Social Media Incident Streams

A Text Retrieval Conference (TREC) Challenge

Ian Soboroff, ITL, NIST

Any mention of companies or services in this talk is for information only;
it does not imply recommendation or endorsement by NIST.

#PSCR2020

NIST

PSCR

DISCLAIMER

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately.

Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

*** Please note, unless mentioned in reference to a NIST Publication, all information and data presented is preliminary/in-progress and subject to change**

Jason R. Hunter Retweeted

Rob Davis
@RobDavisWx

RT @RobDavis_Wx: LANDSLIDE!... road blocked in Costa Rica after M7.6 earthquake twitpic.com/are653 (courtesy: @skasoul) #temblorcr

11:53 AM · Sep 5, 2012 · Twitter Web Client

John Hofileña
@johnhofilena

ALERT LEVEL 4, Marikina River at 18meters. Marikina residents near the river, you guys need to evacuate NOW. #maringPH #floodPH

3:28 AM · Aug 20, 2013 · Twitter Web Client

JR Santos
@jr0403

Replying to @DickGordonDG

@ChairmanGordon @philredcross sir until now no rescue at the place i tweeted earlier at mercedes homes3 biñan laguna. Thank you! #RescuePH

2:48 AM · Aug 19, 2013 · Twitter for iPhone

Philippine Red Cross @philredcross · Aug 19, 2013
Replying to @jr0403
@jr0403 @ChairmanGordon may WASAR teams na po kami nakadeploy

1

JR Santos @jr0403 · Aug 19, 2013
@philredcross @ChairmanGordon thank you po. Nagmesaage kasi sila ulit kanina.

linn Retweeted

Christian Garcia
@dude_funk

RT @dude_funk: #rescuePH Dela Paz, Biñan(?) Laguna. Newly kidney operated Tita currently on rooftop with other family due to flood. Please ...

10:52 PM · Aug 18, 2013 · Twitter for iPhone

Mike Allen
@MikeAllen_YMM

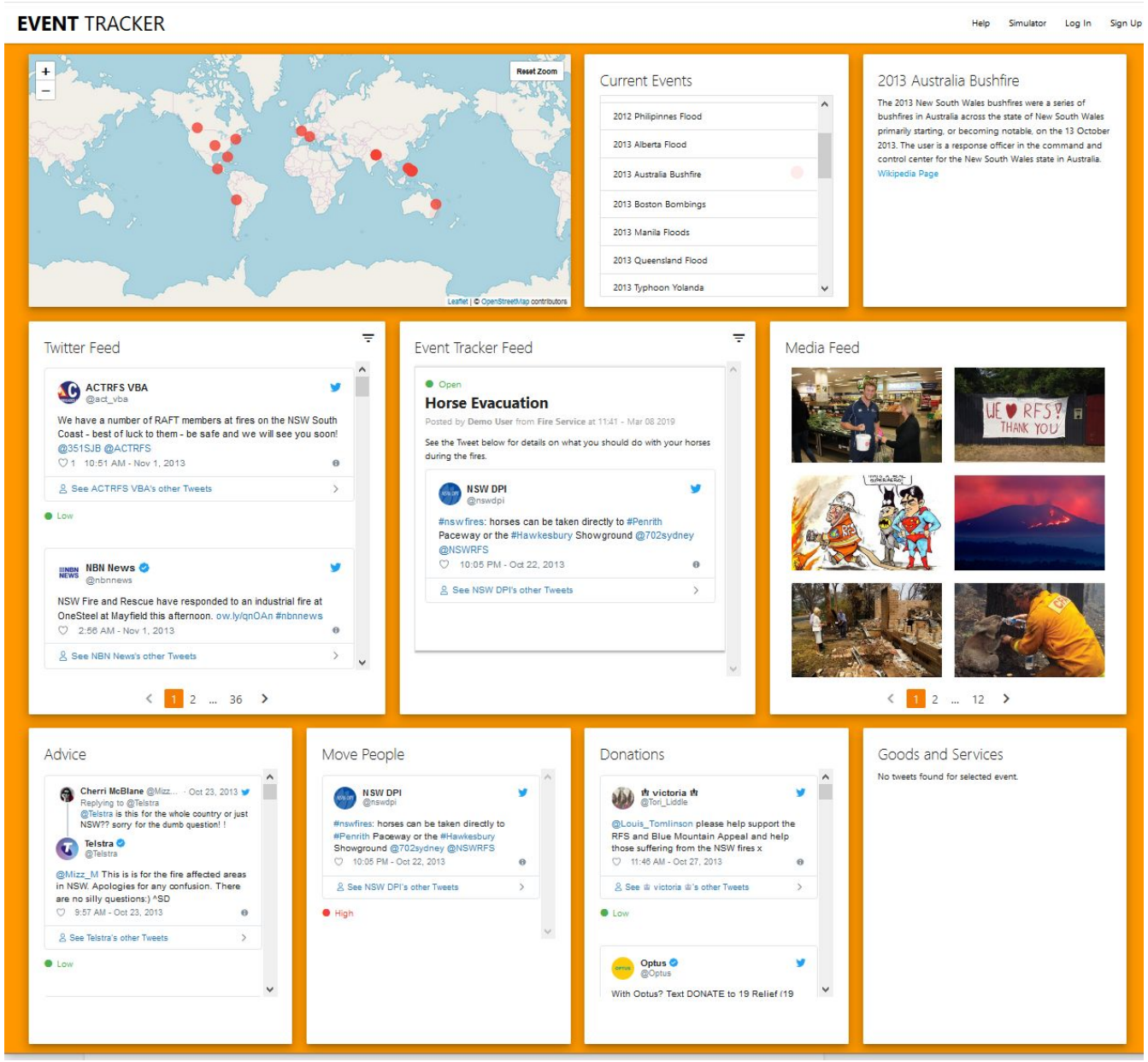
#ymmfire has breached the hill and coming down towards #hwy63 and Grayling Terrace #ymm #

5:49 PM · May 3, 2016 · Twitter for iPhone

(actual tweets from TREC-IS incident collections)

Problem

- People take to Twitter during crises, but no one can monitor it.
  - Hashtags and keywords have high volume, spam, irrelevant information.
  - Flood of hopes and prayers.
- Can computer systems find critical, actionable tweets in this mess?
- Can they get them to the right people in time?



