Where the River Rises: A River of Words.

Source Going Deeper or Flatter: Connecting Deep Mapping, Flat Ontologies and the Democratizing of Knowledge

Humanities **2015**, *4*(4), 623-636; doi:10.3390/h4040623
http://selinaspringett.com/wp/?p=166

Weather events identification in social media streams: tools to detect their evidence in Twitter

Alfonso Crisci (1)

Valentina Grasso (1,2), **Imad Zaza** (3), **Federica Zabini** (1,2), **Paolo Nesi** (3) and **Gianni Pantaleo** (3)

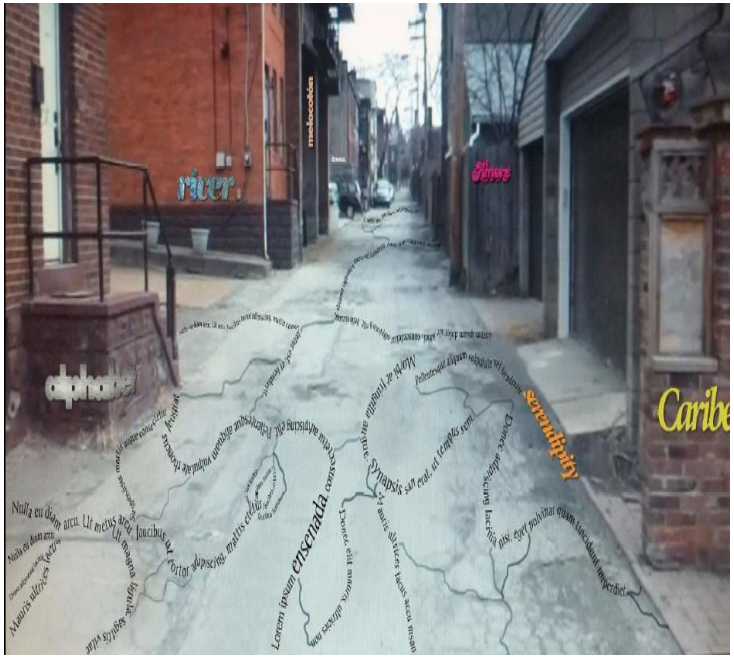
(1) CNR Ibimet, Italian National Research Council, Florence, Italy (v.grasso@ibimet.cnr.it)
(2) LaMMA Consortium, Italian National Research Council, Florence, Italy.
(3) DISIT Lab, Distributed [Systems and internet | Data Intelligence and] Technologies Lab, Dep. of Information Engineering (DINFO), University of Florence, Italy

OGRS'16, October 12-14, 2016, Perugia, Italy

Open Source Geospatial Research & Education Symposium OGRS is a meeting dedicated to sharing knowledge, new solutions, methods, practices, ideas and trends in the field of geospatial information through the development and the use of free and open source software in both research and education.

Background

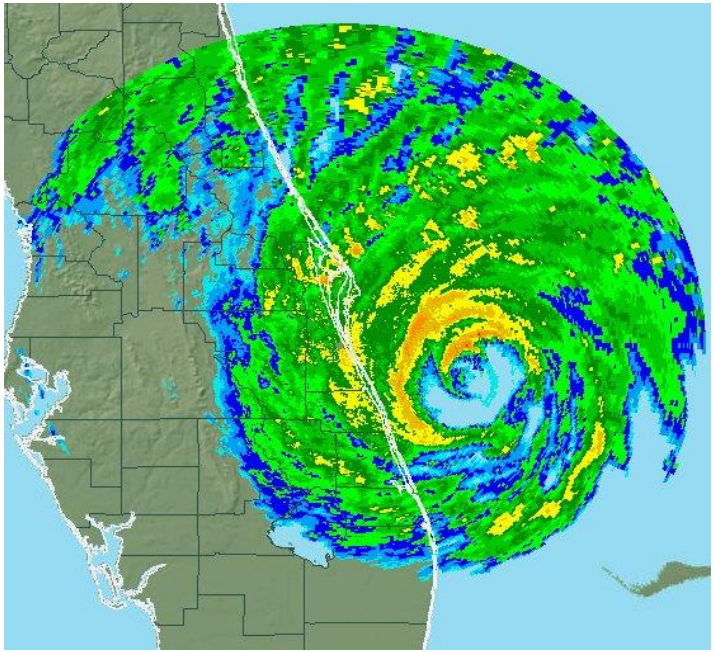
- Social media data become a powerful real time informative source to monitor the impacts of meteoclimatic events.
- The reliability of social media activity and content on weather related natural risk is a current topic.
- Weather events are bounded in time and space and could have a deep impact on civil society
- Performing real-time events detection from social media sources requires tools and a cross-disciplinary framework.
- Methods to detect and track natural hazards by using public informative continue to increase in coverage, resolution and reliability.



River of Words is a project by writer Israel Centeno and visual artists Carolina Arnal and Gisela Romero Pittsburg

Aims of work

- Report a 4 year experience on analysis of Twitter social media (SM) streams related to weather.
- Share background needed for an effective information extraction from public SM flows paying attention to a fast and real detection of impacting weather events in order to increase situational awareness.
- Showing some empirical evidences to help implement a real open toolchain leading to a visual dashboard for regional/national weather services.



