Transforming Social Big Data into Timely Decisions and Actions for Crisis Mitigation and Coordination

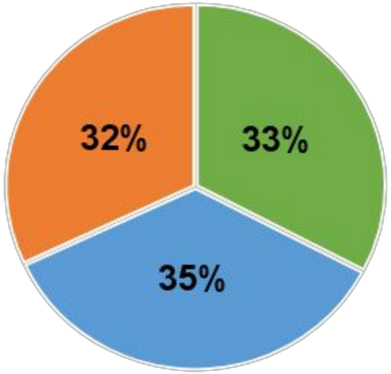
Keynote @ Exploitation of Social Media for Emergency Relief and Preparedness (SMERP)
Co-located with: The Web Conference 2018 (formerly WWW)
Lyon, France. 23 April 2018

Prof. Amit Sheth
LexisNexis Ohio Eminent Scholar
Exec. Dir. - Kno.e.sis @ Wright State University

Kno.e.sis: Ohio Center of Excellence in Knowledge-enabled Computing & BioHealth Innovation

NSF
- ★ Harassment on Social Media
- ★ Citizen & Physical Sensing
- ★ Twitris - Collective Intelligence
- ★ Aerial Surveillance
- ★ Visual Experience
- ★ Web Robot Traffic

NIH
- ★ kHealth - Asthma
- ★ eDrugTrends
- ★ eDarkTrends
- ★ Depression on Social Media
- ★ Drug Abuse Early Warning



DoD & Industry
- ★ Metabolomics & Proteomics
- ★ Medical Info Decisions
- ★ Human Detection on Synthetic FMV
- ★ Sensor & Information
- ★ Material Genomics
- ★ Cardiology Semantic Analysis

15 faculty from 4 colleges + ~60 Funded Students
- ★ **40 PhD**
- ★ **16 MS**
- ★ **10 BS**

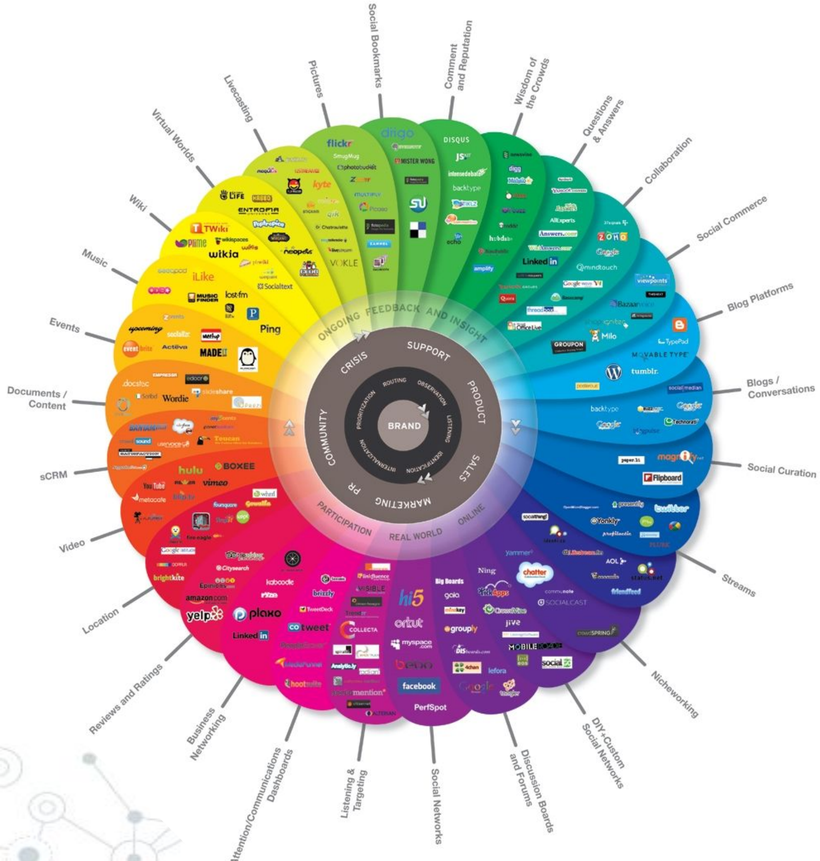
Kno.e.sis conducts research in AI techniques that convert physical-cyber-social big data into smart data, enabling building of intelligent systems for clinical, biomedical, policy, and epidemiological applications.

Example clinical/healthcare applications include major diseases such as asthma, obesity, depression, cardiology, dementia and GI.

This is complemented by social and development challenges such as marijuana legalization policy, harassment on social media, gender-based violence, and disaster coordination.

Kno.e.sis' research in World Wide Web ranks Wright State University among the top 10 organizations in the world based on 10-yr impact [MAS: 2016]. Its total budget for currently active projects is $13+ million [2017]. World-class interdisciplinary research is complemented by exceptional student outcomes and commercialization with local economic impact.

Never before humanity is so connected



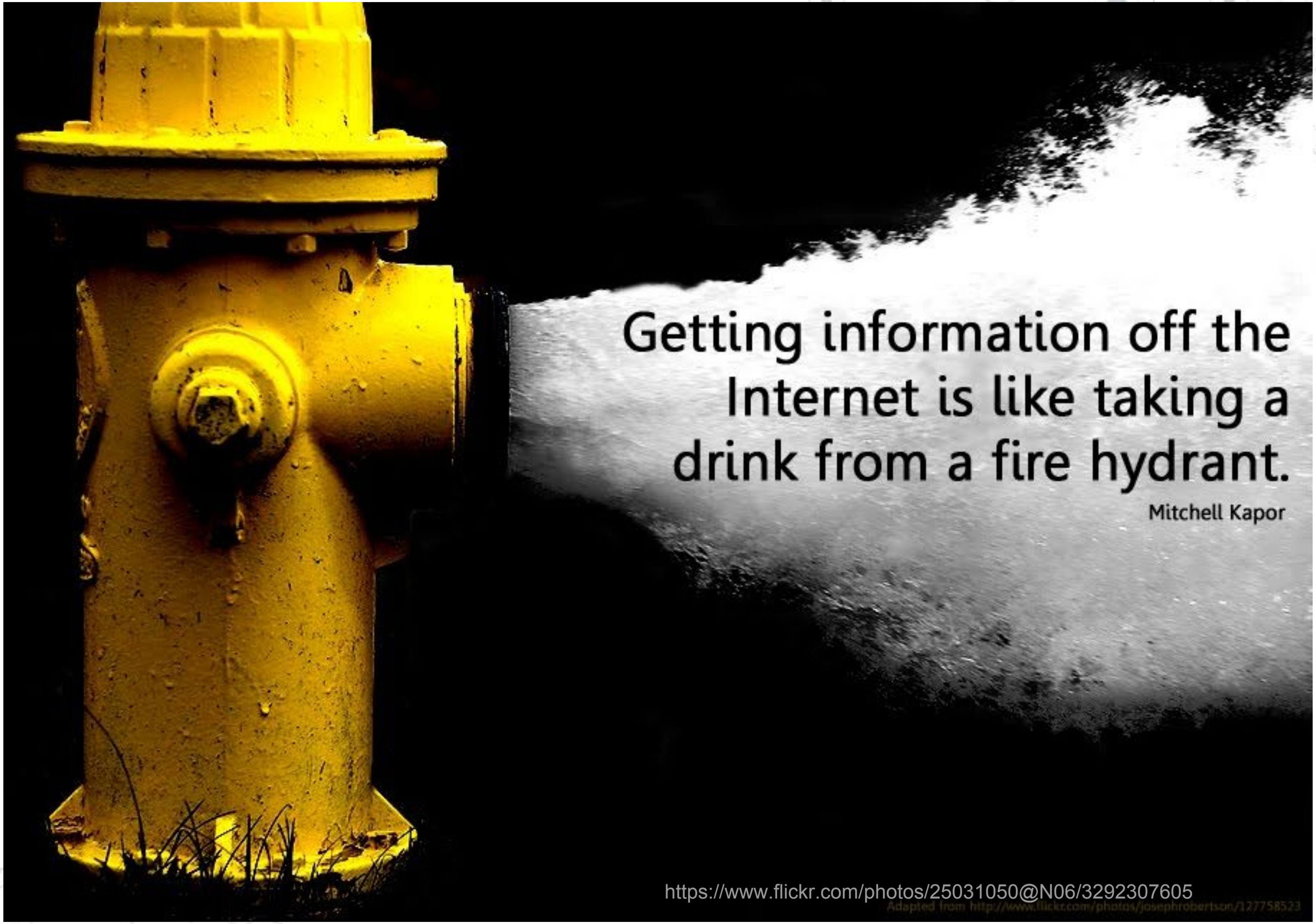
https://upload.wikimedia.org/wikipedia/commons/7/7c/Conversationprism.jpeg


Semantics & Services

Citizen Sensing, Social Signals, and Enriching Human Experience

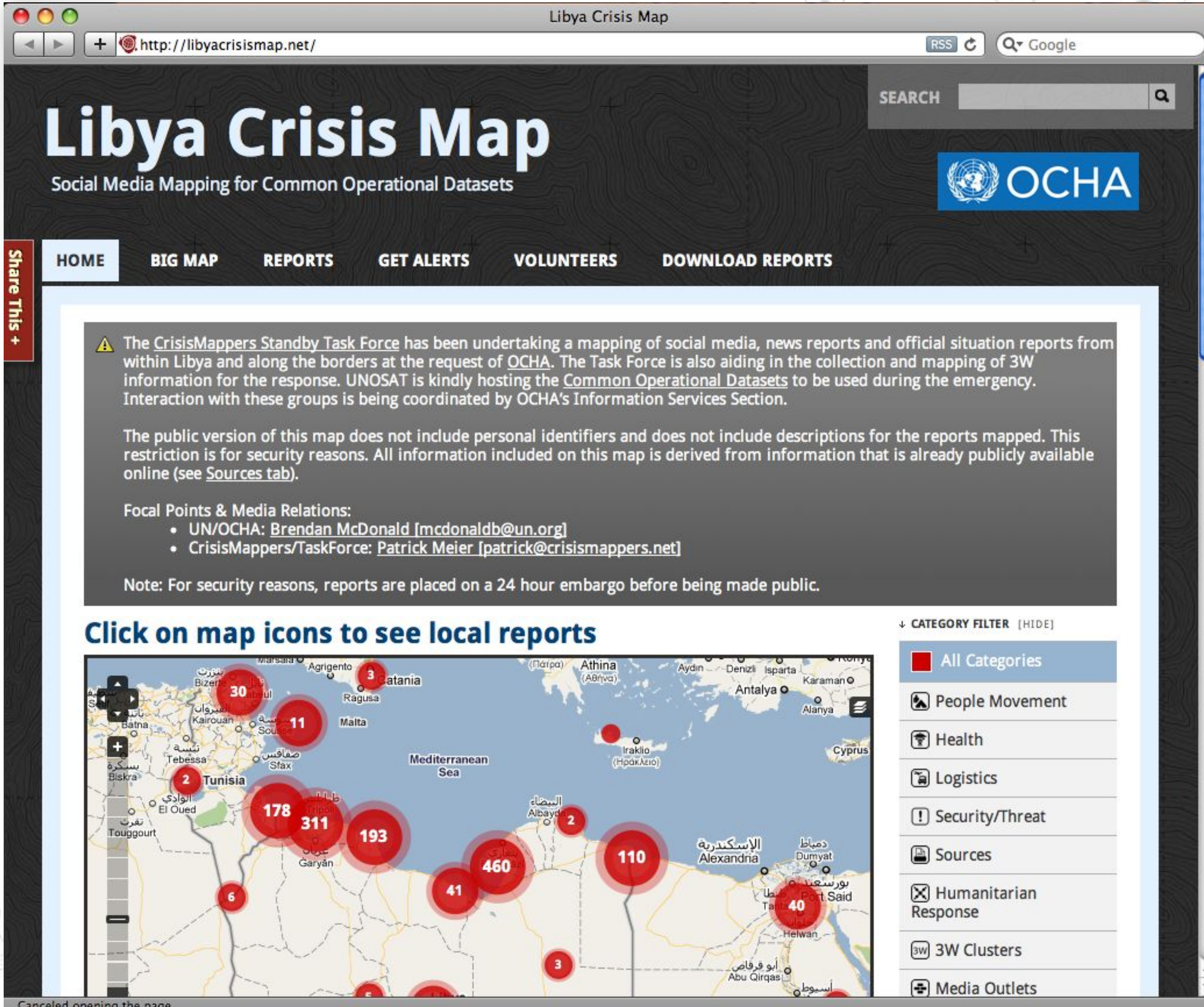
Amit Sheth • Kno.e.sis Center, Wright State University