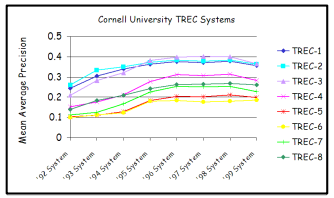
- AI systems can be built to
  - filter out noise,
  - identify critical tweets,
  - prioritize them,
  - and route them to the right people
- But AI depends on high quality training data.

Text Retrieval Conference

trec.nist.gov

- TREC is an evaluation workshop series started by NIST in 1992.
- TREC features a set of *tracks* that pose data challenges around different problems in search, information filtering, and information access.
- Each track creates a dataset for the open participant community: universities and industry research labs who sign up to attempt the challenge.
- The community participation process is leveraged to create ground truth and simultaneously measure the effectiveness of participant solutions to the track challenge.

Improves the state of the art



The TREC data revitalized research on information retrieval. Having a standard, widely available, and carefully constructed set of data laid the groundwork for further innovation in the field. The yearly TREC conference fostered collaboration, innovation, and a measured dose of competition (and bragging rights) that led to better information retrieval.

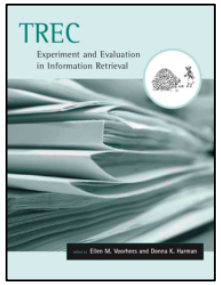
Hal Varian
Google Chief Economist
March 4, 2008

Forms/solidifies a research community

This project [the TREC Legal track] can be expected to identify both cost effective and reliable search and information retrieval methodologies and best practice recommendations, which, if adhered to, certainly would support an argument that the party employing them performed a reasonable ESI search, whether for privilege review or other purposes.

Magistrate Judge Paul Grimm
Victor Stanley v. Creative Pipe

Establishes the research methodology



TREC is an annual benchmarking exercise that has become a de facto standard in Information Retrieval evaluation.

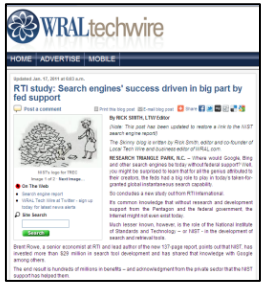
Stephen Robertson
Microsoft
SIGIR 2007

Facilitates technology transfer

TREC has proven to be a valuable forum in which IBM Research has contributed to an improved understanding of search, while at the same time the insights obtained by participating in TREC have helped to improve IBM's products and services.

Alan Marwick, et al.
IBM chapter of the TREC book
2005

Amortizes the costs of infrastructure



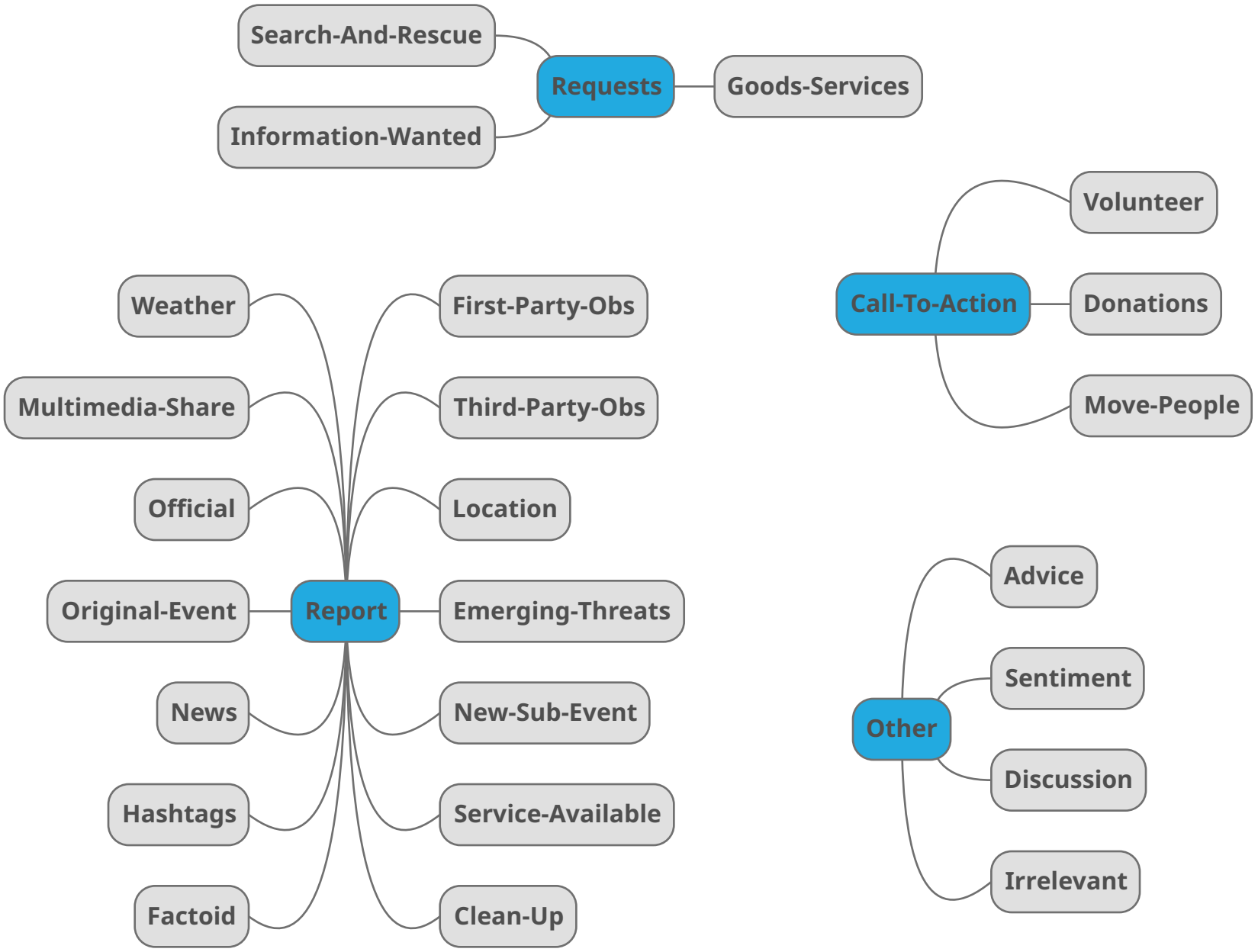
In other words, for every $1 NIST and its partners invested in TREC, at least $3.35 to $5.07 in benefits accrued to IR researchers…These responses suggest that the benefits of TREC to both private and academic organizations go well beyond those quantified by this study's economic benefits.