Met Office ha ritwittato

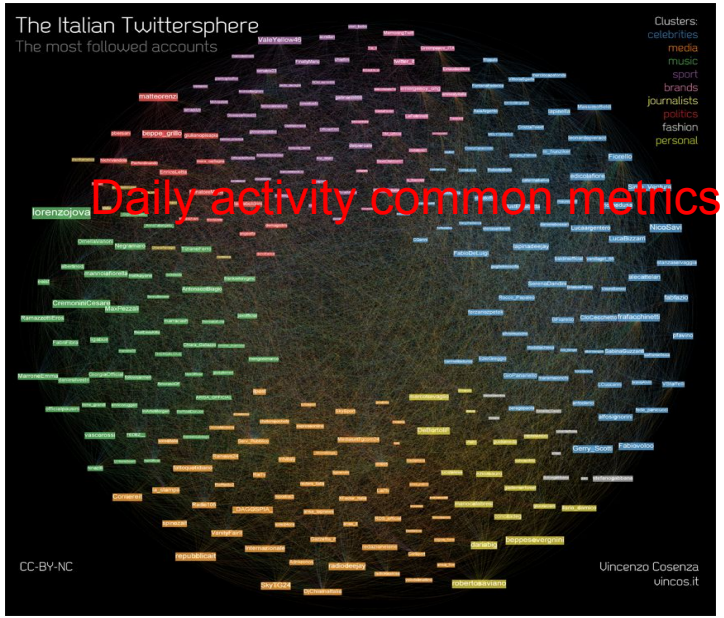
Met Office Storms @metofficestorms · 59 min

The eye and strongest winds of #Hurricane #Matthew moving perilously close to the #Florida coast. Peak winds 120 mph around the eyewall.

Social Media Public Sources: crowd-sensing by using Twitter data

On-line multilanguage platform for social-networking and microblogging.
Twitter data perform significant crowd-sensing.

6.4 million of active users in Italy (2015)



Who feed information in ?

Many and various

Citizen, Institution, Institutional Public Services, Business Companies, Community - NGO organization, Media,Conversational Bots and Sensors Bots.

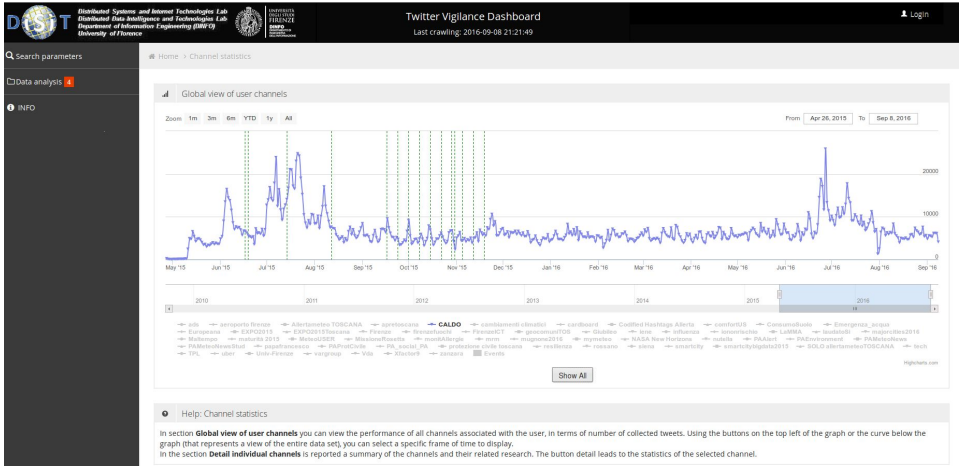
Metrics of activity

RTW_TW: N° of tweets & retweets
RTW: N° of tweets & retweets
TW: N of native tweets & retweets
U_native_users: N° of native TW authors
U_full_user: N° authors of TW & RTW
U_unique_hashtag: N° of hash-TAG

Open Source tools

https://github.com/alfcrisci/rTwChannel

DATA EXTRACTION by word query : the Twitter Vigilance DISIT Model



http://www.disit.org/tv/

Twitter VIgilance platform is an environment developed by DISIT University of Florence that:

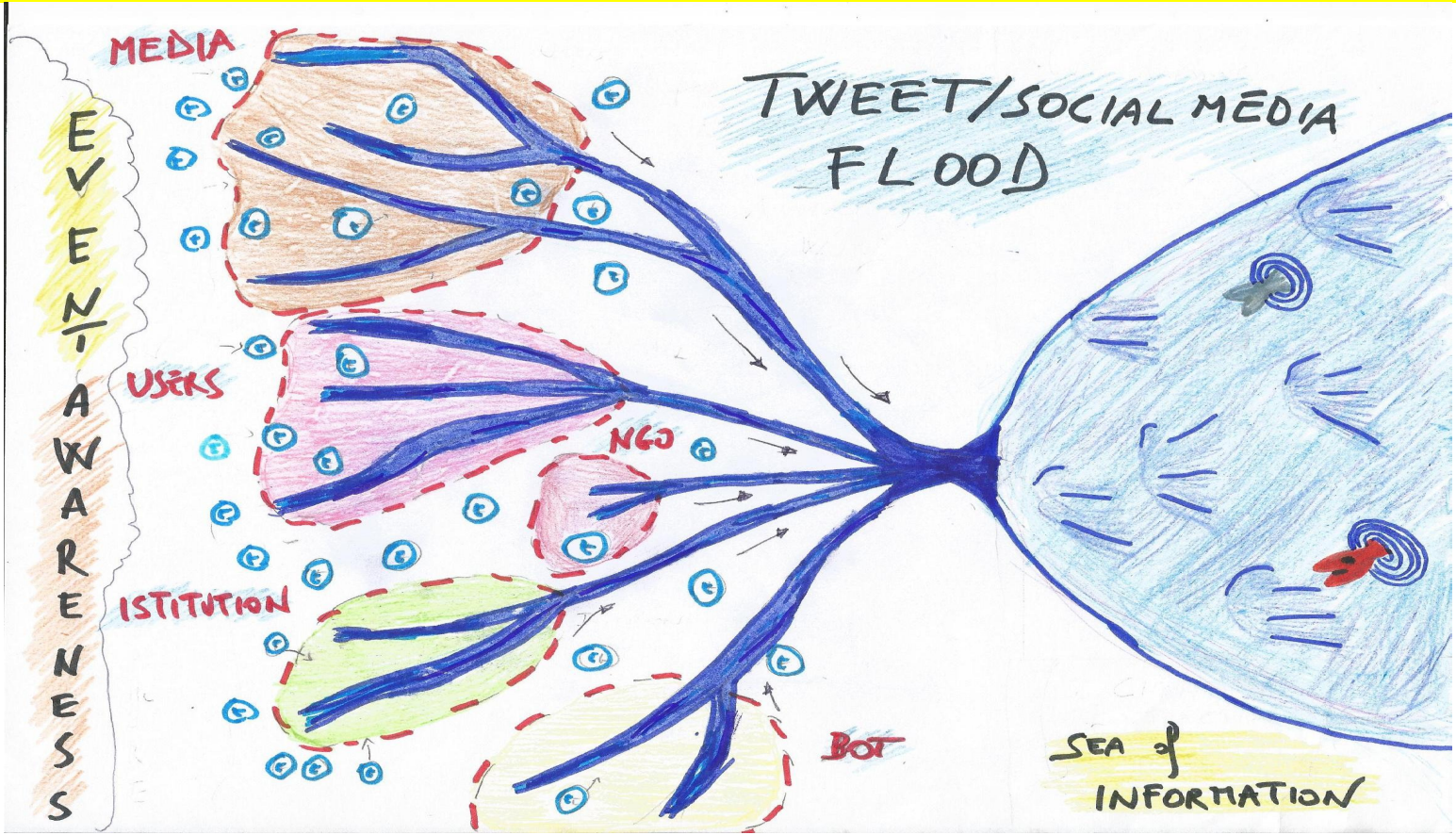
- Manage multiple queries in twitter API
- Store the data of messages collected by user defined queries -> channel
- It is a dashboard able to visualize data collecting process & analytics of twitter metrics of channel

Social media data: Channel “CALDO”

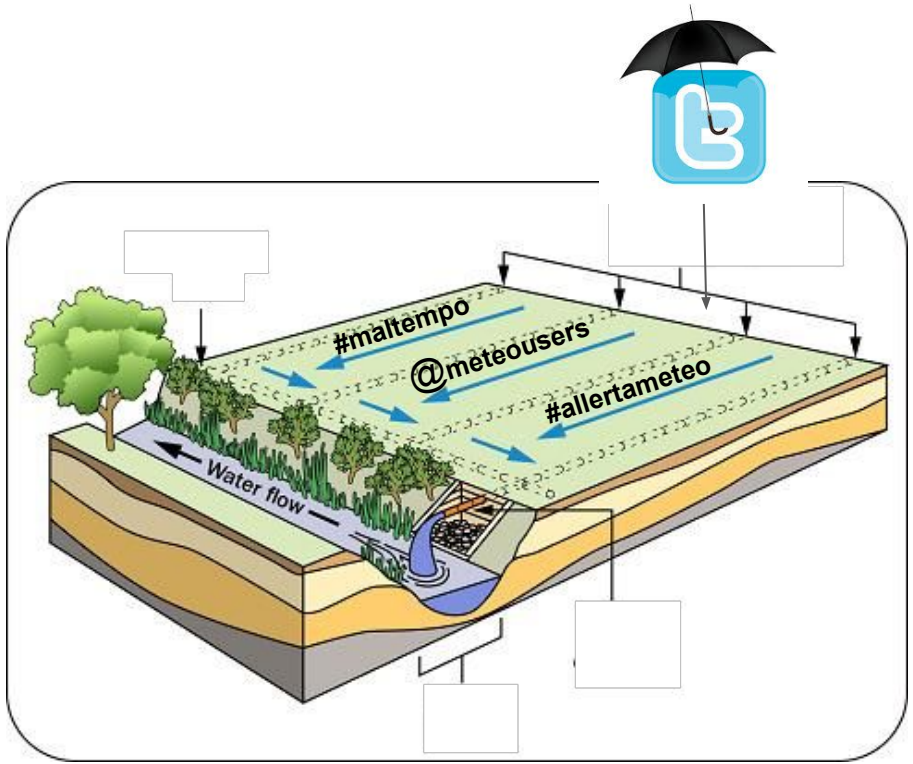
http://www.disit.org/tv/index.php?p=chart_singlechannel&canale=CALDO

