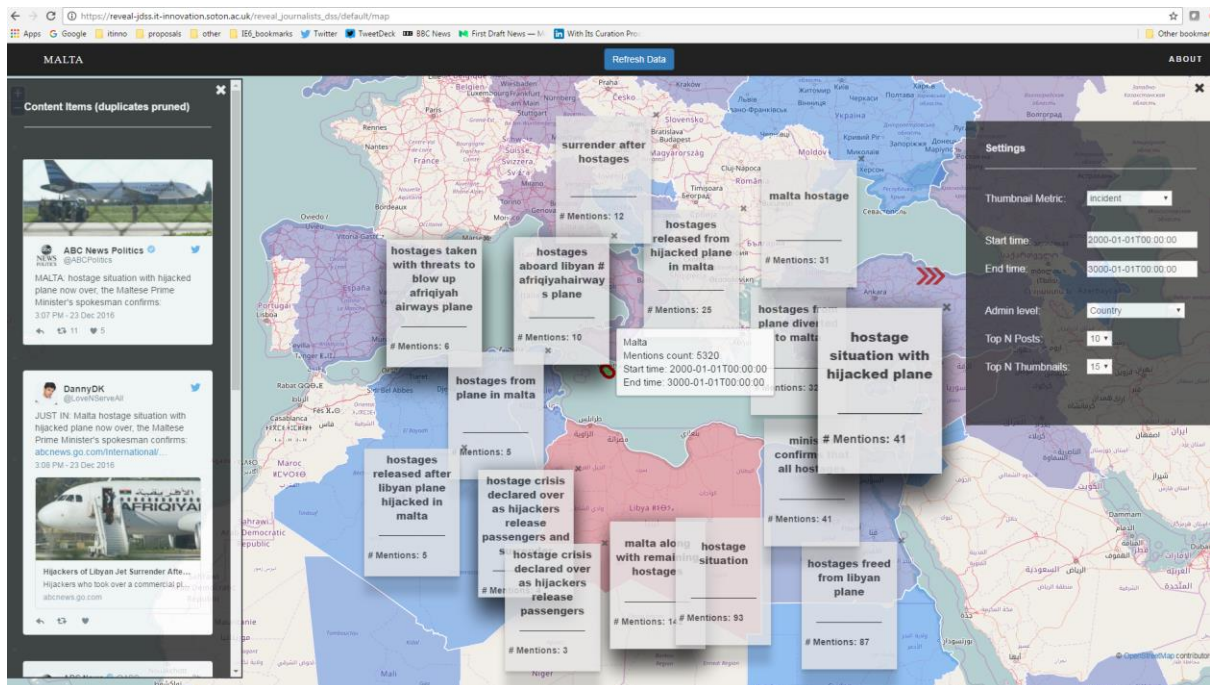


18th Jan 2017

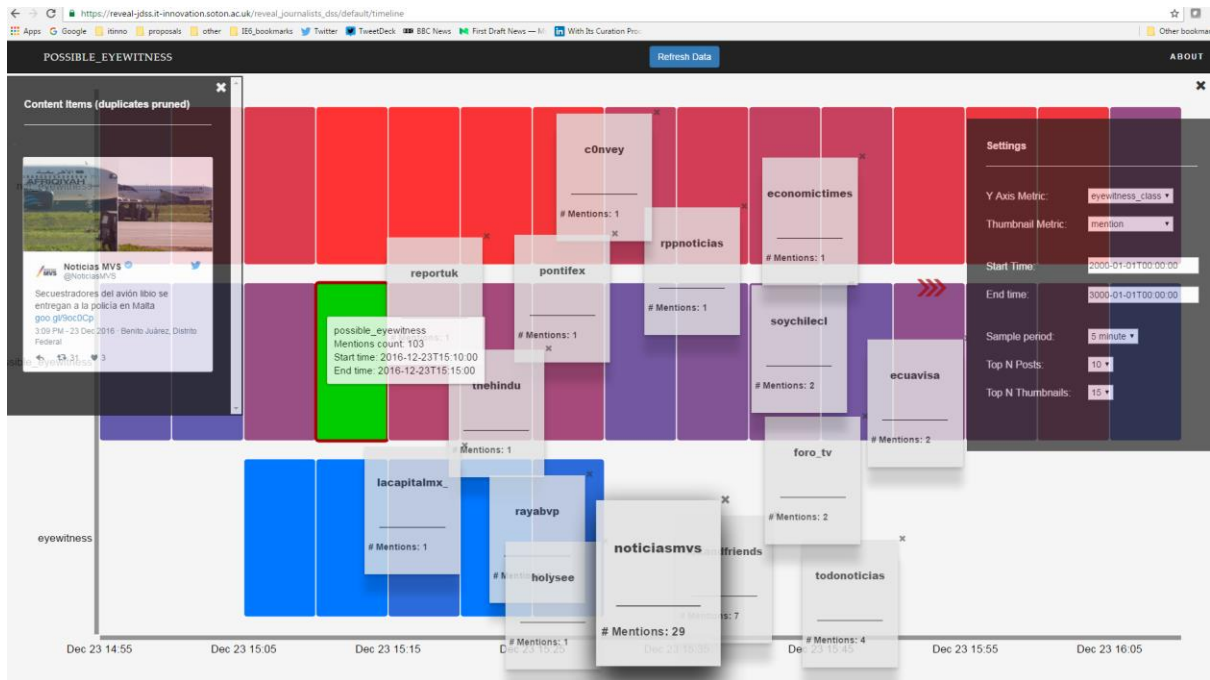
Journalist Decision Support System (JDSS)

The Journalist Decision Support System is a free scalable Twitter analytics platform allowing journalists to crawl Twitter for posts and find user generated content (UGC) relevant to verification tasks. Up to 19 journalists can use JDSS simultaneously, each interactively browsing 10,000's of posts in real-time. Background analytics are automatically run on all posts, including sentiment analysis, fake and eyewitness media labelling and newsworthy claim extraction. Journalists can interactively explore posts, clustering and sub-clustering the data to quickly find groups of contextual posts highly related to the event or claim being verified.