RTI International
Economic Impact Assessment of NIST's TREC Program
December 2010

NIST National Institute of Standards and Technology

TREC Incident Streams track

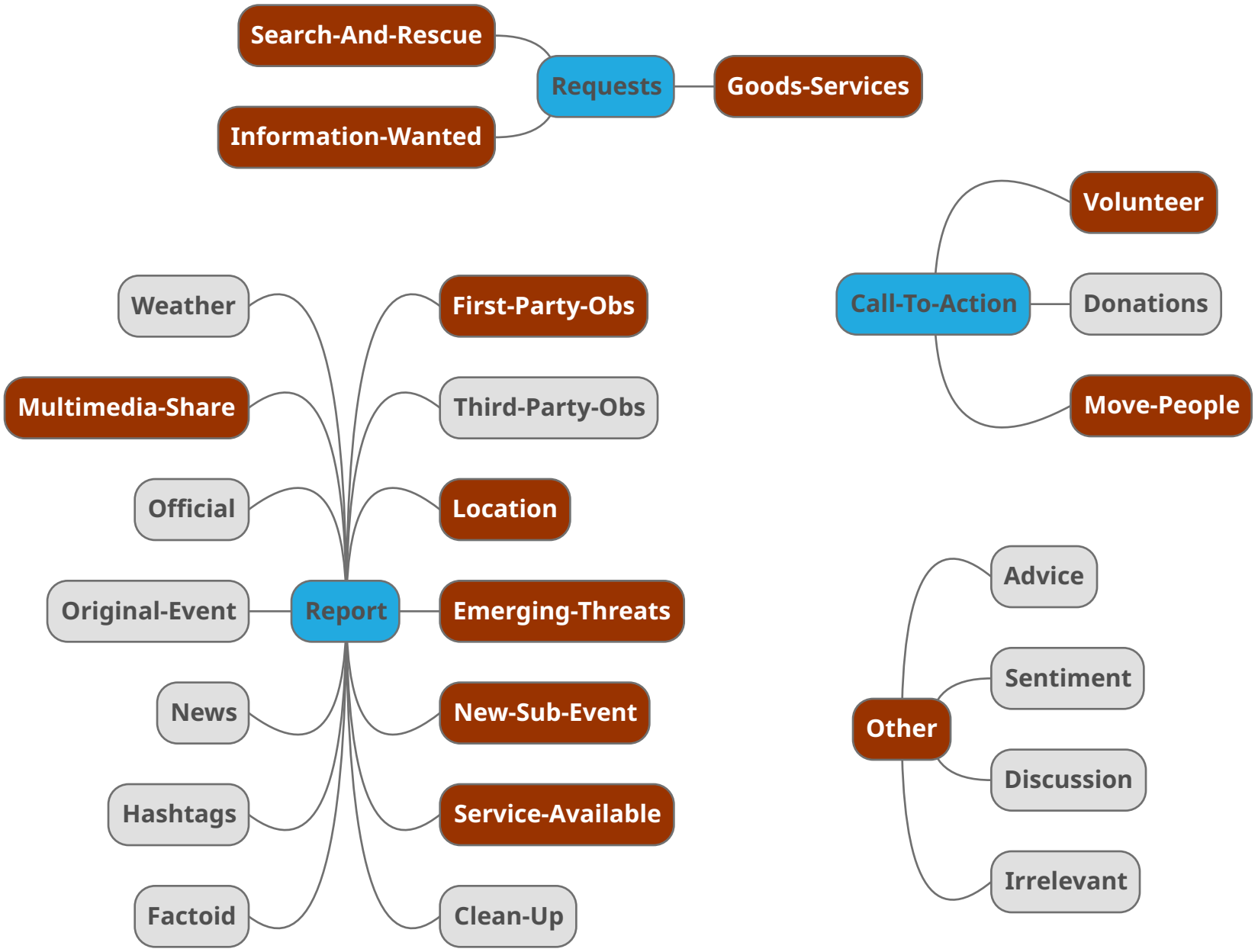
- Started in 2018.
- Provides 33 Twitter datasets from earthquake, wildfire, hurricane, flood, bomb and shooting events.
- Each tweet is labeled to indicate:
  - **Relevance**: Does it contain actionable information?
  - **Categories**: What kind of information does it contain?
  - **Criticality**: How important is it that public safety should see this tweet?