Channel	Related research	Total	N° tweets	N° tweets(%)	N° retweets	N° retweets(%)	Details	Analysis
CALDO	#caldo #sole afa allerta allerta caldo anticiclone caldissimo caldo canicola disagio caldo emergenza caldo estate rovente flegetonte ondata di calore rischio rovente temperature torrido	3344111	1851650	55.37%	1492461	44.63%	From 2009-10-23 To today	From 2015-05-15 To2015-09-15 NLP SA

EVENT DETECTION in social media: fishing words/tags in a FLOOD of in dropped messages



EVENT DETECTION in social media: words/tags creates “word drainage” channels



Different queries correspond to different data extraction volumes.

Each one can be summarized as daily statistics by using key activity metric and obtaining different representable flows in function of the querying parameters (terms, hashtags (#) or users (@).

DATA EXTRACTION : Twitter corpora have the highest lexical diversity

	WIKIPEDIA	TWITTER	BOOKS	WEB
SIZE (ratios)	1 X	~1.2 X	~100 X	~200 X
LEXICAL DIVERSITY	483 k	736 k	135 k	206 k
CURATION	Very High, Peer Reviewed, Updated Frequently	None. High rate of typos and non standard language	Professionally edited.	Mix
REGISTER	Very Formal, Reporting Fashion	Very Informal, Colloquial	Formal, Narrative style	Mix
OBJECTIVITY	Completely Factual	More Opinions	More Fictional	More factual

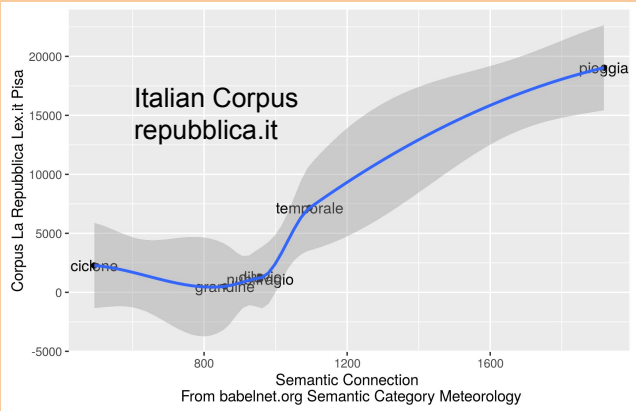
Twitter shows many OOV out-of-vocabulary terms that change in function of the events.

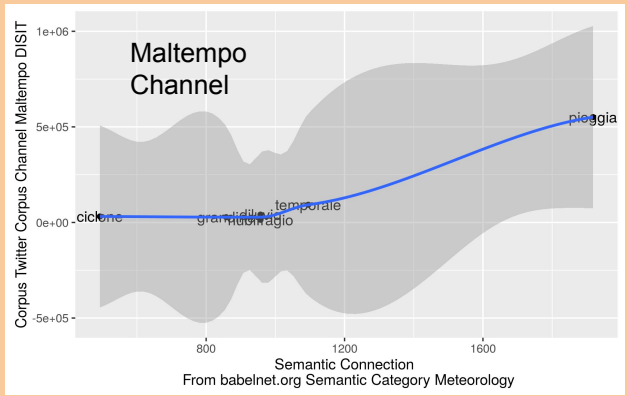
Considering specific features of Twitter textual data it is ever hard to mine conversations.

Modeling Word Meaning: Distributional Semantics and the Corpus Quality-Quantity Trade-Off
S Sridharan, B Murphy
Proceedings of the 24th International Conference on Computational Linguistics

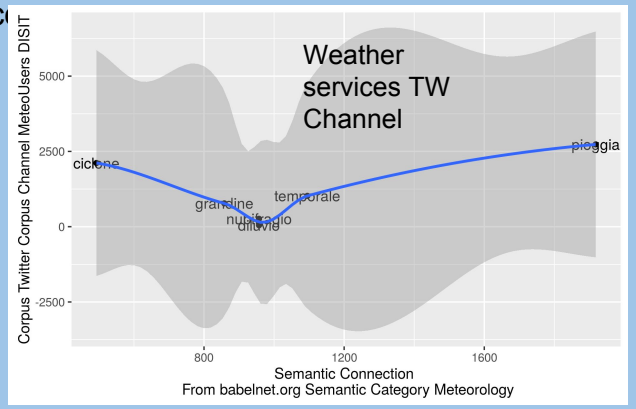
DATA EXTRACTION : Terms of channel have different extractive capabilities