IEEE Internet Computing, 2009



https://www.flickr.com/photos/25031050@N06/3292307605

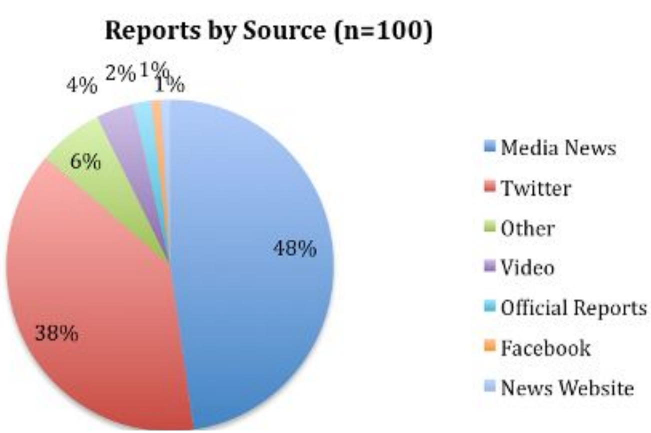

twitter
Login Join Twitter!
Social media is critical for #humanitarian work & now you can see why. Crisis Map of #Libya is now public: http://bit.ly/g8xCtm #UN #OCHA
about 1 hour ago via Tweet Spinner
Retweeted by 100+ people
UN
United Nations
© 2011 Twitter About Us Contact Blog Status Resources API Business Help Jobs Terms Privacy

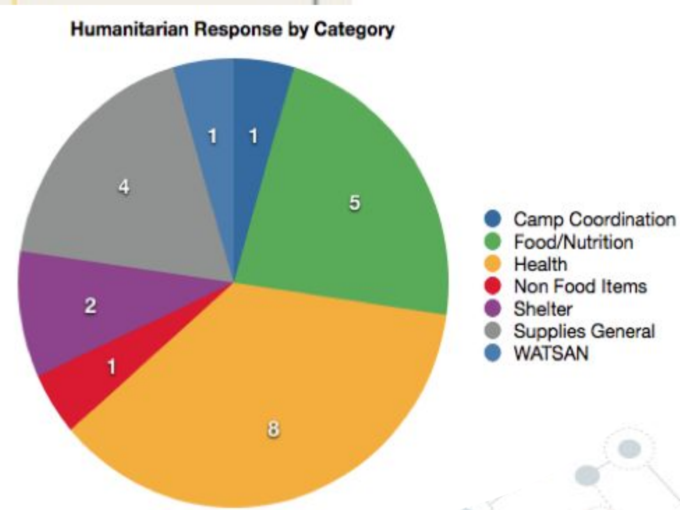


Manual Activity


Incidents of armed attacks, armed threats, injured people, violence, and deaths
Medenine
Tataouine
Al-Khums
Albayda
Derna
Tubruq
Mizdah
District
Bani Waled
Surt
Libya


Border crossings and Evacuations
Mediterranean
Sea


Reports by Source (n=100)
4%
2%
1%
1%
6%
48%
38%
Media News
Twitter
Other
Video
Official Reports
Facebook
News Website


Humanitarian Response by Category
1
1
4
5
2
1
8
Camp Coordination
Food/Nutrition
Health
Non Food Items
Shelter
Supplies General
WATSAN

http://bit.ly/nP1E4q

News

Twitter, Facebook become lifelines after Japan quake

Social networks at their best as people turn to them after massive quake, tsunami

http://cnet.co/jdQgME

MAY 6, 2011 5:51 PM PDT

Japan radiatio

by Eric Mack

Recommend 91

Safecast.org seeks to aggregate worldwide sensor information

(Credit: Screen capture by Eric Mack/CNET)

Image:http://bit.ly/fl4gEJ

ake

Mar. 2011

Currently tracking about 4300 records.

Short URL : http://goo.gl/sagas (Mobile OK)

Additional Partners

Other Resources

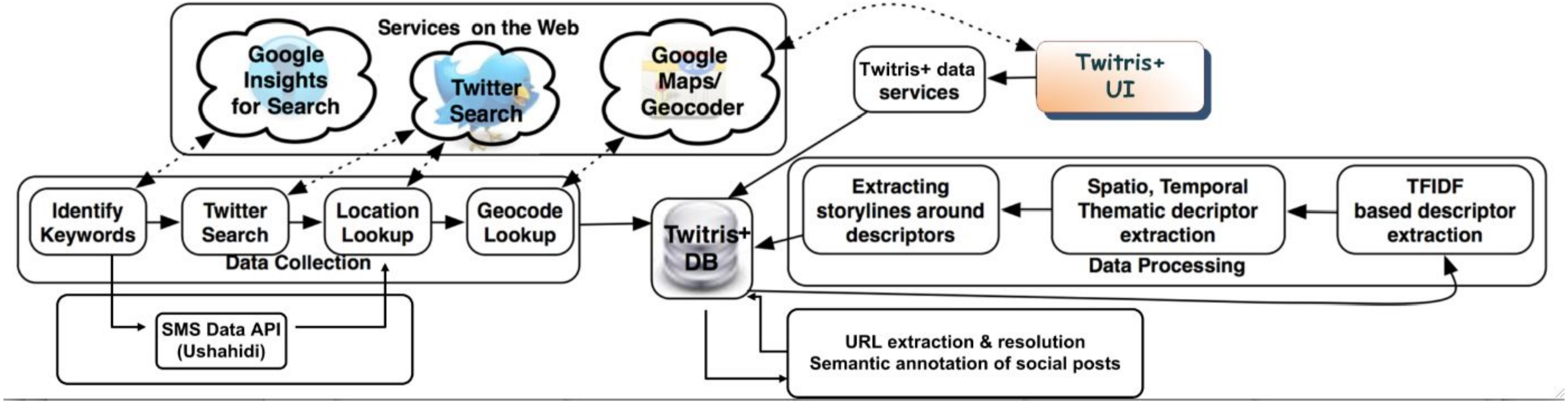
http://bit.ly/gWboib

Project: Social Media Enhanced Organizational Sensemaking in Emergency Response

Twitris v1: Spatio-Temporal-Thematic

Facilitates understanding of multi-dimensional social perceptions over SMS, Tweets, multimedia Web content, electronic news media

Twitris v1: Architecture



Meenakshi Nagarajan, Karthik Gomadam, Amit Sheth, Ajith Ranabahu, Raghava Mutharaju and Ashutosh Jadhav, Spatio-Temporal-Thematic Analysis of Citizen-Sensor Data - Challenges and Experiences,' Tenth International Conference on Web Information Systems Engineering, 539 - 553, Oct 5-7, 2009.

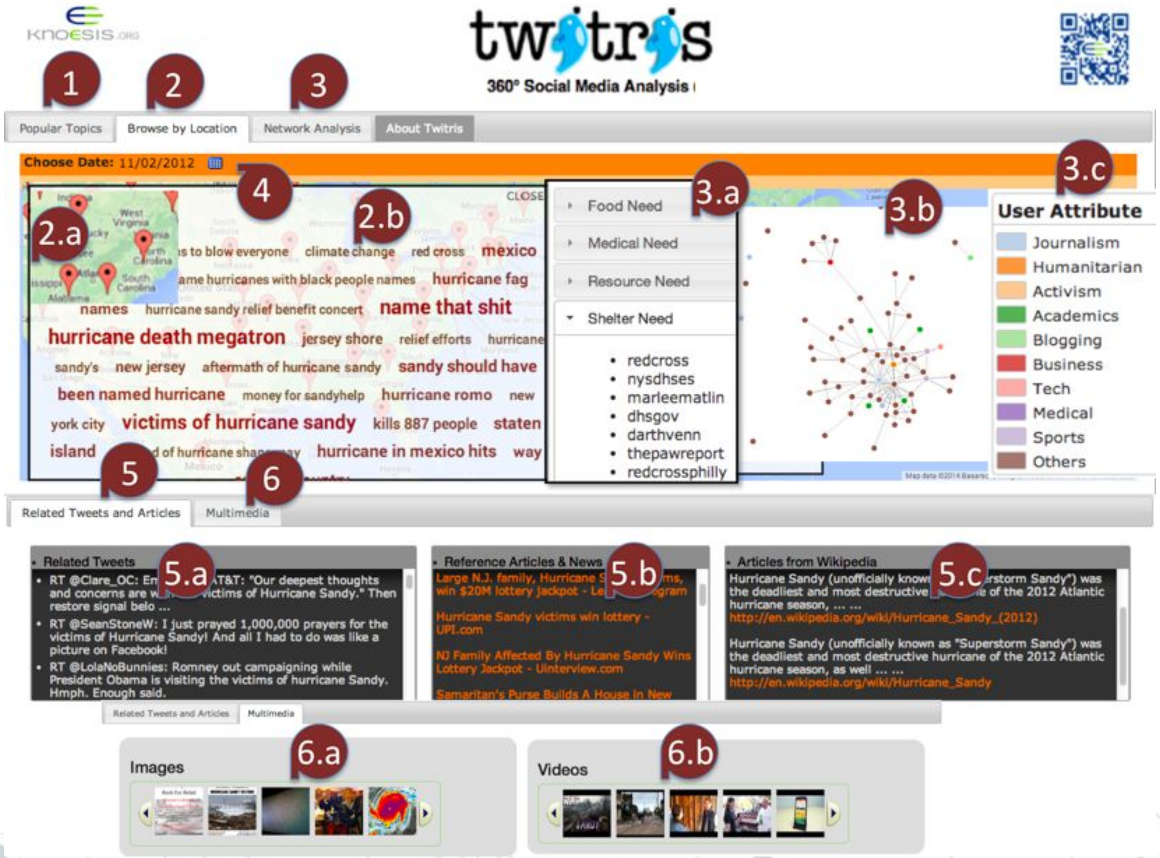
But the amount of data has grown substantially

	Description	# Tweets (# Days)	
Social data collected mainly by Twitris in the teams' prior research	Haiti Earthquake 2010	0.6 million (57)	
	Hurricane Sandy 2012	4.9 million (12)	
	Oklahoma Tornado 2013	2.8 million (10)	
	Chennai Flood 2015	0.41 million (37)	
	Houston Flood 2016	1.7 million (8)	
	Louisiana Flood 2016	0.42 million (10)	
	Hurricane Harvey 2017	4 million (41)	
	Hurricane Irma 2017	4.5 million (35)	
	Earthquake Mexico City 2017	0.4 million (35)	
DEEP	**Crisis**	**# Docs**	**# Excerpts**
	Haiti Disaster	297	863
	Syria Conflict	1788	3988
	South Sudan Conflict	1633	2961
	Democ. Repub. of Congo Disaster	1425	2610
	Nigeria Complex Emergency	1235	2089
	Somalia Disaster	1081	2471
	Yemen Conflict	1034	2223
	Sudan Complex Emergency	1015	1875
	Libya Conflict	954	2024

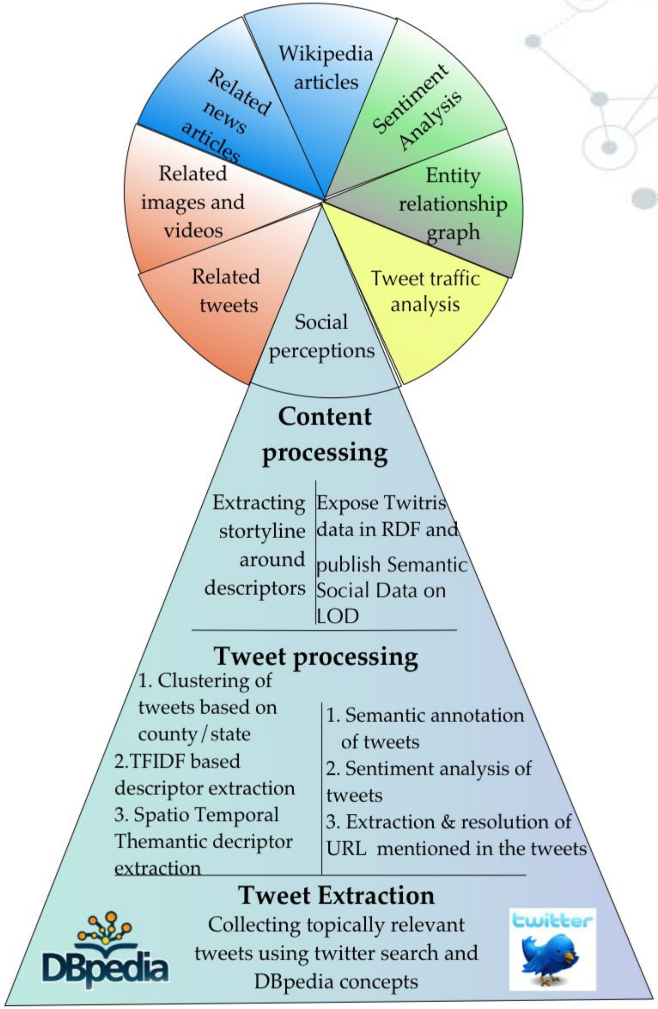
Humanitarian agency reports

DEEP is a platform for collaborative secondary data collection, analysis and dissemination for humanitarian crises. DEEP (thedeep.io) is a joint initiative by seven key humanitarian organizations: UNOCHA, UNHCR, OHCHR, IDMC, JIPS, ACAPS and IFRC.