25 categories in 4 major groups:

- Immediate needs
- or, Useful as metadata
- or, Useful post-event
- or, Useful for research

Derived from existing taxonomies as well as research surveys of social media use during emergencies.



A variant of the task focuses on a subset of 12 categories.

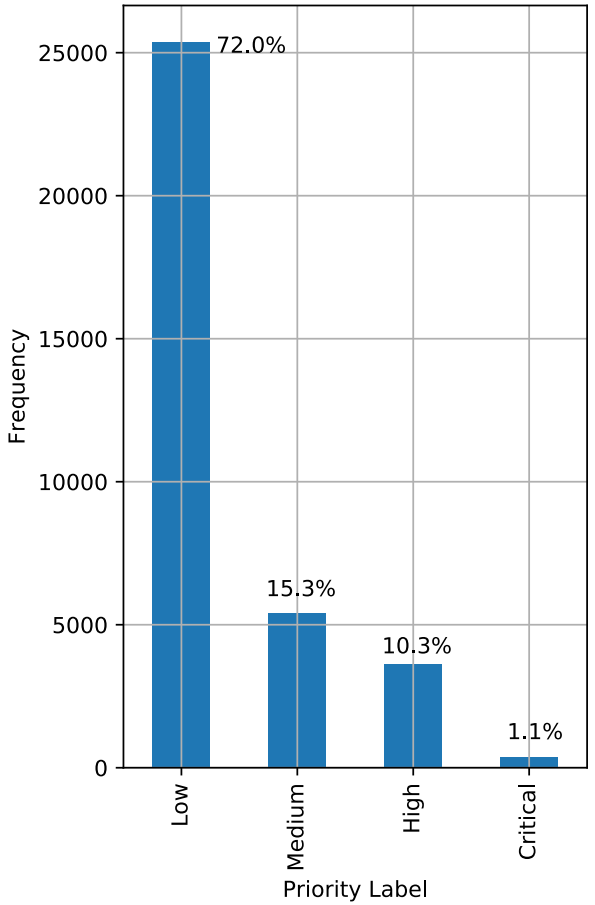
For this task, all the “Other” categories are collapsed.

These categories are those most likely to be the most useful to public safety personnel during an emergency situation.

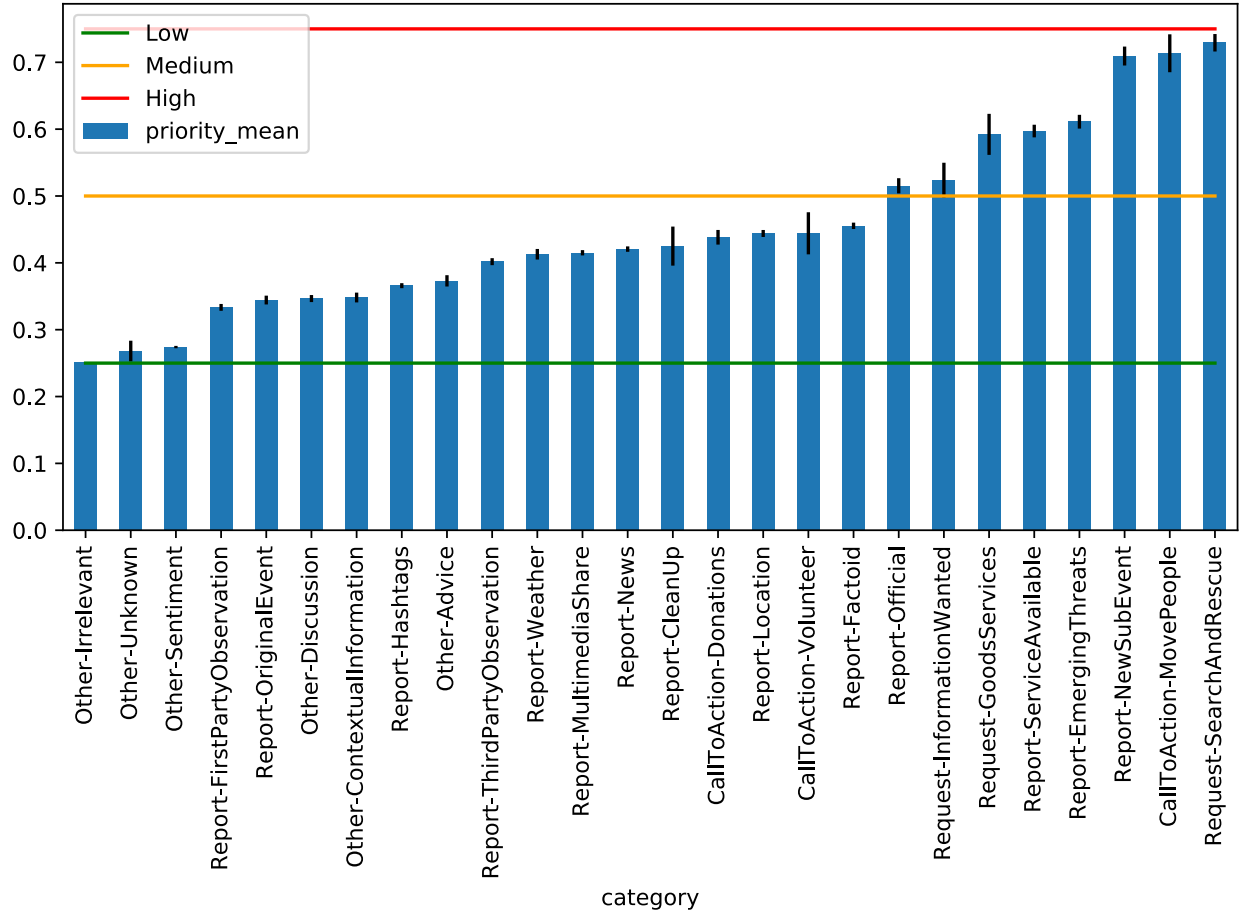
- After assigning all pertinent categories, the assessor indicates the tweets **criticality**.
- Does this tweet need immediate attention from emergency personnel, or can it wait?