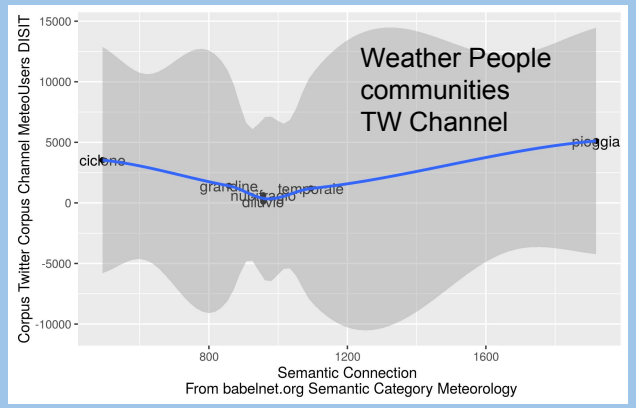
More “crowd” corpora





More “oriented community based” c

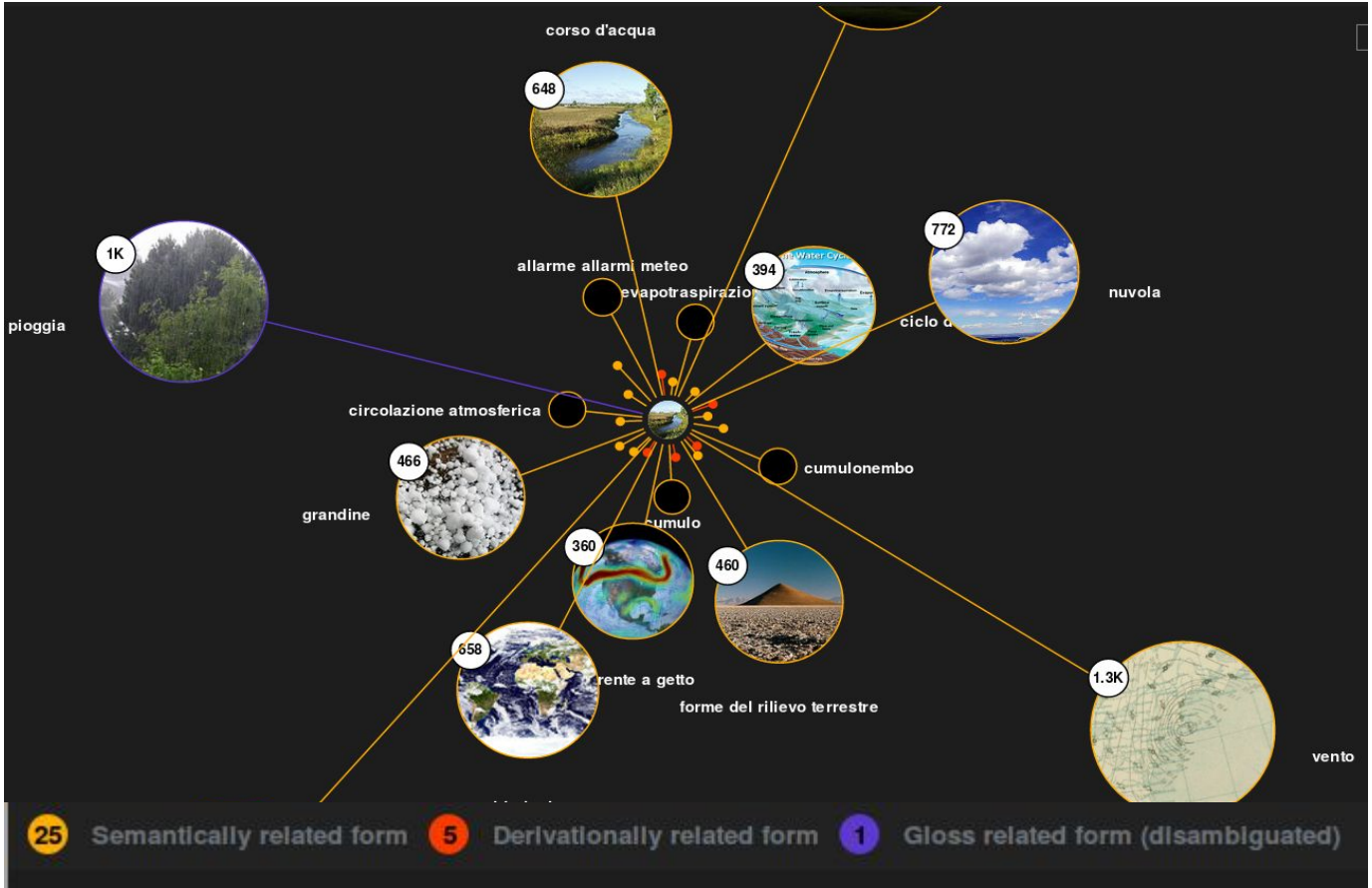




Semantic capabilities (N of connection) of words are different
and their power of extraction in the Twitter streams differs in function to Users sets.
In popular streams wide sense words seems also high frequency words and are able extract more messages.

Babelnet.org metrics are used for meteorology category.

