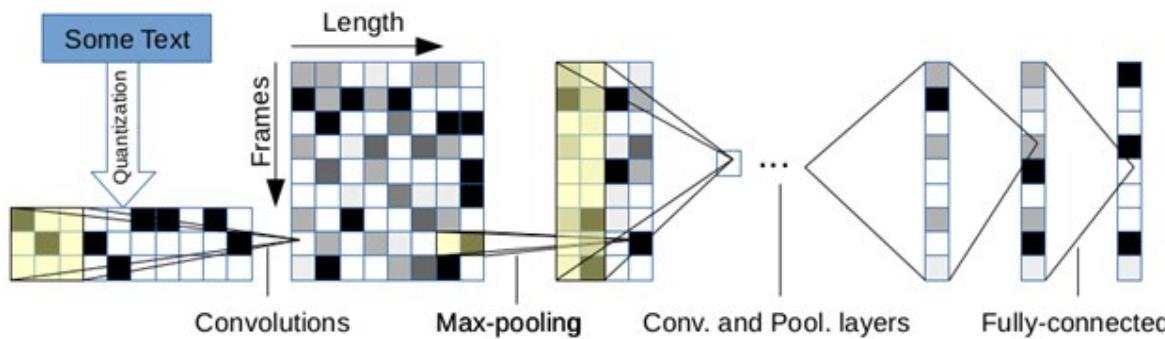
Image Processing for Damage Assessment



Tasks

- **Image categorization**
 - E.g. building, bridge, road damage
- **Damage and severity assessment**
 - Given a damage image, identify severity of damage (low, mild, high)

Deep Learning to Improve Classification



Tasks

- Improve text classification performance
 - Availability of big data
 - Automatic features learning
 - Binary and multi-class classification
 - Testing data from multiple past events

Transfer Learning

Differences in classification tasks:

- Different classification tasks
- Different types of disasters, stakeholders, information needs

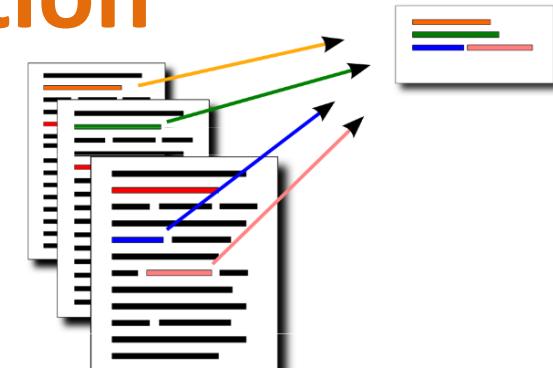
Task:

- **Learn from source to classify target**
 - Semantic similarity between tasks
 - Instances similarity between domains
 - Instance weighting

Summarization and Prioritization of Actionable Information

Information needs & problem:

- Different stakeholders
- Different goals, requirements, and info. needs



General situational awareness vs. Target situational awareness

- High-level general updates from an event
- Specific updates (infrastructure damages)

Resources, Datasets, And Tools

Towards Standard Baselines and Datasets



- Access to **52 million tweets**
- Around **50k labeled tweets** into humanitarian categories
- Largest **word2vec embeddings** trained on 52m crisis-related tweets
- **Out-of-vocabulary dictionaries**

CrisisNLP.qcri.org

Towards Standard Baselines and Datasets

[Home](#)[Download Collections](#)[Create a Collection](#)[Download Lexicons](#)[Create a Lexicon](#)

CrisisLex.org

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.