Critical *(Notify immediately)*
High *(Should be viewed by officer)*
Medium *(Can be viewed later)*
Low *(Can be safely ignored)*

Criticality



(a) Criticality Distribution

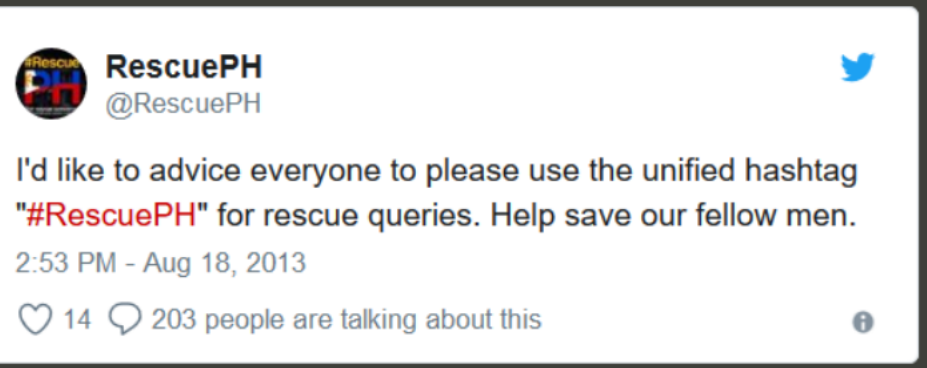


(b) Average Criticality by Information Type (95% Confidence)

SYSTEM STATUS / [404 REMAINING] AWAITING INPUT: SELECT RELEVANT CATEGORIES OR MARK AS UNINFORMATIVE

SEVERE TROPICAL STORM TRAMI, KNOWN IN THE PHILIPPINES AS TROPICAL STORM MARING, WAS A TROPICAL CYCLONE THAT BROUGHT HEAVY RAINS TO TAIWAN AND EAST CHINA DURING MID-AUGUST 2013. THE USER IS A RESPONSE OFFICER RESPONSIBLE FOR METRO MANILA, ONE OF THE THREE DEFINED METROPOLITAN AREAS OF THE PHILIPPINES. WIKIPEDIA PAGE

Tweet Rendered by Twitter



Tweet Text

RT @RescuePH: I'd like to advice everyone to please use the unified hashtag "#RescuePH" for rescue queries. Help save our fellow men.