Twitris v2: People-Content-Network



Twitris v2: Functional Overview



Functional overview of Twitris 2.0

Sample of Real-World Impact & Media Coverage

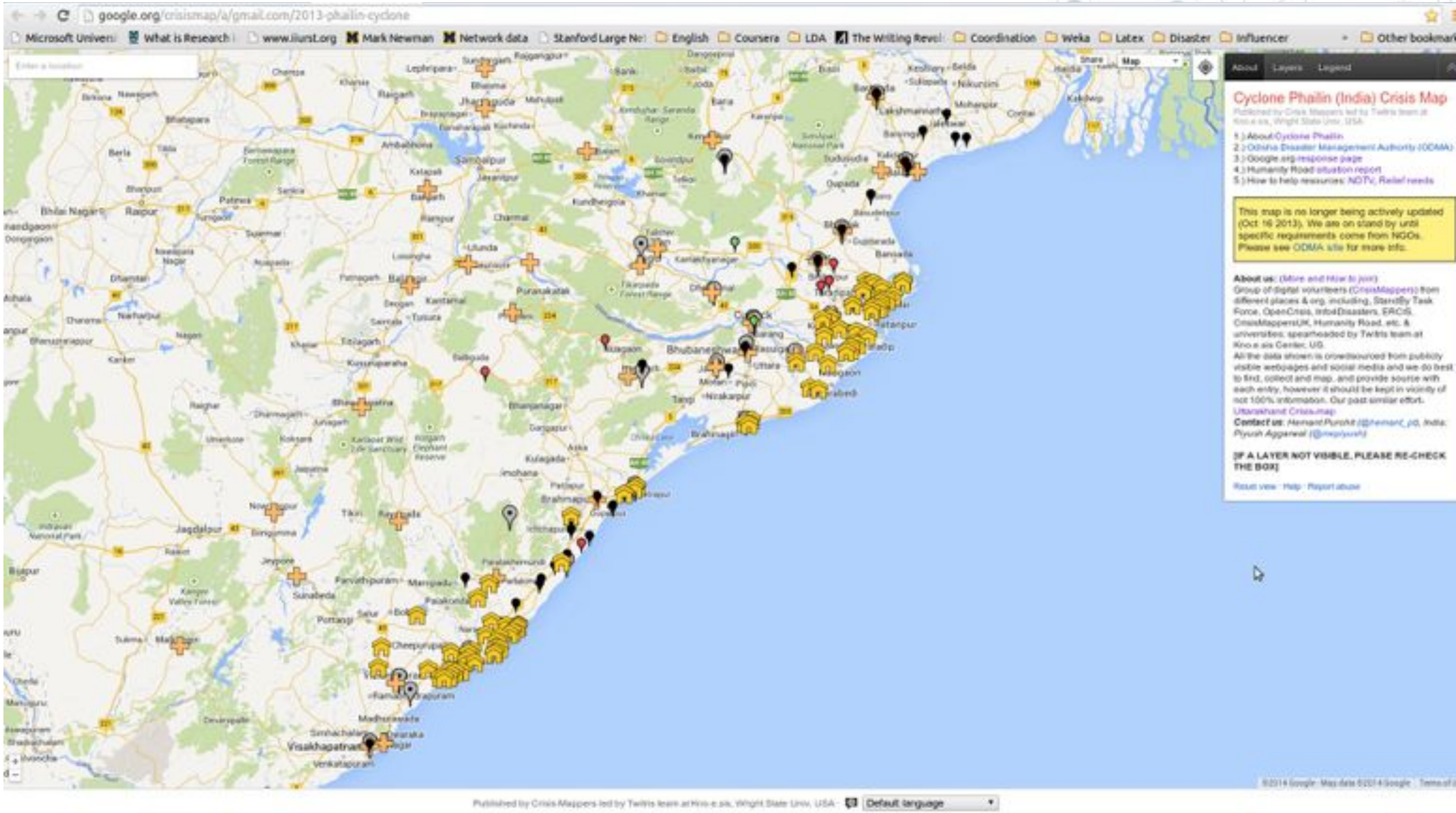
- Chennai floods: How social media and crowdsourcing helps people on ground, OneIndia, 12/2015
- Digital soldiers emerge heroes in Kashmir flood rescue, HindustanTimes, 09/2014
- Google's Person Finder and Google Crisis Response Map for Phailin to help with crisis information, DNA, Oct 12, 2013
- Using crisis mapping to aid Uttarakhand, The Hindu, Jun 27, 2013
- Twitris: Taking Crisis Mapping to the Next Level, Tech President, June 24, 2013
- Could Twitris+ Be Used for Disaster Response? iRevolution, September 11, 2012

Also tracked: Japan Earthquake, Haiti Earthquake, Pakistan Floods, Oklahoma Tornado, Hurricane Sandy, Uttarakhand Floods, Houston Floods,...and many more. And many other topics: Emoji, Religion, Gun Violence, Public Policy, Smart City, Health, Election (currently predicting: Election2012, Brexit, Election2016, ALSenate): http://knoesis.org/amit/media/

Sample of Real-World Impact & Media Coverage

- Sep 11, 2012 — [Could Twitris+ Be Used for Disaster Response?]
- Jun 24, 2013 — [Twitris: Taking Crisis Mapping to the Next Level]
- Jun 27, 2013 — [Using crisis mapping to aid Uttarakhand]
- Oct 12, 2013 — [Google's Person Finder and Google Crisis Response Map for Phailin to help with crisis information]
- Sep 9, 2014 — [Digital soldiers emerge heroes in Kashmir flood rescue]
- Dec 7, 2015 — [Chennai floods: How social media and crowdsourcing helps people on ground]

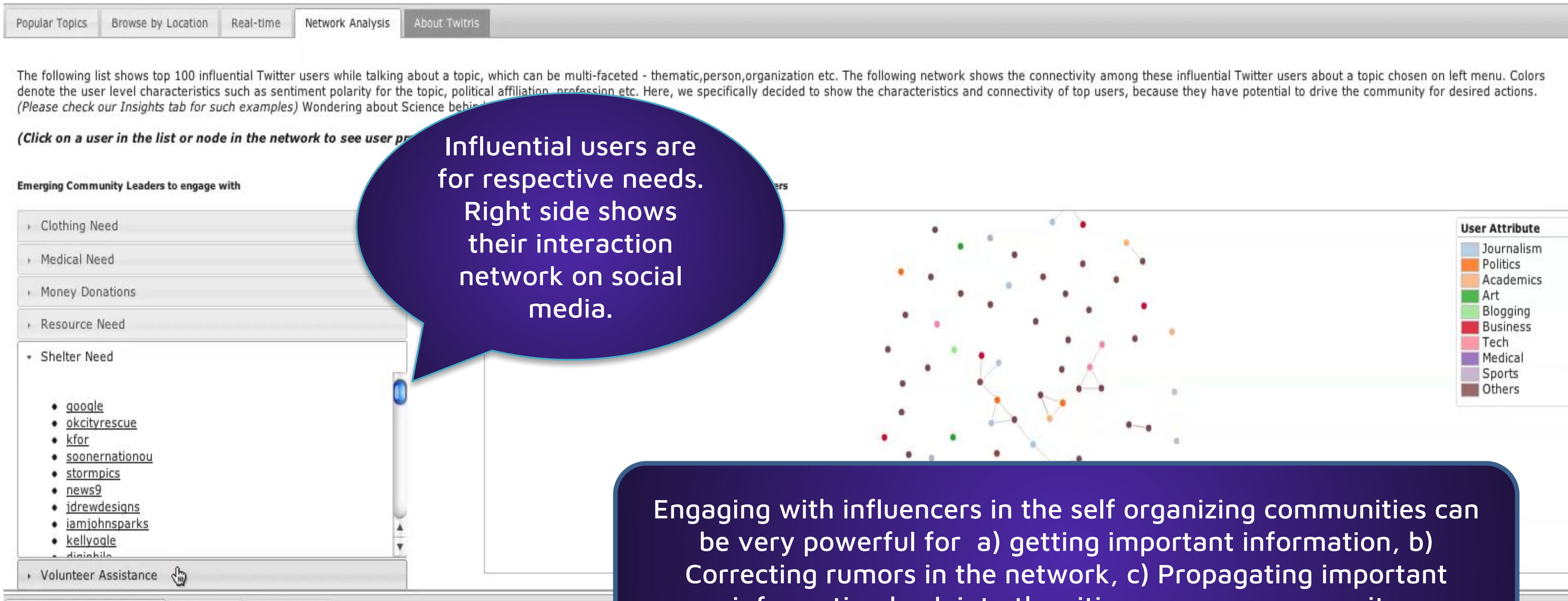
And many other topics: Emoji, Religion, Gun Violence, Public Policy, Smart City, Health, Election (currently predicting: Election2012, Brexit, Election2016, ALSenate): http://knoesis.org/amit/media/



Google Crisis Map for Hurricane Phalin, which used data from international participants spearheaded by Twitris team at the Kno.e.sis center.

Twitris: Real Time Information

Influencers to engage with, for specific needs



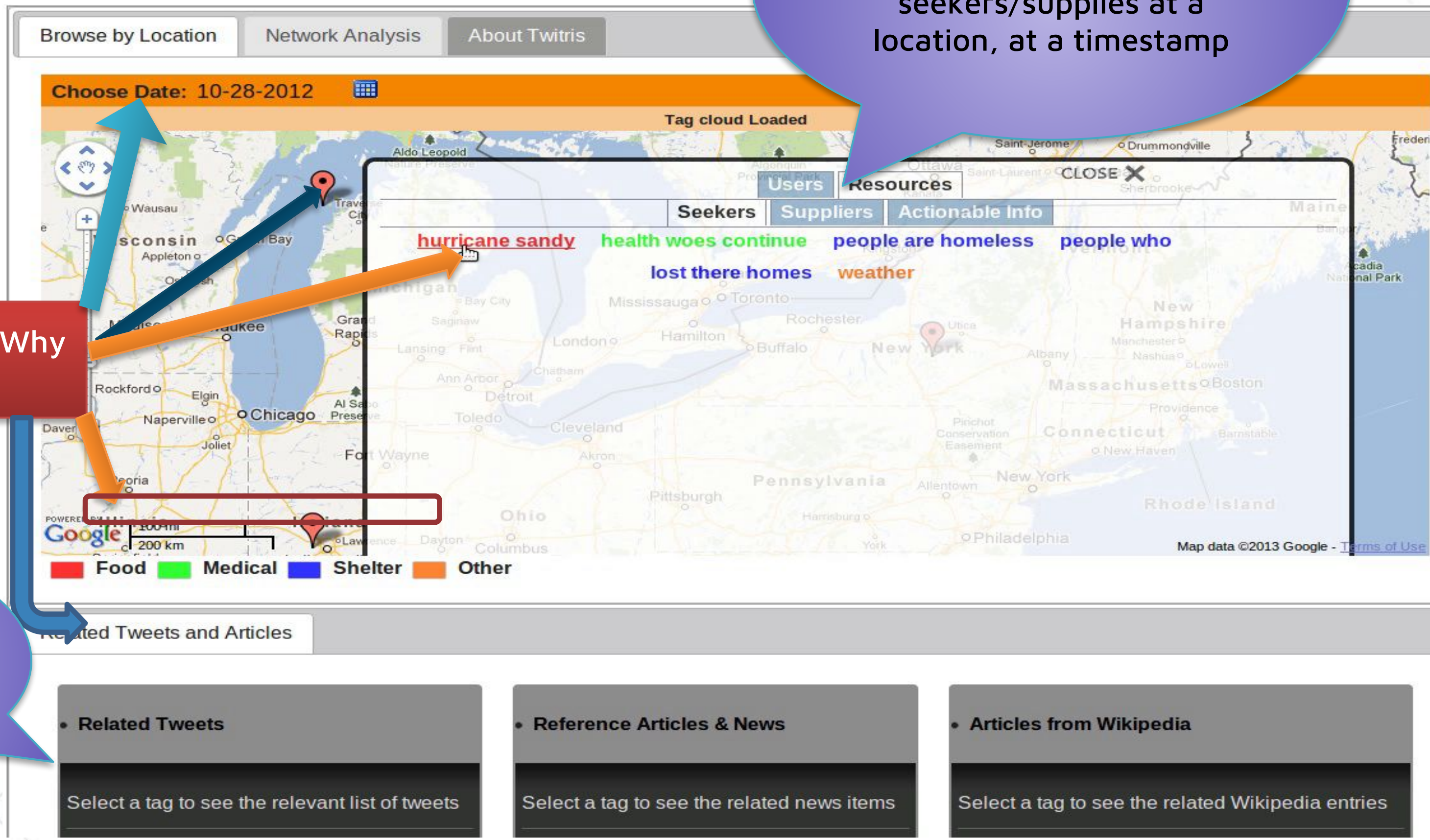
Influential users are for respective needs. Right side shows their interaction network on social media.