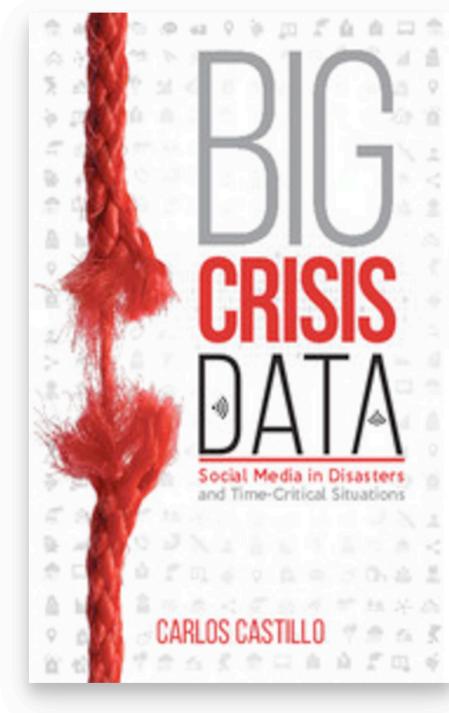
[Collections of Crisis Data »](#)[Lexicons of Crisis Terms »](#)

Upcoming Book

BIG CRISIS DATA

Social Media in Disasters and Time-Critical Situations

Upcoming book by Carlos Castillo. Cambridge University Press, 2016.



Volume: Data Acquisition,
Storage, and Retrieval

Volunteers: Humanitarian
Crowdsourcing

Vagueness: Natural
Language and Semantics

Veracity: Misinformation
and Credibility

Variety: Classification and
Clustering

Validity: Biases and Pitfalls
of Social Media Data

Virality: Networks and
Information Propagation

Visualization: Crisis Maps
and Beyond

Velocity: Online and Real-
Time Processing

Values: Privacy and Ethics

Digital Humanitarians: Book

DIGITAL HUMANITARIANS

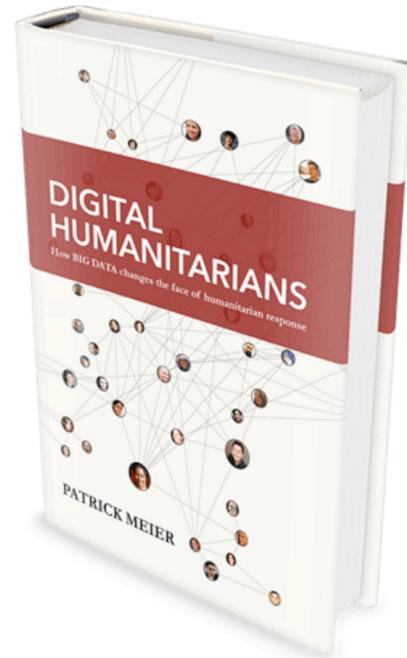
How Big Data is Changing the Face of Humanitarian Response

The overflow of information generated during disasters can be as paralyzing to humanitarian response as the lack of information. This flash flood of information is often referred to as Big

Data, or Big Crisis Data. Making sense of Big Crisis Data is proving to be an impossible challenge for traditional humanitarian organizations, which is precisely why they're turning to Digital Humanitarians.

Written by: [@PatrickMeier](#)

Book hashtag: [#DigitalJedis](#)



ACM Computing Survey

Processing Social Media Messages in Mass Emergency: A Survey [Imran et al. 2015]



ACM Computing Surveys (CSUR)

publishes comprehensive, readable tutorials and survey papers that give guided tours through the literature and explain topics to those who seek to learn the basics of areas outside their specialties.



Conclusions

- **Information bestows power for disaster response**
 - People need information as much as water, shelter, and food
 - Disasters are unavoidable, but planning can lessen their effects
- **Social media as time-critical information source**
 - Early warnings, event detection, event monitoring
 - Availability of information opens new opportunities
- **Artificial Intelligence for Disaster Response**
 - Applied research at its best
 - AI + humans-in-the-loop can enable rapid crisis response
 - AI techniques useful for:
 - Situational awareness
 - Actionable information extraction
 - Summarization

THANK YOU!

Muhammad Imran

 <http://mimran.me/>

 Follow @mimran15



AIDR.qcri.org



CrisisNLP.qcri.org