ID: 369099410526855170

TWEET CONTAINS NO RELEVANT INFORMATION (DELETE) | SKIP TWEET

Request

GOODSSERVICES | SEARCHANDRESCUE | INFORMATIONWANTED

CallToAction

VOLUNTEER | DONATIONS | MOVEPEOPLE

Report

FIRSTPARTYOBSERVATION | THIRDPARTYOBSERVATION | WEATHER

EMERGINGTHREATS | SIGNIFICANTEVENTCHANGE | MULTIMEDIASHARE

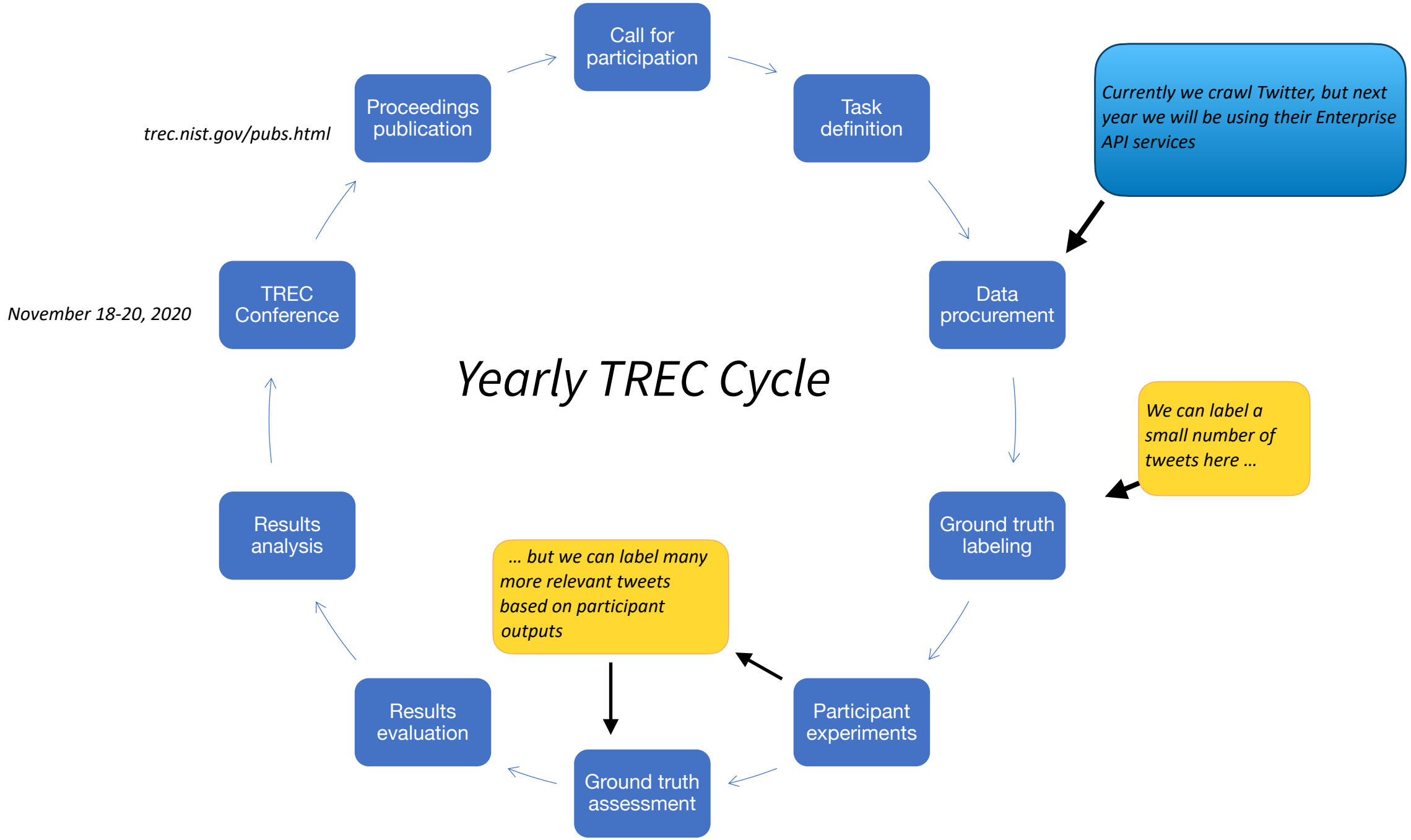
SERVICEAVAILABLE | FACTOID | OFFICIAL | CLEANUP | HASHTAGS

Other

PASTNEWS | CONTINUINGNEWS | ADVICE | SENTIMENT | DISCUSSION

IRRELEVANT | UNKNOWN | KNOWNALREADY

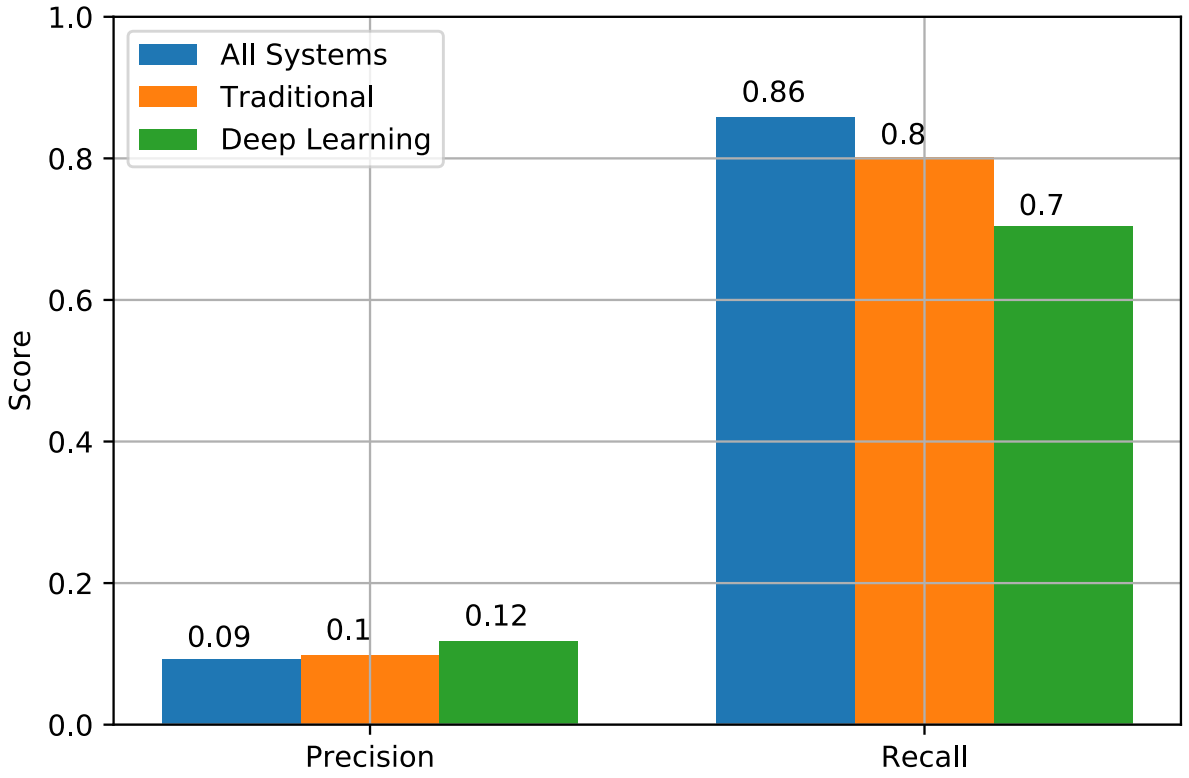
SAVE CATEGORIES


Call for participation
Task definition
Data procurement
Ground truth labeling
Participant experiments
Ground truth assessment
Results evaluation
Results analysis
TREC Conference
Proceedings publication
Yearly TREC Cycle
Currently we crawl Twitter, but next year we will be using their Enterprise API services
We can label a small number of tweets here …
… but we can label many more relevant tweets based on participant outputs
trec.nist.gov/pubs.html
November 18-20, 2020