Engaging with influencers in the self organizing communities can be very powerful for a) getting important information, b) Correcting rumors in the network, c) Propagating important information back into the citizen sensors community

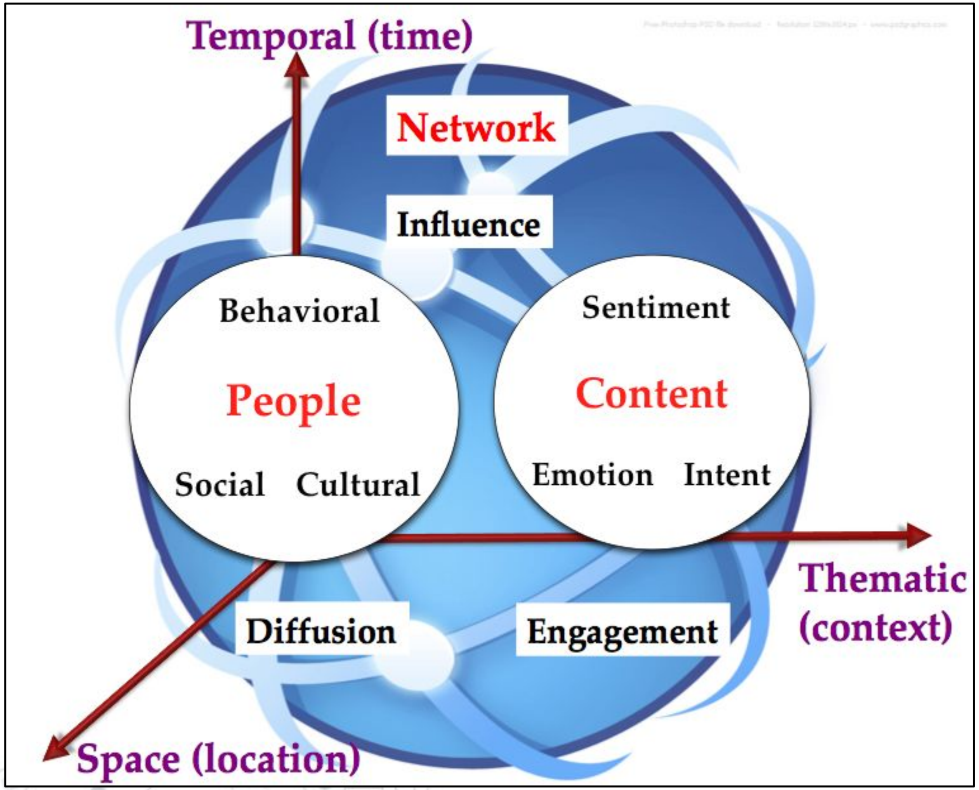
Influential users to engage with and resources for seekers/supplies at a location, at a timestamp

What-Where-How-Who-Why Coordination

Contextual Information for a chosen topical tags



TWITRIS' technical Approach to Understand & Analyze Social Content



Social Data is incredibly rich.

Real-time analysis of
v1: Spatio-Temporal-Thematic
v2: People-Content-Network
v3: Sentiment-Emotion-Intention
v4: Semantic filtering/knowledge graph, IFTTT, scalability, robustness

Commercial: Cognovi Labs

Twitris Technology:
Real-time, Actionable Insights from Social-media

Spatio-Temporal-Thematic

- Provides thematic context through analysis of place and time.

People-Content-Network

- Analyzes influential users and identifies who is being listened to.

Sentiment-Emotion-Intent

- Extracts and assigns structured sentiment and emotion scoring from unstructured content to understand motivation, feelings, opinion and intent.

Key Differentiators:

- **Comprehensive (above)**
- **Semantic Processing:** use of public and proprietary knowledge.
- **Real-time processing:** used in live blogging of election debate; coordination during disasters.
- **Scalable**: deployed on a large cloud (864 CPUs, 17 TB main, 435 TB disk).

Twitris Technology:
Real-time, Actionable Insights from Social-media

S-T-T | **P-C-N** | **S-E-I**

Spatio-Temporal-Thematic
Provides thematic context through analysis of place and time.

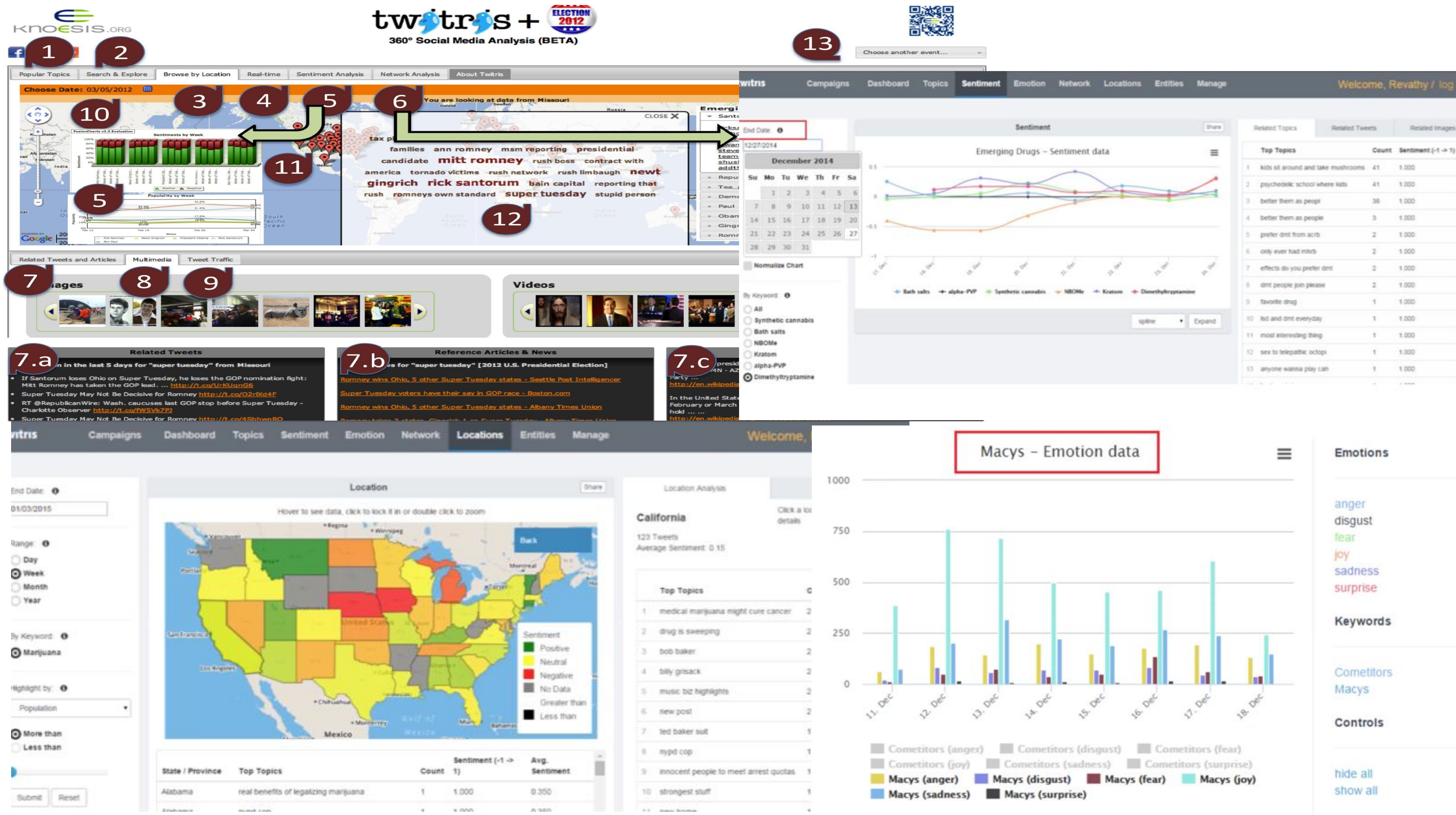
People-Content-Network
Analyzes influential users and identifies who is being listened to.

Sentiment-Emotion-Intent
Extracts and assigns structured sentiment and emotion scoring from unstructured content to understand motivation, feelings, opinion and intent.

Key Differentiators:

- **Comprehensive (above)**
- **Semantic Processing:** use of public and proprietary knowledge.
- **Real-time processing:** used in live blogging of election debate; coordination during disasters.
- **Scalable**: deployed on a large cloud (864 CPUs, 17 TB main, 435 TB disk).

Snapshot of Some Real-world Applications/Trials



Domains: Branding, Disaster Coordination, Social Movements, Election, Development, Epidemiology,..

Some of the significant human, social & economic development applications we work on at Kno.e.sis

- Coordination during disasters (QCRI, Microsoft Research NYC, CrisisNET, UN)
- Harassment on social media (WSU cognitive scientists)
- Prescription drug and opioid abuse, Cannabis & Synthetic Cannabinoid epidemiology (Center for Interventions, Treatment and Addictions Research, ….)
- Depressive disorders (Weill Cornell Med)
- Gender-based violence (UNFPA), Zika Spread
- and extensive applications in personalized digital health, public health (Dayton Children's Hospital, Wright St Physicians, …)

Highly multidisciplinary team efforts, often with significant partners, with real world data, intended to achieve real-world impact

Some of the topics on Online Social Media at Knoesis

- Named Entity Recognition, Implicit Entity
- Relationship Extraction (E.g., ADR)
- Language usage in Social Media
- Exploration of People, Content and Network dynamics
- Sentiment, Emotion, Intent extraction; Opinion mining
- Trust
- Integrated exploitation of Multimodal data (text, photo-satellite images), sensor/IoT-web-social data and knowledge (PCS applications)

All embodied in Twitris technology, commercialized as Cognovi Labs

Why People-Content-Network + Spatial-Temporal-Thematic metadata?

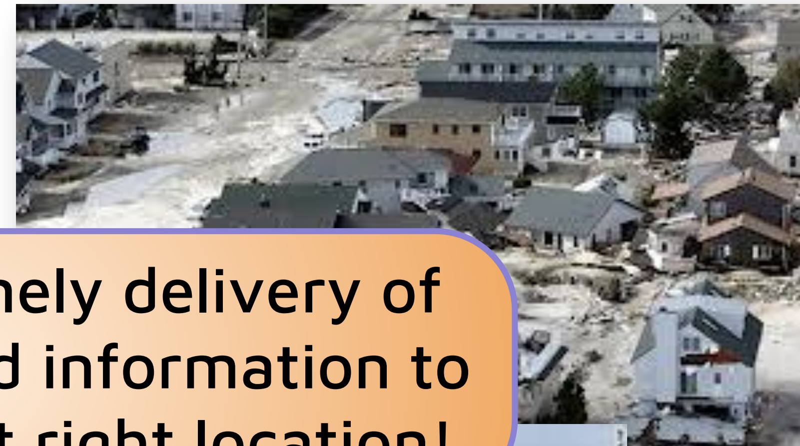
(Example of Understanding Crisis Data)

	Data generated at the disaster location	Data generated around the world
Who generates the data? (**People**)	Affected people, NGO volunteers (witness)	People not directly affected by the disaster (Passionate observers, domain experts)
What data is generated? (**Content**)	1. Reports about medical emergency, needs for resources, current situation 2. Complains about robbery, unavailability of resources, help etc	1. Opinion, concerns, sympathy 2. Sharing of related news, personal experience, desire for help. 3. Discussions about management (coordination), role of NGO and government and environmental, economical effects
How data is generated? (**Network**)	Primarily by sending SMS and Web reports to involved NGOs and government organizations	Primarily through tweets, Facebook messages, blogs
Why data is generated? (**Intention**)	Ask for help, Offer help, etc.	Sharing personal view-points on the disaster related incidents
When data is generated? (**Time**)	Primarily in recovery and rebuild phase	Primarily immediately after the disaster

Content Analysis: Typical Sub-tasks

- Recognize key entities mentioned in content
  - Information Extraction (entity recognition, anaphora resolution, entity classification..)
  - Discovery of Semantic Associations between entities
- Topic Classification, Aboutness of content
  - What is the content about?
- Intention Analysis
  - Why did they share this content?

Mining Actionable Information to Support Disaster Coordination



ACTIONABLE: Timely delivery of right resources and information to the right people at right location!