Semantic network of word sense: word “PIOGGIA” from Babelnet.org



EVENT IDENTIFICATION: Social media and real-world event condition for linking

Identification and Characterization of Events in Social Media

Hila Becker
Thesis
Columbia University 2011

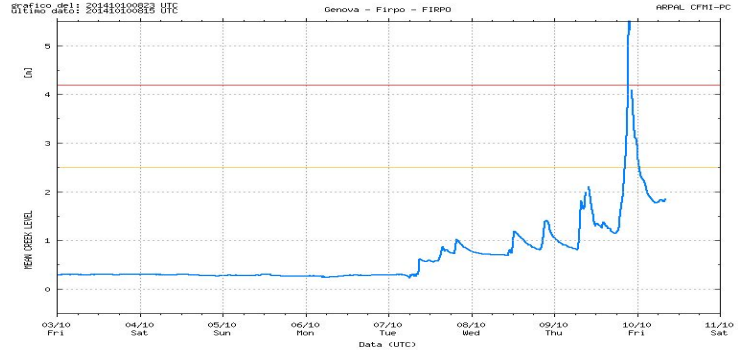
http://www.cs.columbia.edu/~hila/publications.html

- **Document stream**: a time ordered sequences of featured documents D_{ev}
- **Trending time**: a trending time period for a feature over a document stream is time period where document frequency of the features in document stream is substantially higher than expected.
- **Trending event** : is a real-world occurrence with an (1) *associated time* (T_{ev}),(2) a *stream document* (D_{ev}) about the occurrence and published during T_{ev} and (3) one or more features that describes the occurrence and for which T_{ev} is trending time period over document stream (D_{ev}).

EVENT DETECTION in social media: what is a Twitter channel?

- Channel is a conversation stream obtained by a recursive API query of tweets (Twitter API 1.1)
- Channel produces a dataset with a time dimension.
- Channel has a set of contributors (original unique authors or amplifiers by re-tweet mechanism).
- Channel is a virtual conversational space among different communities.
- Channel is a textual corpus with specific linguistic and semantic properties.
- Channel has proper activity dynamics and behaviours regarding social media mechanisms (retweeting, mentioning, tagging, media and linking content)
- Channel could be representative of a specific web-community or not.
- Semantic oriented channels could represent a set of topic specific informations.

EVENT DETECTION. Social media and real-world event: conditions for their link.



As a flood needs rain, in social media real-impact events needs documents.

Quantitative relation relies on some factors:

- Population density of the area impacted
- Level of awareness of people involved
- Digital literacy of impacted population
- Institutional preparedness to weather related risks