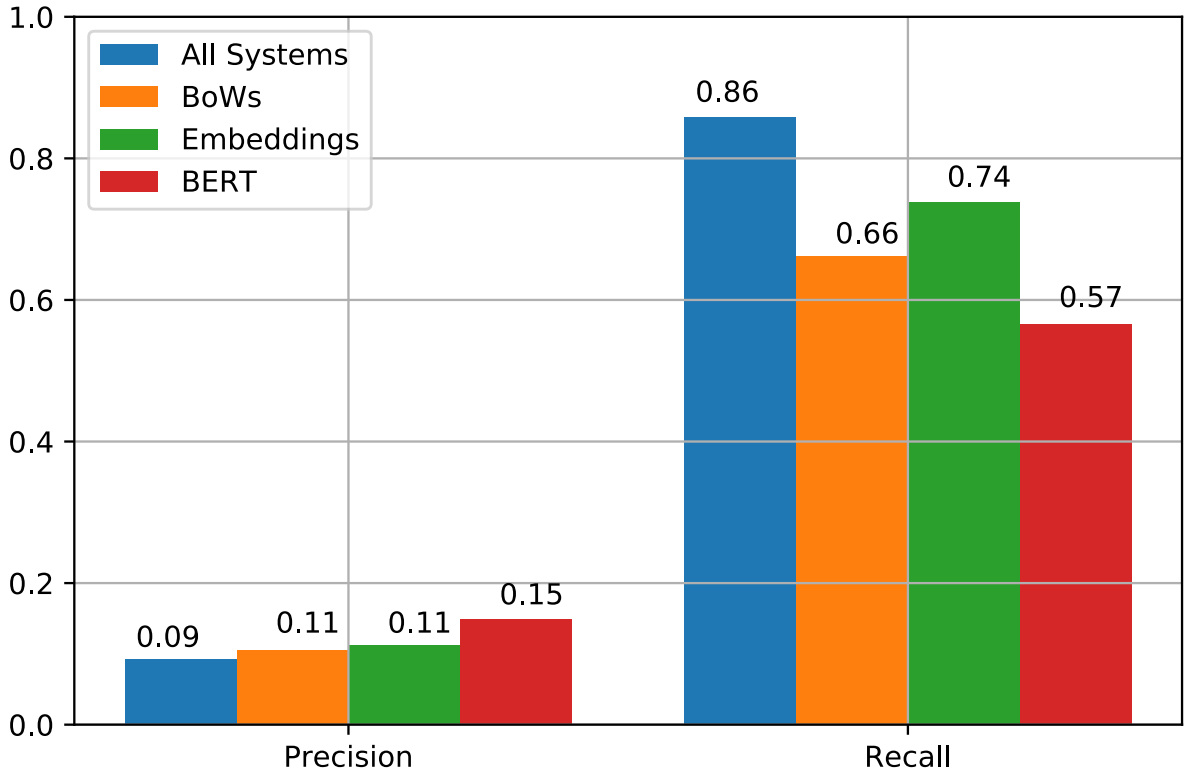
F1 scores

	Request-GoodsServices	Request-SearchAndRescue	Request-InformationWanted	CallToAction-Volunteer	CallToAction-Donations	CallToAction-MovePeople	Report-FirstPartyObservation	Report-ThirdPartyObservation	Report-Weather	Report-Location	Report-EmergingThreats	Report-NewSubEvent	Report-MultimediaShare	Report-ServiceAvailable	Report-Factoid	Report-Official	Report-CleanUp	Report-Hashtags	Report-OriginalEvent	Other-ContextualInformation	Other-Sentiment	Other-Discussion
cbnuC1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cbnuS1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DLR_BERT_R	0	0	0.12	0.05	0.05	0.25	0.11	0.15	0.45	0.08	0.17	0.01	0.29	0.19	0.42	0.2	0.13	0.4	0.21	0.04	0.32	0.06
DLR_Fusion	0	0	0	0	0	0.25	0.03	0.06	0.22	0.04	0.01	0	0.14	0.13	0.33	0	0	0.05	0.29	0.05	0.25	0.03
DLR_MeanMaxAAE_Regress	0	0	0.07	0	0	0.16	0.07	0.05	0.02	0.07	0.01	0	0.12	0.03	0.33	0.04	0.19	0.11	0.29	0.05	0.21	0.02
DLR_SIF_R	0	0	0	0.32	0.32	0.11	0.07	0.15	0.03	0.03	0.03	0	0.14	0.07	0.32	0.06	0.11	0.09	0.25	0.07	0.18	0.03
DLR_USE_R	0.12	0	0.31	0.14	0.14	0.28	0.08	0.11	0.38	0.08	0.16	0.03	0.5	0.2	0.46	0.06	0.26	0.41	0.33	0.06	0.32	0.08
ict_dl	0	0	0	0	0	0	0.09	0.27	0	0.01	0.14	0	0.47	0.07	0.29	0.05	0	0.24	0	0	0.17	0.09
IITBHU_run1	0	0	0	0	0	0.02	0.01	0.19	0.26	0.01	0.02	0	0.41	0.01	0.1	0.01	0	0.03	0.04	0	0.15	0.09
IITBHU_run2	0	0	0.01	0	0	0.02	0.06	0.15	0	0.02	0.15	0.01	0.01	0	0.34	0.02	0.02	0	0	0.07	0.01	0.01
Informedia-nb	0.07	0	0.11	0	0	0.37	0.11	0.08	0.44	0.23	0.13	0	0.21	0.28	0.54	0.26	0	0.46	0.31	0.06	0.35	0.09
Informedia-rf1	0	0	0	0	0	0.19	0.1	0.13	0.04	0	0	0	0.49	0.18	0.44	0	0	0.53	0	0.04	0.32	0.08
Informedia-rf2	0.05	0	0	0	0	0.11	0.08	0.19	0.05	0	0.07	0	0.38	0.26	0.37	0	0	0.44	0	0.03	0.33	0.08
Informedia-rf3	0.02	0	0	0	0	0.02	0.01	0.06	0	0	0.18	0.01	0	0.15	0.21	0	0	0.01	0	0	0.31	0.03
IRITrun1	0	0	0.11	0	0	0.03	0.11	0.28	0.12	0.04	0.05	0.01	0.61	0.02	0.39	0.03	0	0.64	0.32	0.05	0.26	0.08
IRITrun2	0	0	0.1	0	0	0.03	0.1	0.28	0.12	0.03	0.04	0	0.5	0.05	0.47	0.06	0	0.55	0.32	0.05	0.27	0.08