Coordination of Actionable Information Needs at Varied Levels

Personal Level: Chennai Floods


Shonali Krishnaswamy
17 hrs ·

Many, many thanks to Prof. Amit Sheth and his team for their Twitter Analytics work ...this crowdsourced information is the only source of info. at a difficult time like this. I have not been able to contact my parents and sister in law since yesterday..but his team was able to tell me those areas have no flooding...but of course like all of chennai no electricity, no mobile networks...A very difficult time. Thanks to all crowd sourced social media info.!

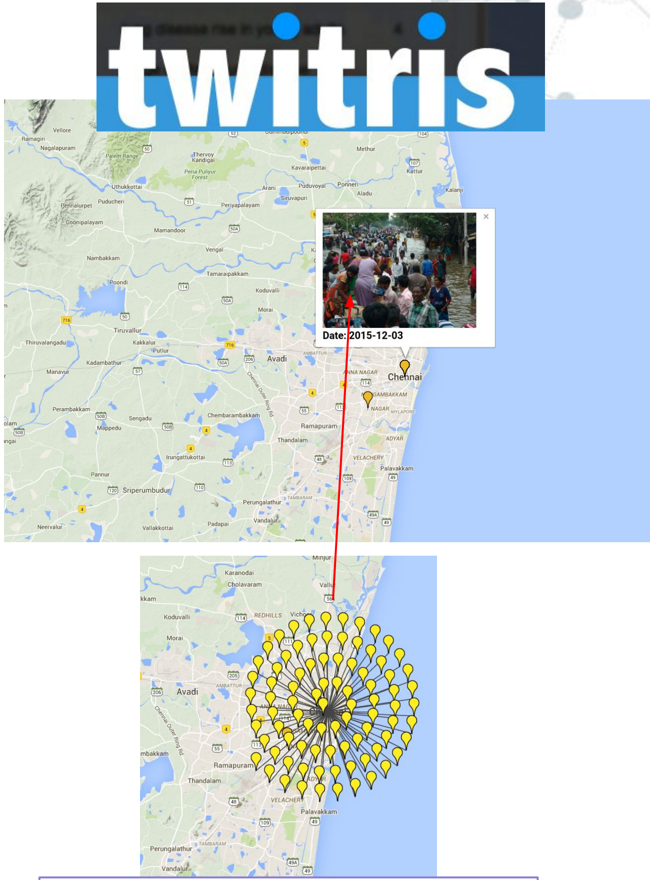
Like Comment Share Buffer

Parul Sheth, Pavan Kapanipathi, Rajaraman Kanagasabai and 20 others like this.

Srinath Sridharan That's the toughest part of this disaster (any disaster for that matter). No power, no network reception, cannot reach our family by any means. I just hope there are no further rains to ease the rescue and recovery operations.
Like · Reply · 1 · 16 hrs · Edited

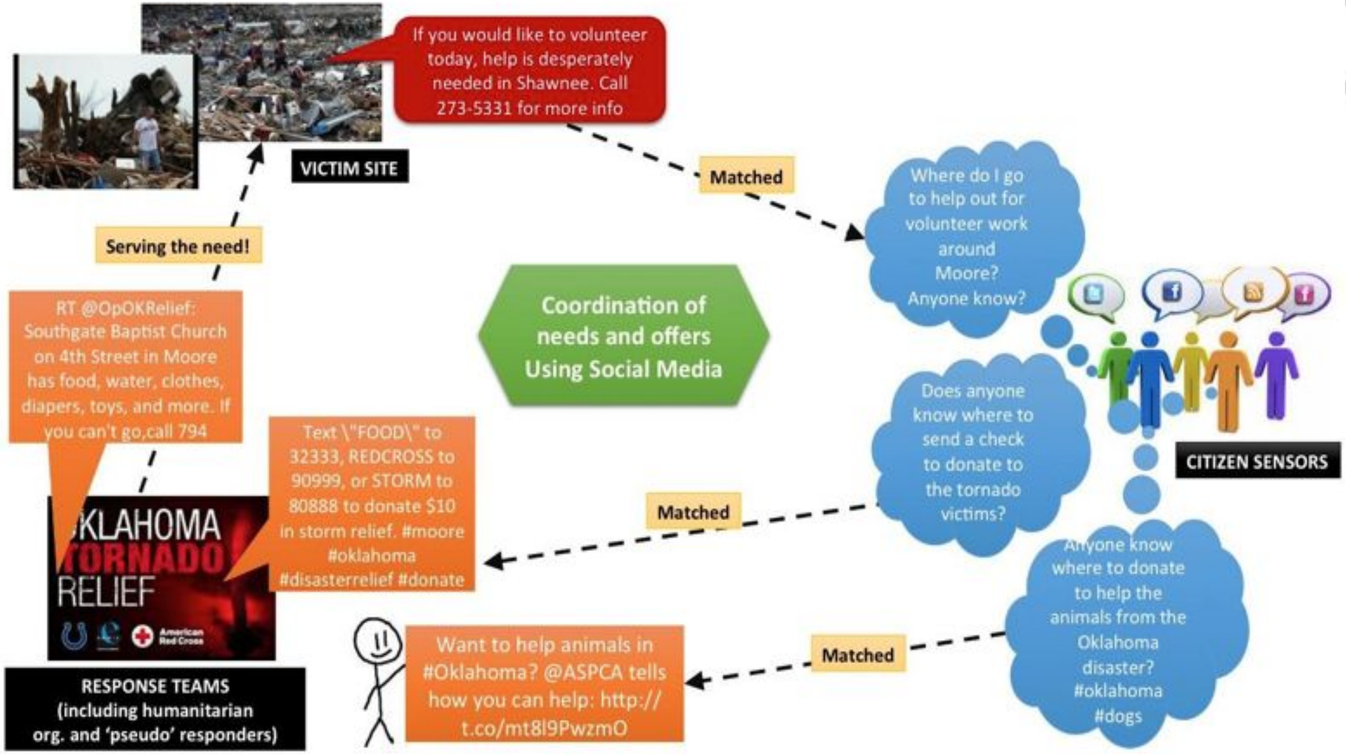
Amit Sheth Shonali Krishnaswamy: emailed you all tweets from the areas of your interest (Valmiki Nagar, Indira Nagar, Thiruvanmiyur or Seaward Road). Also sent you login/access to Twitris ChennaiFloodRelief campaign that you can use to do real-time (and past) monitoring of all tweets that mention terms (e.g., Valmiki) of your interest.
Like · Reply · 2 · 16 hrs

Shonali Krishnaswamy Thanks so much Amit!
Like · Reply · 16 hrs



Twitris Chennai Flood Map

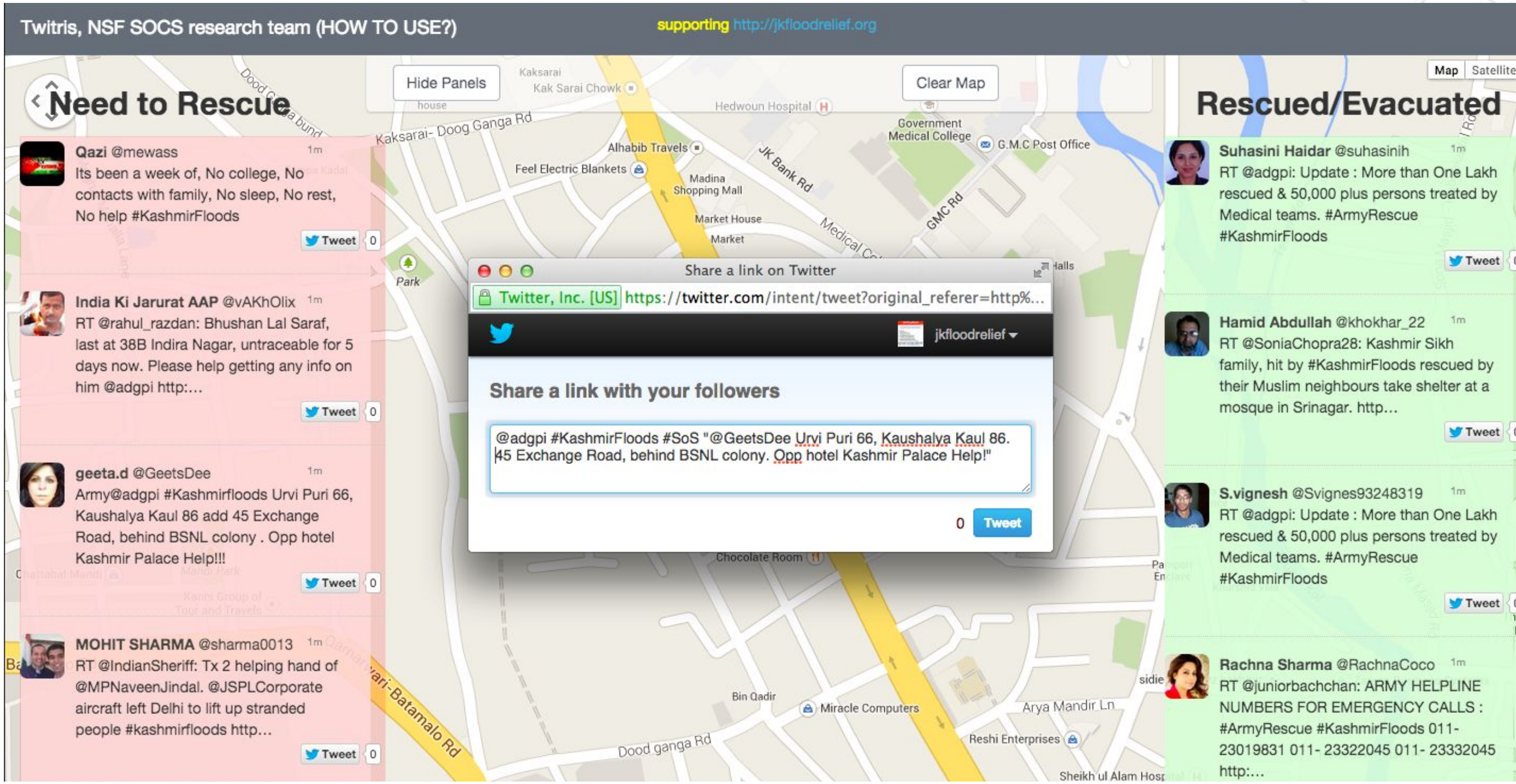
Community Level: Oklahoma Tornado



Example of coordination during #Oklahoma-tornado response based on automatically matching need-offer pairs of community members.

More: Purohit et al. [Emergency-relief coordination on social media: Automatically matching resource requests and offers](). First Monday, 2014.

Regional Level: Kashmir Floods



Rescue and Evacuation Stream Map during the historic Jammu & Kashmir Floods in Sep/2014. Twitris supported the scalable relief effort of JKFloodRelief.org initiative.

More: Purohit et al. Empowering Crisis Response-Led Citizen Communities: Lessons Learned from JKFloodRelief.org Initiative. IGI Global book, 2016.

Regional Level: Kashmir Floods

Moriam Nessa
Sep 8th, 11:09am

I do not know if anyone will be reading this message and if this will be of any help. But I have my sister who is stranded in Srinagar. She is 9 months pregnant and they need help. We have been trying to get through the help lines but nothing is working. Somebody please help ...

ADGPI - Indian Army
Sep 8th, 3:24pm

Jawahar Nagar is heavily flooded. Rescue teams will be going there. So dont worry they will be all right. Indian Army is there

Moriam Nessa
Sep 8th, 7:28pm

Hey, Thank you so much for your effort in rescue operations. But I'm writing to you again about update of the situation in Jawahar Nagar. My concern is that a young girl there is pregnant.

Moriam Nessa
Sep 9th, 8:27am

Thank you so much again for your relief work. My sister has been rescued. All thanks to you and your team.

Finding Actionable Information NUGGETS for Responders

For responders, most important information to manage coordination dependencies is to know: *WHO-WHAT-WHERE-WHEN*

- Scarcity of resources → Demand
- Availability of resources → Supply

Demand-Supply Identification: Oklahoma Tornado

Really sparse Signal to Noise:
2M tweets during the first 48 hrs. of #Oklahoma-tornado-2013
→ 1.3% as the precise resource donation requests to help
→ 0.02% as the precise resource donation offers to help

- *Text REDCROSS to 909-99 to donate to those impacted by the Moore tornado! http://t.co/oQMljkicPs* (REQUEST)
- *Please donate to Oklahoma disaster relief efforts.: http://t.co/crRvLAaHtk* (REQUEST)

- *Anyone know how to get involved to help the tornado victims in Oklahoma??#tornado #oklahomacity* (OFFER)
- *I want to donate to the Oklahoma cause shoes clothes even food if I can* (OFFER)

Blog by our colleague Patrick Meier on this analysis: http://irevolution.net/2013/05/29/analyzing-tweets-tornado/

Questions for social media tool to support Disaster Response Coordination

Which needs are of utmost importance?