EVENT DETECTION : 12-08-2015 Twitter channels for Rossano (Calabria) flood event

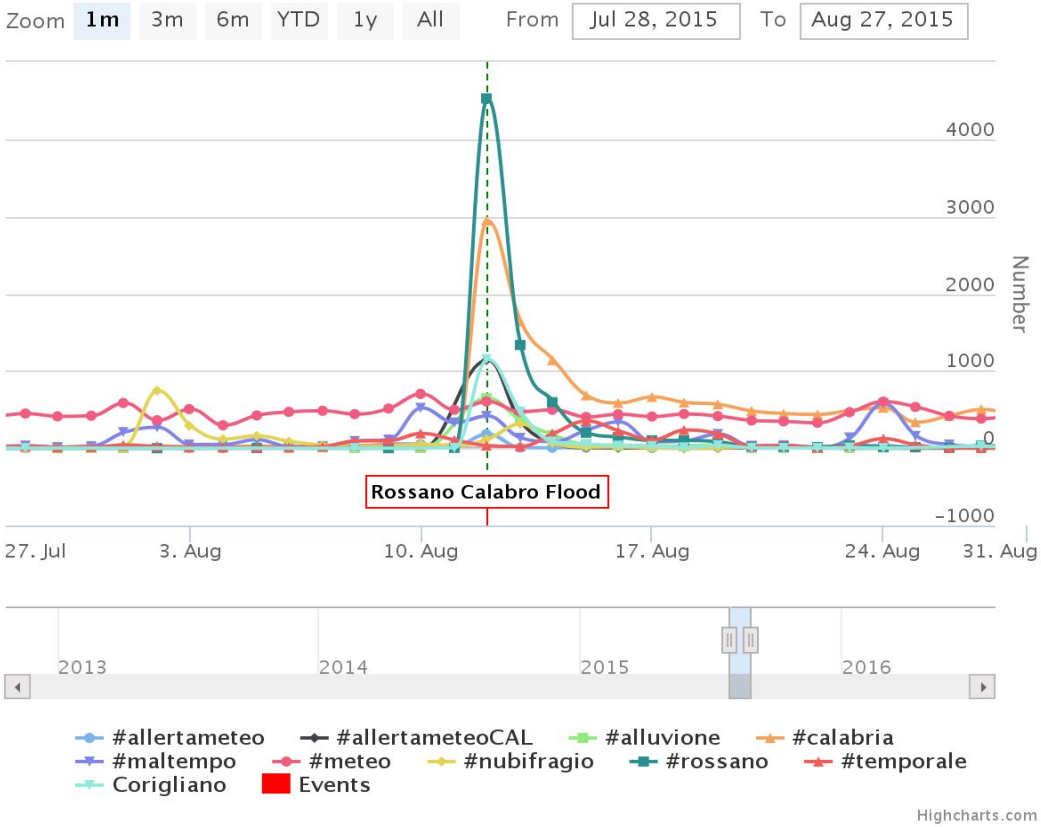


Image source :
http://www.caritasitaliana.it/pls/caritasitaliana/

EVENT DETECTION : 01-10-2015 Twitter channels for Olbia (Sardinia) flood event

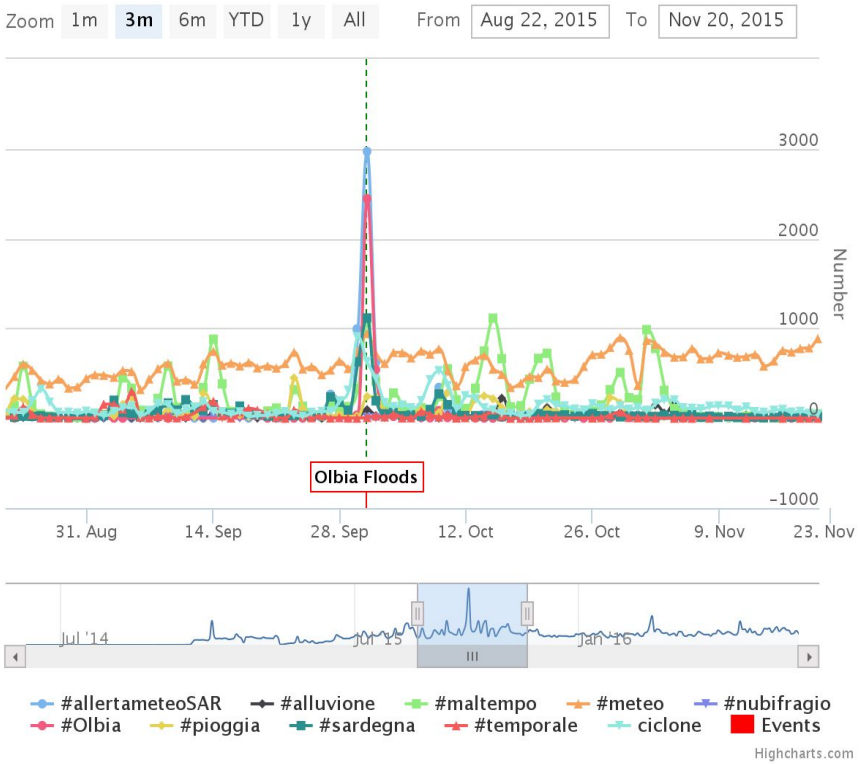


Image source
http://www.ilpost.it/2013/11/19/foto-alluvione-sardegna/alluvione-sardegna-14/

EVENT DETECTION : 01-08-2015 twitter channels for Firenze Toscana burst event

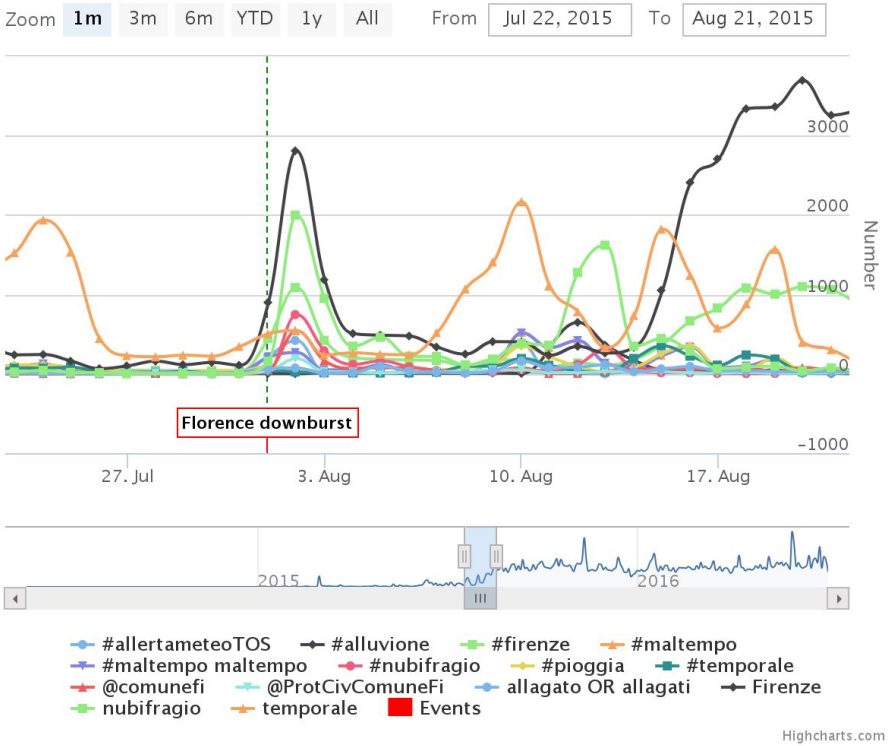
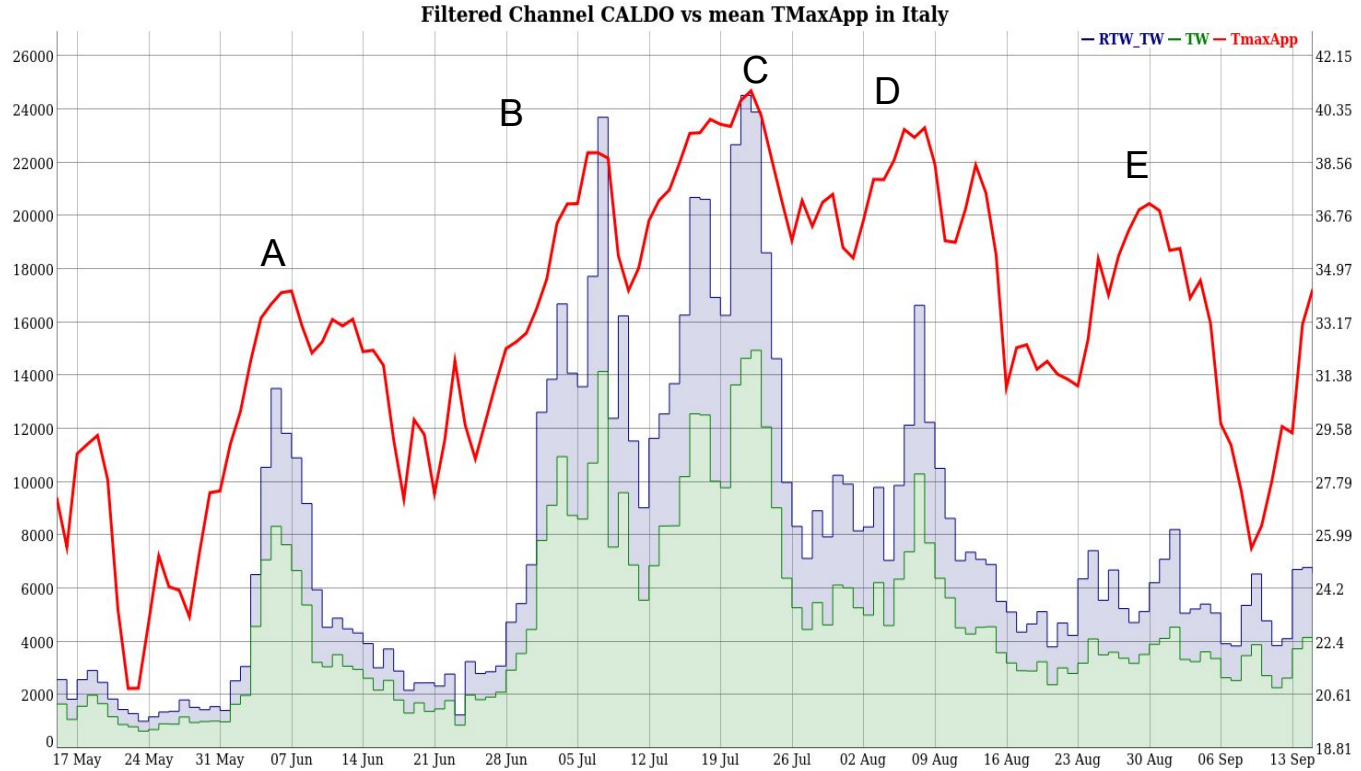