IRITrun3	0	0	0.06	0	0	0.02	0.1	0.02	0	0	0	0	0.66	0	0.24	0.03	0.04	0.7	0.03	0.06	0.22	0.05
IRITrun4	0	0	0.06	0	0	0	0.08	0.31	0	0	0	0	0.71	0	0.4	0.02	0	0.43	0.03	0.05	0.16	0.07
nyu.base.multi	0.04	0.01	0.05	0.05	0.05	0.16	0.09	0.26	0.42	0.38	0.22	0.06	0.54	0.22	0.5	0.25	0.05	0.55	0.37	0.05	0.33	0.1
nyu.base.sing	0.05	0	0.11	0.12	0.12	0.16	0.09	0.03	0.34	0.1	0.12	0.01	0.15	0.13	0.25	0.18	0.03	0.15	0.3	0.05	0.3	0.1
nyu.fast.multi	0	0	0.09	0	0	0.22	0.06	0.11	0.25	0.43	0.24	0.01	0.62	0.17	0.54	0.26	0.03	0.69	0.13	0.05	0.33	0.13
nyu.fast.sing	0	0	0.04	0.07	0.07	0.19	0.01	0.09	0.08	0.3	0.06	0.05	0.17	0.11	0.43	0.15	0	0.16	0.01	0.07	0.23	0.14
run1_baseline	0.02	0	0	0	0	0.04	0.02	0.01	0.11	0	0.03	0.03	0.06	0	0	0	0	0	0	0	0.18	0.11
run2_negative	0.02	0	0	0	0	0.04	0.01	0.01	0.1	0	0.03	0.02	0.06	0	0	0	0	0	0	0	0.16	0.1
run3_irn	0.02	0	0	0	0	0.04	0.02	0.01	0.11	0	0.03	0.03	0.06	0	0	0	0	0	0	0	0.18	0.11
run4_all	0.02	0	0	0	0	0.04	0.01	0.01	0.1	0	0.03	0.02	0.06	0	0	0	0	0	0	0	0.16	0.1
UCDbaseline	0.08	0.01	0.04	0.06	0.06	0.17	0.09	0.15	0.31	0.58	0.22	0.05	0.48	0.31	0.65	0.22	0.06	0.53	0.35	0.06	0.28	0.1
UCDbcnelmo	0.03	0	0.06	0.03	0.03	0.26	0.13	0.07	0.3	0.13	0.19	0.01	0.26	0.26	0.59	0.21	0	0.02	0.21	0.06	0.4	0.06
UCDbilstmalpha	0.08	0	0	0.07	0.07	0.27	0.1	0.09	0.36	0.11	0.05	0.03	0.2	0.21	0.6	0.08	0	0.37	0.31	0.06	0.32	0.04
UCDbilstmbeta	0.09	0	0.07	0.18	0.18	0.18	0.09	0.04	0.37	0.11	0.25	0.04	0.2	0.27	0.35	0.15	0.03	0.15	0.26	0.05	0.34	0.08
UPB-BERT	0.08	0.02	0.22	0.05	0.05	0.24	0.14	0.17	0.44	0.01	0.24	0.09	0.48	0.28	0.5	0.24	0.15	0.41	0.36	0.05	0.31	0.1
UPB-FOCAL	0.14	0.03	0.13	0.09	0.09	0.36	0.13	0.19	0.45	0	0.16	0.09	0.46	0.28	0.59	0.19	0.16	0.38	0.33	0.04	0.3	0.11

Comparison of approaches



(a) Learning Paradigm Groups



(b) Text Featurization Groups

Outcomes

- 33 emergency events collected and annotated.
- 11 teams participated in the 2018 challenge,
  10 teams in 2019,
  expecting about the same for 2020.
- Papers about the effort published in ISCRAM 2019 and 2020.
- All event datasets freely available from trecis.org.
- Papers by teams at trec.nist.gov under Publications ⇢ Proceedings.

#PSCR2020