Who are the people to engage with in the evolving ad-hoc social community?

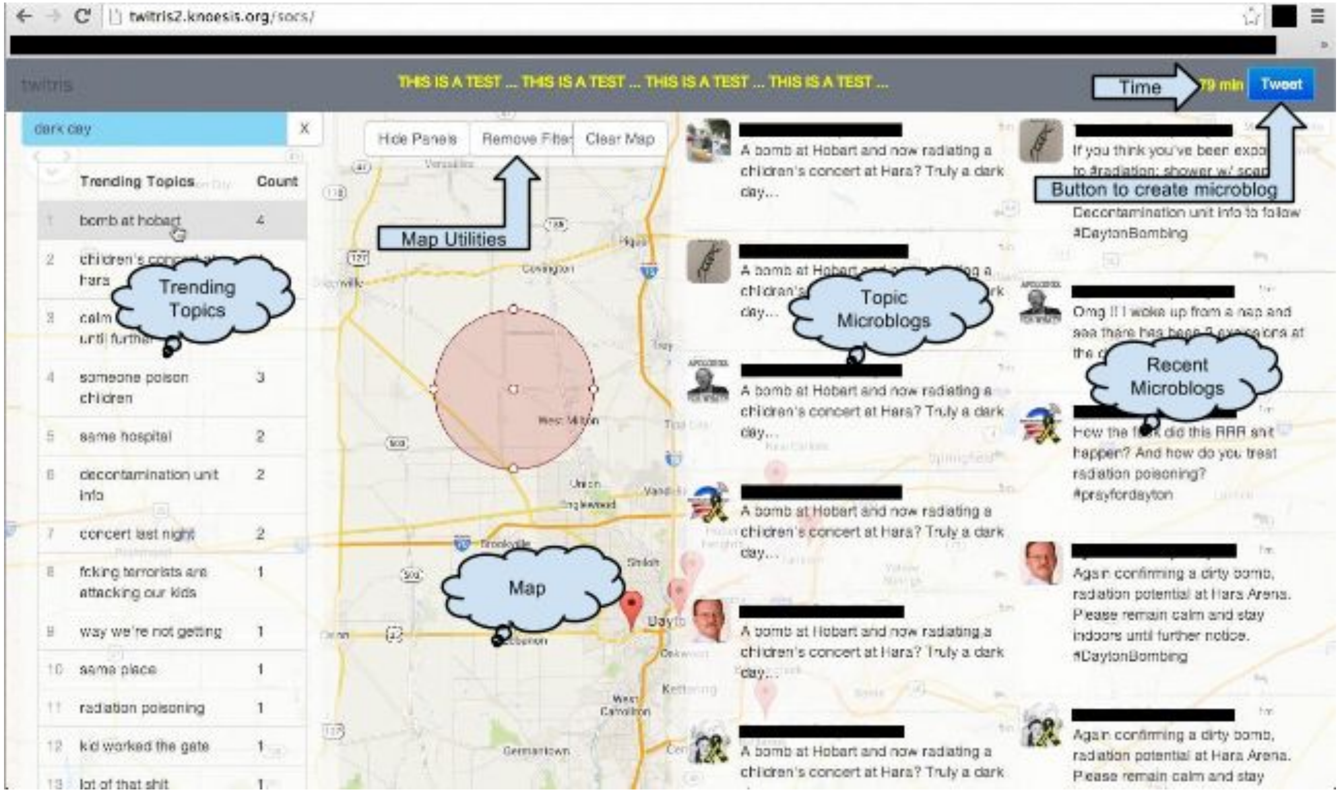
Who are the resource seekers and suppliers?

Where can I go for volunteering at my location?

How and Where can one donate?

Actionable information improves decision making process.

Designing Real-time Coordination Tools: Dayton Regional EM Exercise



<u>Snapshot</u>
Twitris-based simulation tool for filtering social system

Used in a functional exercise of emergency response organizations
City: Dayton
Date: 5/28/14

Data used for this simulation was based on repurposing of 2013 Boston Bombing dataset.

More: Hampton et al. [Constructing Synthetic Social Media Stimuli for an Emergency Preparedness Functional Exercise.](#) ISCRAM-2017.

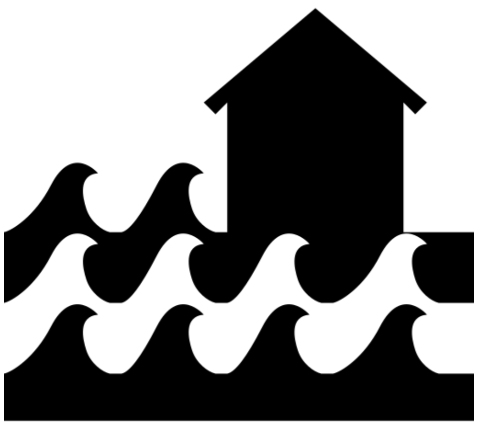
Designing Real-time Coordination Tools: Oklahoma Tornado

Incoming Tweets with need types to give quick idea of what is needed and where currently #OKC

Legends for

Clicking on a tag brings contextual information–relevant tweets, news/blogs, and Wikipedia articles

Chennai Flood 2015



Situational Awareness involving multimodal data (text+images from social media, also satellite images)

Real Example: use case of images

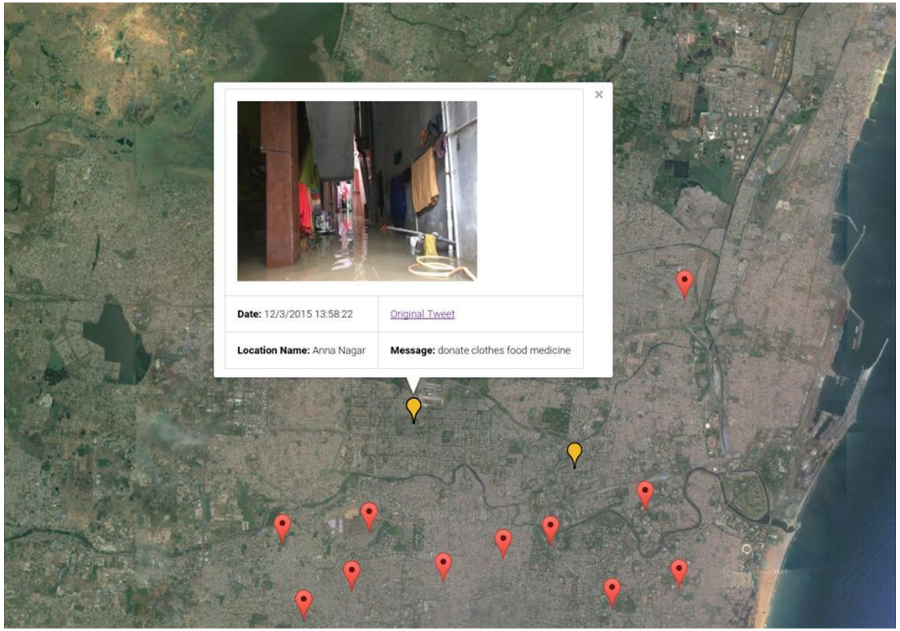
Social media is indispensable for supporting relatively simple but extremely important user needs during crisis. This is an illistration.

Visualization of Images

A drilled down view of a photo from a neighborhood of user's interest.

Powerful metadata extraction capabilities of Twitris can be combined

→ to give more information on the photo.



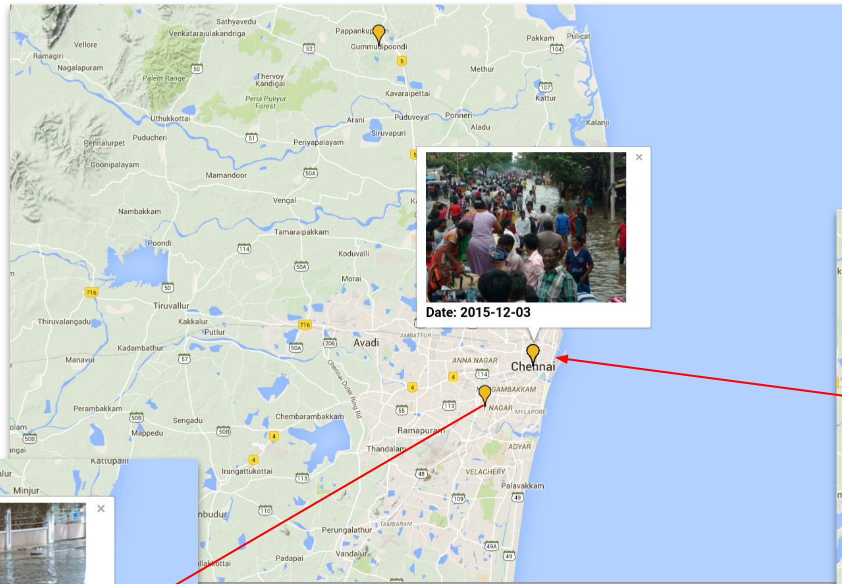
Some Statistics of Images

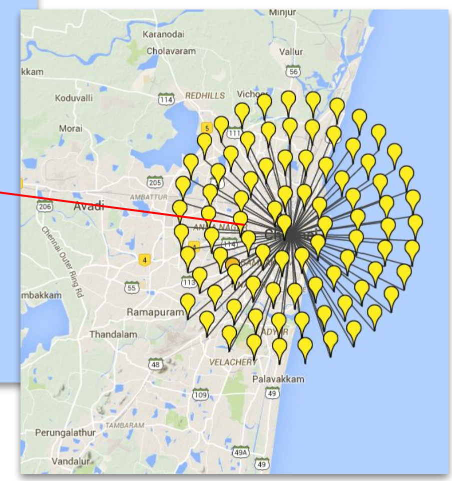
	Tweets	Images	Percentages
All	229,384	78,207	34%
India	105,200	35,609	34%
Tamil Nadu	34,387	10,622	31%
Chennai	21,260	6,195	29%

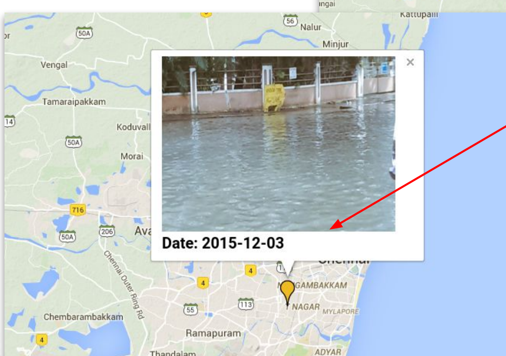
Geo-Tagged Tweets	55%*

* Tweets with longitude and latitude values of Twitter user-- this is very approximate as most of them is based on location (e.g., Chennai) provided in the profile of the poster, unless tweet itself has a geocode-- which would be a lot more precise (Twitter tweets have only 1% that are geotagged).

Plotting Images on the Map







* Twitris geotagged tweets using the user profile information when the geo-location information is missing mapping to the center of the city. That's why many tweets have the same location.

Examples of Images

Crowdsourcing Image Location

A	B	C	D	E	F	G	H	I
Tweet_Nbr	Tweet_URL	Lat	Long	Created_At (GMT)	Image_URL		Nearset Location by neighbourhood name (please click on arrow at the top right of the cell to get options for most likly and nearest neighbourhood to pick from)	More Details
79	https://twitter.com/srive	80.2203058	13.0416707	12/3/2015 16:32:46	https://pbs.twimg.com/ext_tw_video		Kilpauk	Kelly's Road
80	https://twitter.com/srive	80.2203058	13.0416707	12/3/2015 16:33:35	https://pbs.twimg.com/media/CVUIC		Kilpauk	Kelly's Road
81	https://twitter.com/GAU	80.2588243	13.1264673	12/3/2015 1:48:44	https://pbs.twimg.com/media/CVQ9		Kodungaiyur East	
82	https://twitter.com/A_fc	80.27072144	13.08267975	12/3/2015 6:19:58	https://pbs.twimg.com/media/CVR8i		Maduravoyal	

Crowdsourcing Location Features

A	B	C	D	E	F	G	H	I
Tweet_Nbr	Tweet_URL	Lat	Long	Created_At (GMT)	Image_URL		Nearset Location by neighbourhood name (please click on arrow at the top right of the cell to get options for most likly and nearest neighbourhood to pick from)	More Details
79	https://twitter.com/srive	80.2203058	13.0416707	12/3/2015 16:32:46	https://pbs.twimg.com/ext_tw_video		Kilpauk	Kelly's Road
80	https://twitter.com/srive	80.2203058	13.0416707	12/3/2015 16:33:35	https://pbs.twimg.com/media/CVUIC		Kilpauk	Kelly's Road
81	https://twitter.com/GAU	80.2588243	13.1264673	12/3/2015 1:48:44	https://pbs.twimg.com/media/CVQ9		Kodungaiyur East	
82	https://twitter.com/A_fc	80.27072144	13.08267975	12/3/2015 6:19:58	https://pbs.twimg.com/media/CVR8i		Maduravoyal	

- Two annotators from our team (one of them is a Chennai local)
- They relied on:
  - Their knowledge of the area.
  - The textual and metadata content about the images.
  - Their friends and other sources of help and information
  - Directly contact the authors on Twitter.