The Journalist Decision Support System quickly gets journalists to the right content so they can make UGC verification decisions quicker. You need a Chrome browser and a Google login - then just click [here](#) to get started.



JDSS map showing posts from Malta, sub-clustered by a specific claim of a hostage situation



JDSS Timeline showing posts in a 5 minute sample window, with an embedded or linked image/video classified as a 'possible eyewitness'

Real-time Twitter Searching

There is no limit the number of searches that can be run, but only 3 searches can be active at the same time for any user. Each search is limited to a maximum of 10,000 result posts before it finishes. All result posts are added to the user's JDSS post collection and rendered on the interactive map and timeline for browsing. Globally 19 searches can be active in parallel at any one time.

Searches are based on a set of [keywords](#) (e.g. 'Trump OR Hillary'), a Twitter List of usernames or a Twitter Collection of posts. The free Twitter search API is used behind the scenes so there are rate limits applied by Twitter. Only posts within 7 days can be searched and the search throughput is limited to 400 posts per minute for each parallel search. This means a user running 3 parallel searches can get 10,000 posts in 8 minutes.

Interactive Browsing and Clustering

Posts found are rendered on a real-time interactive map and timeline. Posts appear as they are found and the views are updated live.

Each post is geoparsed (using [geoparsepy](#)) to extract each location mention. These locations are used to spatially group posts on a map view. Different administrative levels can be viewed, from country level all the way down to local suburbs. Each location can be clicked on the map and all the posts mentioning this location, or any sub-location, are shown on a sidebar sorted so earliest post appears first. For example clicking on London will show posts mentioning London, City of London, Westminster, Wimbledon etc. This provides a much more powerful spatial browsing experience than simple keyword search allows (e.g. posts with '#London').

Posts can also be displayed on a timeline, broken down into samples of period 1 day, 1 hour or 5 minutes. All samples are selectable so, for example, it is easy to see posts for a specific 5 minute period during an event.

Lots of analytics are applied automatically to the result posts and popup sub-cluster thumbnails can be shown for any location or time sample. These thumbnails allow a second level of clustering to provide a way to really zoom into groups of posts that match each journalist's specific requirements. For example the location 'London' can be selected and thumbnails popped up for incident report claims within the posts that mention a location within London. Each claim thumbnail can itself be selected and the posts behind it shown sorted by earliest mention on the sidebar.

Newsworthy Claim Extraction

Newsworthy claim patterns [2] are automatically extracted for posts in English, French, Spanish or German. Incident reports (e.g. magnitude 7 earthquake in Nepal) and reports of event actors (e.g. gunmen in Brussels airport) are extracted, along with associated damage reports (e.g. tens dead in Paris). These are all available for thumbnail subclustering when interactively browsing posts. For example a user can see posts in a 5 minute sample that also claim 'tens dead in Paris'.

Eyewitness and Fake Classification

Two automated classifiers [1] [2] are provided to identify posts likely to be either eyewitness media or real/fake content. The eyewitness classifier looks for linguistic patterns in each post's textual content that research has observed are historically associated with UGC from eyewitnesses during major events. The fake classifier does the same thing but for linguistic patterns in posts that are associated with real or fake media. These are also available for subclustering when interactively browsing. For example a user can see posts for locations in Paris which contain embedded or linked media that is 'possible eyewitness' content.

Sentiment Analysis

The [VADER](#) sentiment analysis algorithm is applied to each post, showing how positive/neutral/negative the text is. For example a user can see posts for locations in Paris that are strongly negative.

Acknowledgement

The work presented in this article is part of the research and development in the REVEAL project (grant agreement 610928), supported by the 7th Framework Program of the European Commission.

About the author



Stuart E. Middleton is a senior research engineer at the University of Southampton IT Innovation Centre. His main research interests are social media, sensor systems, data fusion and semantics. Stuart has a PhD in Computer Science from the University of Southampton.

@stuart_e_middle @IT_Innov

<http://www.it-innovation.soton.ac.uk> <http://users.ecs.soton.ac.uk/sem/>

REVEAL project, @RevealEU

<http://revealproject.eu/>

https://reveal-jdss.it-innovation.soton.ac.uk/reveal_journalists_dss/

References

- [1] Middleton, S.E. Krivcovs, V. 2016. *Geoparsing and Geosemantics for Social Media: Spatio-Temporal Grounding of Content Propagating Rumours to support Trust and Veracity Analysis during Breaking News*, ACM Transactions on Information Systems (TOIS), 34, 3, Article 16 (April 2016), 26 pages. DOI=10.1145/2842604
- [2] Middleton, S.E. "Extracting Attributed Verification and Debunking Reports from Social Media: MediaEval-2015 Trust and Credibility Analysis of Image and Video", MediaEval-2015, Wurzen, Germany, Sept 2015