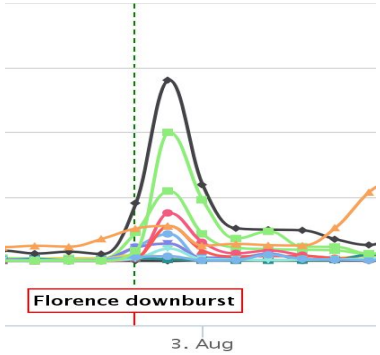
Image Sources:
http://aldopiombino.blogspot.it/2015/08/il-downburst-del-temporale-di-firenze.html

Channel “Caldo” and Heatwave periods 2015 (15 May to 15 September)



Framing conjectures for detection of weather events

- For weather events that have a slow impact “trending periods” of SM activity work well.
- For the detection of fast and sudden events the synchronization of trending periods could be exploited by using search terms with different semantic extents.
- During synchronization time document frequency reach its relative maximum creating a well recognizable pattern and event is clearly detected (half onion effect).
- Geographical terms show their importance because are good proxy of event ‘s situational awareness.
- During severe events trending topics generally contain the name of the places involved.



http://www.shutterstock.com/pic-86262523/stock-photo-a-red-onion-sliced-in-half-isolated-on-white-background.html

Conclusion

- Public & accessible social media data could be considered as a real huge informative data stream concerning severe weather events or other climatic threats.
- Social media (SM) data acquisition, storage and filtering to obtain reliable data collections requires huge work and dedicated tools. It is a real challenge for open source developers.
- Finding appropriate query-terms to obtain suitable data. The semantic tuning of TW channels is ever required. Platforms as TwitterVigilance are suitable example.
- Weather events are different in space, time and atmospheric processes involved. Events detection requires a specific search strategy. Appropriate use of semantic features of words point to interesting directions.

Contacts

Communication analyst

Valentina Grasso
grasso@lamma.rete.toscana.it
v.grasso@ibimet.cnr..it

Federica Zabini
zabini@lammarete.toscana.it

http://www.lamma.rete.toscana.it
http://www.ibimet.cnr.it/

Biometeorology

Alfonso Crisci
a.crisci@ibimet.cnr.it
alfcrisci@gmail.com
@alf_crisci

Informatics Science

Imad Zaza
Imad_zaza@unifi.it
Paolo Nesi
paolo.nesi@unifi.it
Gianni Pantaleo
gianni.pantaleo@unifi.it

http://www.disit.org/
http://www.disit.org/tv/

Data & Code:
https://github.com/alfcrisci/ogrs_2016_weathersocial_paper

Image from :https://commons.wikimedia.org/wiki/File:Thank-you-word-cloud.jpg

This study was carried out in the field of the CARISMAND Project: *Culture And RISkmanagement in Man-made And Natural Disasters*