Follow

ChennaiRains:

Kushal Shah
@globaldesi__

@ Can you please provide the location where the image was captured so that we can put it on a CrisisMap

Answers:

· 4 Dec 2015
Replying to @globaldesi__
@kushalns5 Loyola College, Chennai !

· 4 Dec 2015
Replying to @globaldesi__
@kushalns5 Loc : Vaanagaram.Water has been receding .seems better nw.i heard there was power cable in water which has detached from line.

We don't have we cant communicate with them for 3 days now..

Kushal Shah @globaldesi__
Replying to
@soundarztweet @TimesNow @TwitterIndia can you pl provide the location where the image was captured so we can put it on a crisismap

And most of the time they don't answer or they answer saying there is no water anymore there.

Lessons Learned

1. Crowdsourcing is very **hard** due to human training, coordination, and time difference between team members.

2. **52%** of the images were localized by our annotators to a **neighborhood level.** (But humans can annotate only so many images in a given time-- we need annotations with local knowledge, it is time consuming and costly).

3. Our location extraction tool (LNEx) was able to localize **13%** of images (relying on textual content only) by geoparsing fine-grained location mentions (i.e., finding the geo-coordinates, not just geotagging)

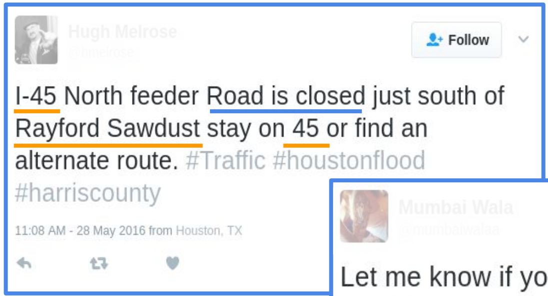
LNEx

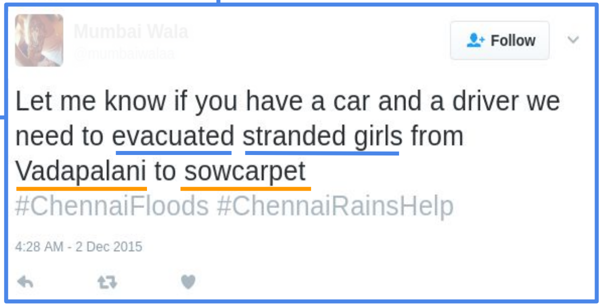
Location Name Extractor (LNEx)

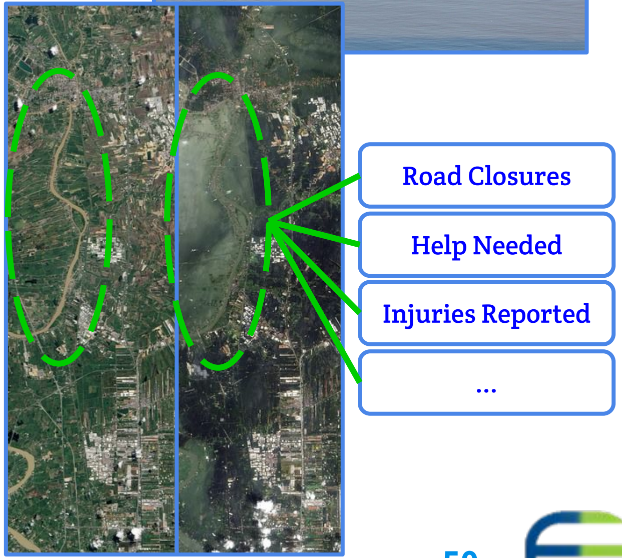
https://arxiv.org/abs/1708.03105

Automatic Extraction of Location Mentions

- Disaster Management (Response and Recovery)
  - Road Closures or Evacuations
  - Disaster Relief (shelters, food, and donations)
- Provide a system for Disaster Assistance Centers







Targeted Twitter Streams

Event-specific tweets. e.g.,

- ➢ Natural Disasters
- ➢ Political/Social Issue Demonstrations

Collected using hashtags

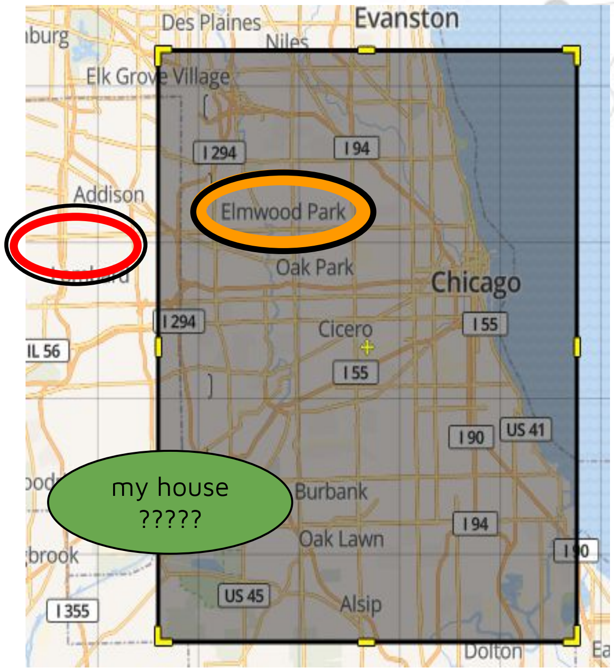
- ➢ #ChennaiFloods
- ➢ #ChennaiRains
- ➢ #houwx
- ➢ #houstonflood
- ➢ #LAWX
- ➢ #LaFlooding

or a bounding box

Wikipedia:File:2016_Louisiana_floods_map_of_parishes_declared_federal_disaster_areas.png

Location Names Categorization

- Previous works categorize location names based on their types (e.g., building, street)
  [Matsuda et al., 2015; Gelernter and Balaji, 2013]
- We categorize the location names based on their geo-coordinates and meaning into:
  - In-area Location Names (inLoc)
    - Location names that are inside the area of interest.
  - Out-area Location Names (outLoc)
    - Location names that are outside the area of interest.
  - Ambiguous Location Names (ambLoc)
    - Ambiguous in nature, need more context or background data for disambiguation



boundingbox.klokantech.com

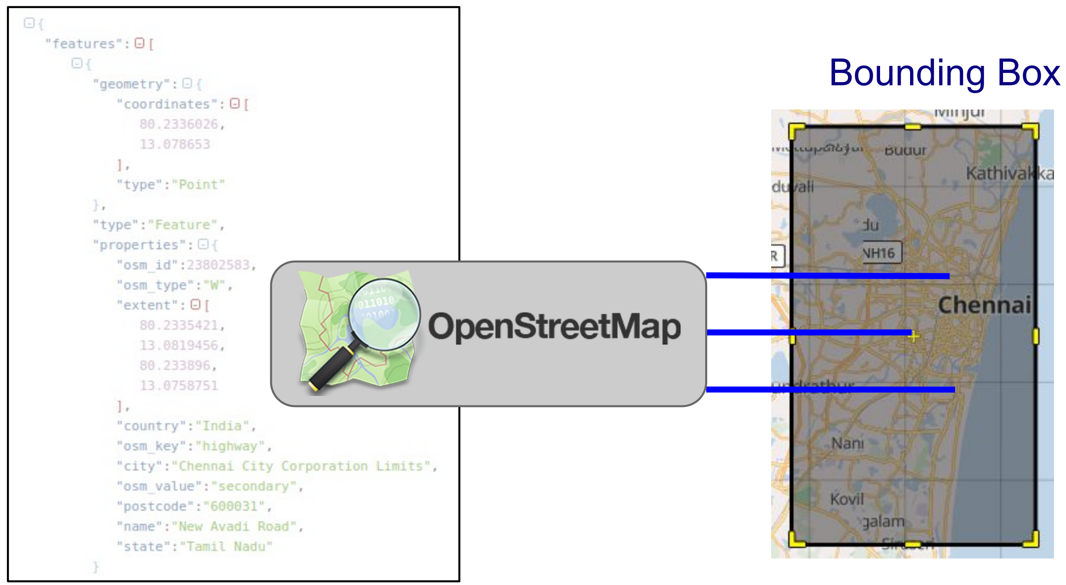
[#] Matsuda et al. 2015. Annotating geographical entities on microblog text. In The 9th Linguistic Annotation Workshop.
[#] Gelernter et al. 2013. Automatic gazetteer enrichment with user-geocoded data. SIGSPATIAL Workshop.

Challenges of Location Extraction

Contractions	Referring to “Balalok Matric Higher Secondary School” as “Balalok School”
Ambiguity	My backyard, My house, Buffalo
Abbreviations	Referring to “Wright State University” as “WSU”
Nicknames	HTown
Mentions in Hashtags	#LouisianaFlood
Misspellings	sou th kr koil street
Word Shape Problems	west mambalam
Ungrammatical Writing	Oxford school.west mambalam

Building Region-Specific Gazetteers

- → Cities
- → Countries
- → Street names
- → Neighborhoods
- → Points of interest
- → Building names
- → Organizations
- → Districts
- → States



```
{
  "features":[
    {
      "geometry":{
        "coordinates":[
          80.2336026,
          13.078653
        ],
        "type":"Point"
      },
      "type":"Feature",
      "properties":{
        "osm_id":23802583,
        "osm_type":"W",
        "extent":[
          80.2335421,
          13.0819456,
          80.233896,
          13.0758751
        ],
        "country":"India",
        "osm_key":"highway",
        "city":"Chennai City Corporation Limits",
        "osm_value":"secondary",
        "postcode":"600031",
        "name":"New Avadi Road",
        "state":"Tamil Nadu"
      }
```


boundingbox.klokantech.com

OpenStreetMap: Our Choice

- Regarded as the Wikipedia of maps.
- Contains more fine-grained locations than any other resource.
- More accurate geo-coordinates in comparison with Geonames [#]
- and, it has a strong volunteer foundation (such as hotosm.org) which maps thousands of locations during a disaster.

HOT Humanitarian OpenStreetMap Team

[#] Gelernter et al. 2013. Automatic gazetteer enrichment with user-geocoded data. SIGSPATIAL Workshop.

Gazetteers Preprocessing

- ★ Capturing different forms of a toponym (to improve recall)
  - ○ "Balalok Matric Higher Secondary School" → "Balalok School"
- ★ Recording alternative names (to improve recall)
  - ○ "Anna Salai (Mount Road)" → "Anna Salai" and "Mount Road"
- ★ Filtering toponym names (to improve recall)
  - ○ Break records: "Tamilnadu Housing Board Road , Ayapakkam"
- ★ Filtering out very noisy toponyms (to improve precision)

Jackson	Clinton	Ball
City in Mississippi	Town in Louisiana	Town in Louisiana

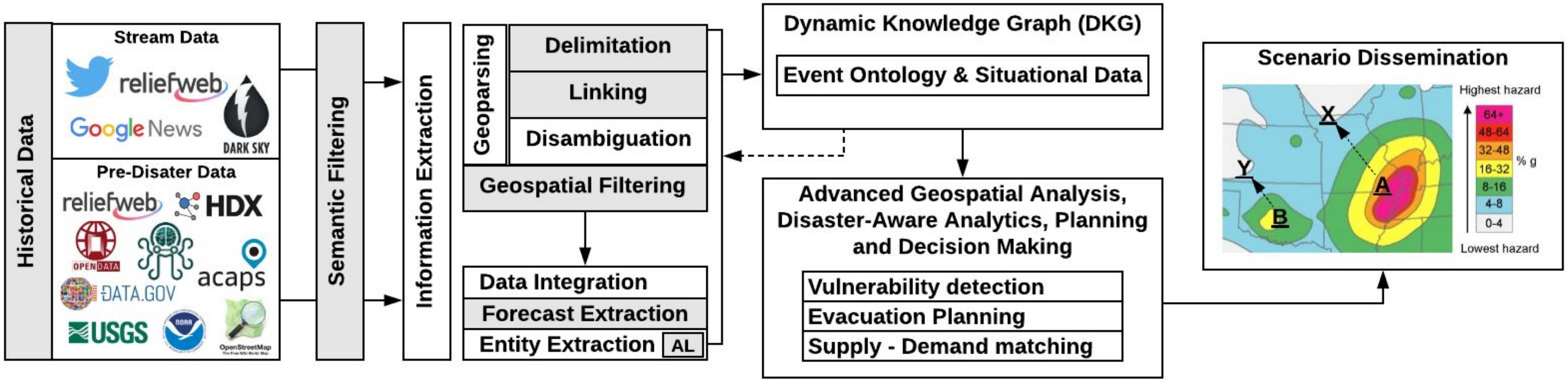
https://foursquare.com/

LNEx Location Extraction Results

LNEx

	Datasets									
	Chennai			Louisiana			Houston			AVG
	P	R	F	P	R	F	P	R	F	F
Google NLP	0.40	0.49	0.44	0.55	0.75	0.64	0.39	0.51	0.44	0.51
OpenCalais	0.43	0.10	0.17	0.81	0.77	0.78	0.62	0.35	0.45	0.47
DBpedia Spotlight	0.31	0.44	0.36	0.57	**0.88**	0.70	0.35	0.53	0.42	0.50
Yahoo! PLaceFinder	0.67	0.39	0.49	0.83	0.80	0.81	0.64	0.42	0.50	0.61
Stanford NER	0.72	0.29	0.41	0.78	0.42	0.55	0.74	0.32	0.45	0.47
OpenNLP	0.55	0.15	0.24	0.62	0.19	0.29	0.60	0.23	0.34	0.29
OSU TwitterNLP	0.74	0.40	0.52	**0.84**	0.69	0.76	0.66	0.39	0.49	0.59
TwitIE-Gate	0.51	0.36	0.43	0.66	0.84	0.74	0.35	0.39	0.37	0.52
Geolocator 3.0	0.43	0.54	0.48	0.32	0.71	0.44	0.38	0.58	0.46	0.46
Geoparsepy	0.41	0.28	0.33	0.45	0.72	0.55	0.44	0.46	0.45	0.45
LNEx-RawGaz	0.80	0.78	0.79	0.51	0.80	0.62	0.63	0.66	0.64	0.69
LNEx-AFGaz	**0.91**	**0.80**	**0.85**	0.83	0.81	**0.82**	**0.87**	**0.67**	**0.76**	**0.81**

Chain of Plausibility (CoP) Pipeline



Partners

DEEP

MILE 2

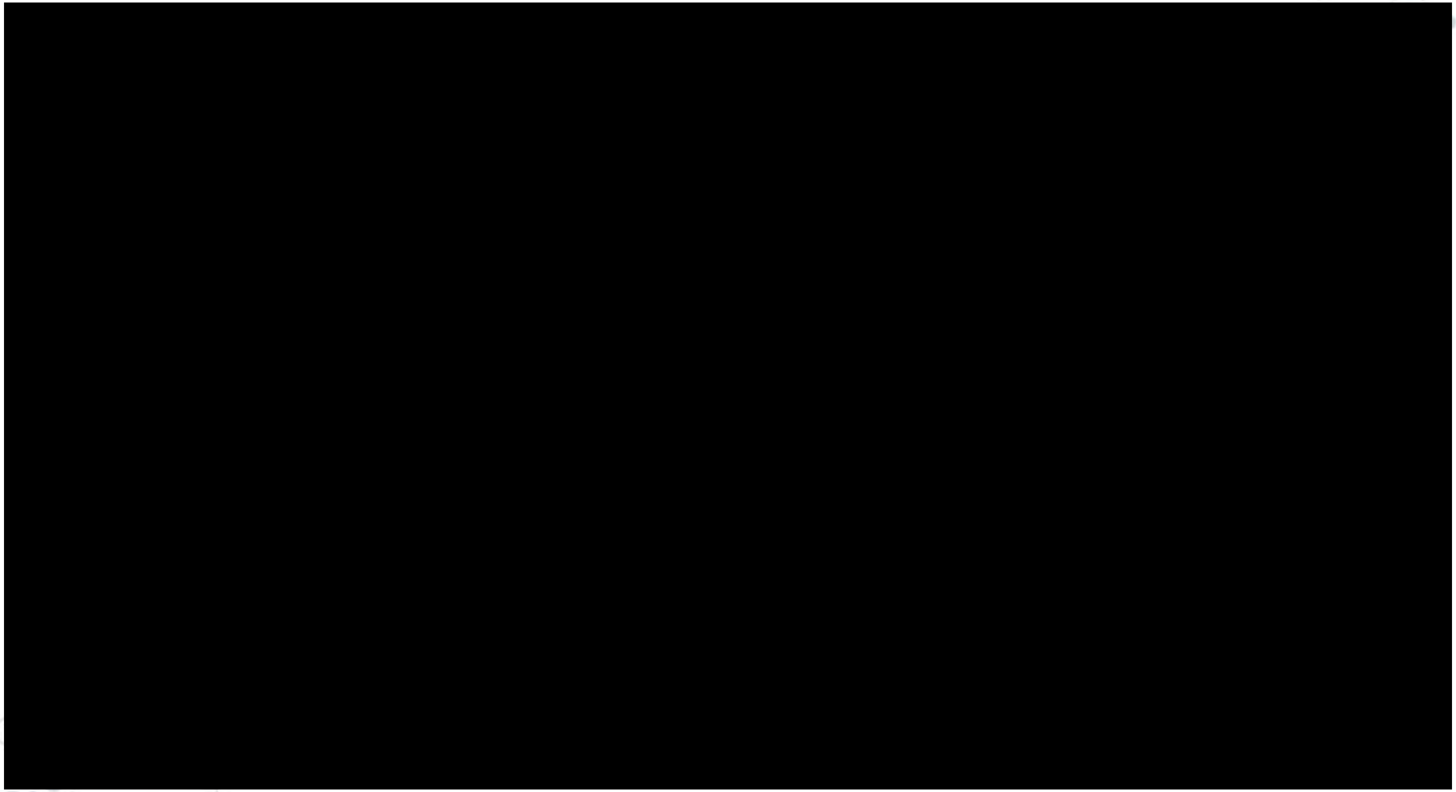
VIRGINIA DEPARTMENT OF EMERGENCY MANAGEMENT

UNITED STATES OF AMERICA
THE DAYTON THINK TANK
OHIO

Sopact

IBM Science for Social Good

Our Chain of Possibility Disaster Response Tool (scenario of supply demand match during flood crisis)

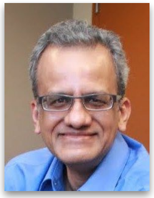


Example Smartphone App for Rescue/Response

- Medical Help
- Food/Shelter
  - Type of help needed
  - Number of people needing help


WRIGHT STATE
UNIVERSITY
KNO.E.SIS
Russ
Engineering

Collaborators



Dr. Krishnaprasad Thirunarayan
(Faculty - WSU)

Valerie Shalin
(Faculty - WSU)

Hemant Purohit
(Faculty - GMU)

Srinivasan Parthasarathy
(Faculty -OSU)

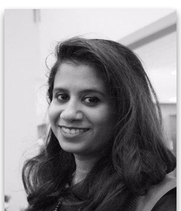
Special Thanks

Hussein Al-Olimat
(Graduate Student)

SoonJye Kho
(Graduate Student)

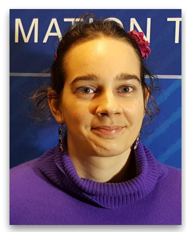
Hazards SEES Current Team Members

Hussein S. Al-Olimat
(Graduate Student)



Shruti Kar
(Graduate Student)

Manas Gaur
(Graduate Student)



Amelie Gyrard
(Post-Doctoral Researcher)

Hazards SEES Past Members:

Sarasi Lalithsena, Pavan Kapanipathi, Siva Kumar, Saeedeh Shekarpour, and Tanvi Banerjee

HazardsSEES

Hazards SEES project is funded by NSF Award#: EAR 1520870

Acknowledgements

Significant components of this talk is from the tutorials given by my team at

- WWW2011: "Citizen Sensor Data Mining, Social Media Analytics and Development Centric Web Applications," with Meena Nagarajan and Selvam Velmurugan.
- ICWSM2013: "Crisis Mapping, Citizen Sensing and Social Media Analytics: Leveraging Citizen Roles for Crisis Response," with Carlos Castillo, Patrick Meier, and Hemant Purohit.
- ISCRAM 2018: "Location Extraction and Georeferencing in Social Media: Challenges, Techniques, and Applications" with Hussein Al-Olimat, Amir Yazdavar, and T.K. Prasad.

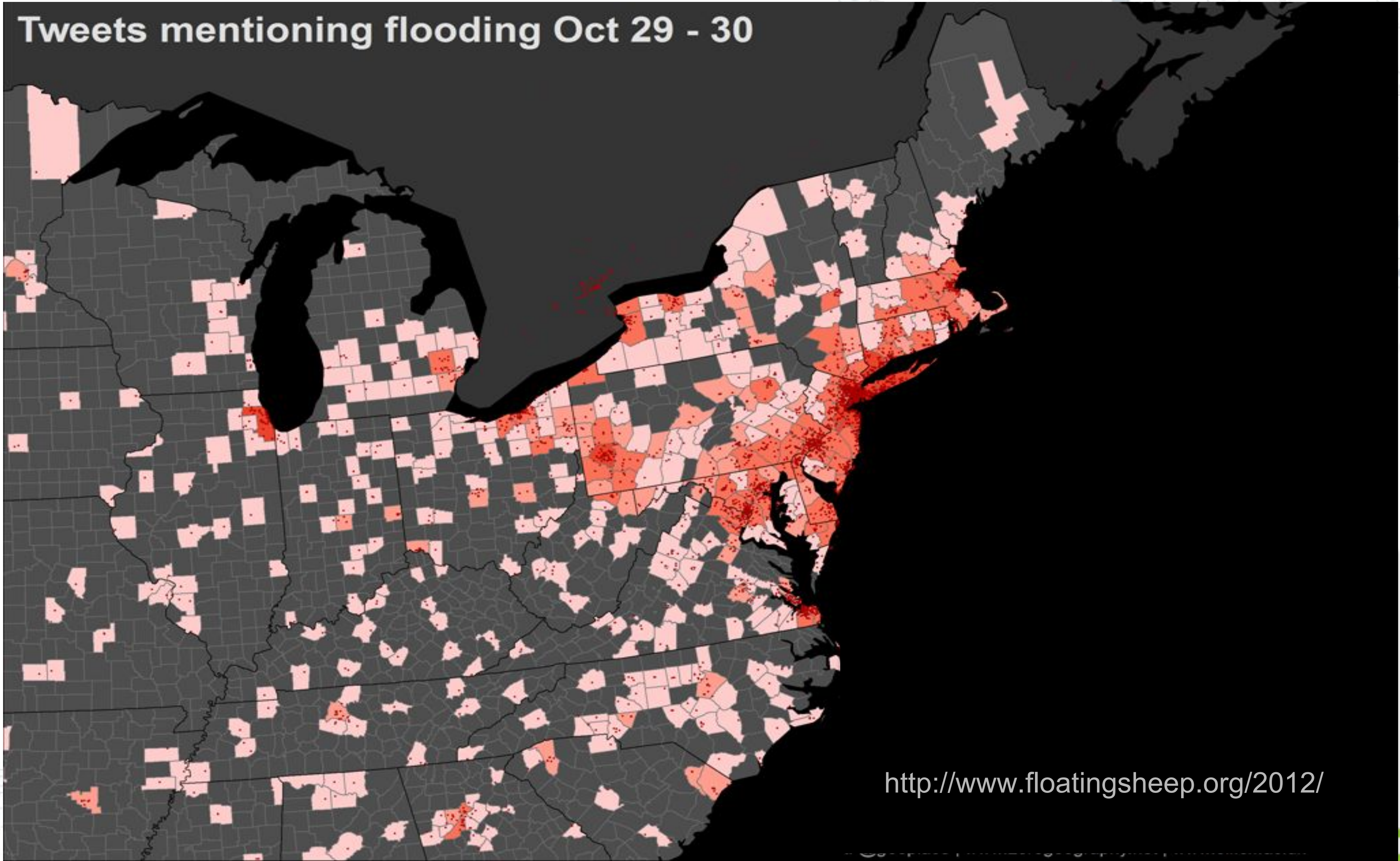
Contributors to Twitris and/or Semantic Social Web Research @ Kno.e.sis: L. Chen, H. Purohit, W. Wang with: P. Anantharam, A. Jadhav, P. Kapanipathi, Dr. T.K. Prasad, and alumni: K. Gomadam, M. Nagarajan, A. Ranabahu

Funding: NSF, AFRL, NIH; Collaborations: IBM, Microsoft

Thank You!

Acknowledgements: Supported in part by: NSF Award#: EAR 1520870: "Hazards SEES: Social and Physical Sensing Enabled Decision Support for Disaster Management and Response".

Backup


Tweets mentioning flooding Oct 29 - 30
http://www.floatingsheep.org/2012/