



Social Media Verification

REVEAL Project - Trust and Credibility Analysis

Stuart E. Middleton

University of Southampton IT Innovation Centre

sem@it-innovation.soton.ac.uk

www.it-innovation.soton.ac.uk



Overview

RDSM 2015 Invited Talk Overview

- REVEAL Project
- Modality Analysis of Social Media Streams
- Geosemantics and Spatio-Temporal Grounding of Rumours
- Knowledge-based Approach to Trust and Credibility Modelling
- Future Work



REVEAL Project

Overview

- Objectives
 - Enable users to reveal hidden 'modalities' such as reputation, influence or credibility of information
- Approach - Modality Extraction and Analysis
 - Real-time modality extraction
 - On-demand analytics capabilities
 - Event-driven architecture using RabbitMQ to communicate
 - Processing based on a scalable STORM cluster (real-time) & standalone HTTP services (on-demand)
- Journalism Use Case
 - Newsgathering - Find newsworthy content and evidence to help verify this content
- Enterprise Use Case
 - Forums - Identify and help newbies, track product feedback & sentiment & emerging trends



Modality Analysis of Social Media Streams

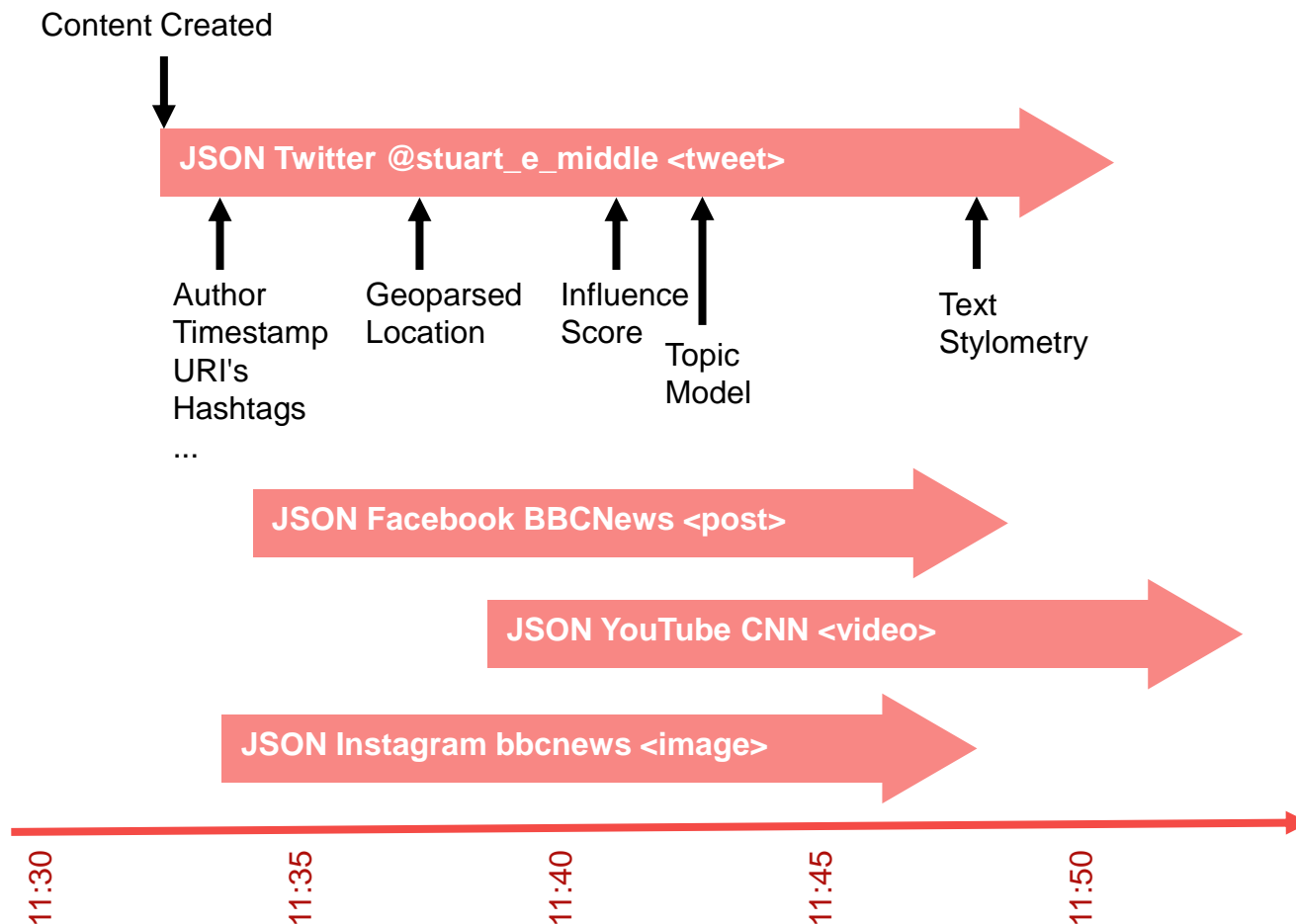
Modality Extraction

- **Social Media Streams**
 - Twitter, You Tube, Instagram, Facebook, Four Square
 - Search (historical) and Stream (real-time)
- **Social Network Analysis**
 - Community Detection, Community Graph Extraction, Community Classification (e.g. topics), Role Analysis (e.g. popular participant), Influence Models ...
- **Content Analysis**
 - Image Feature Extraction (e.g. sky, city), Image Similarity Clustering, Multimedia Indexing, Image Manipulation Detection, Topic Models, Original Content Detection, Text Stylometry ...
- **Geospatial Analysis**
 - Geoparsing, Geosemantic Classification, Image-based Geolocation, Geospatial Topic Model ...
- **Semantic Analysis**
 - Directed Linked-Data Crawler, Semantic Context (e.g. context summaries based on linked data) ...



Modality Analysis of Social Media Streams

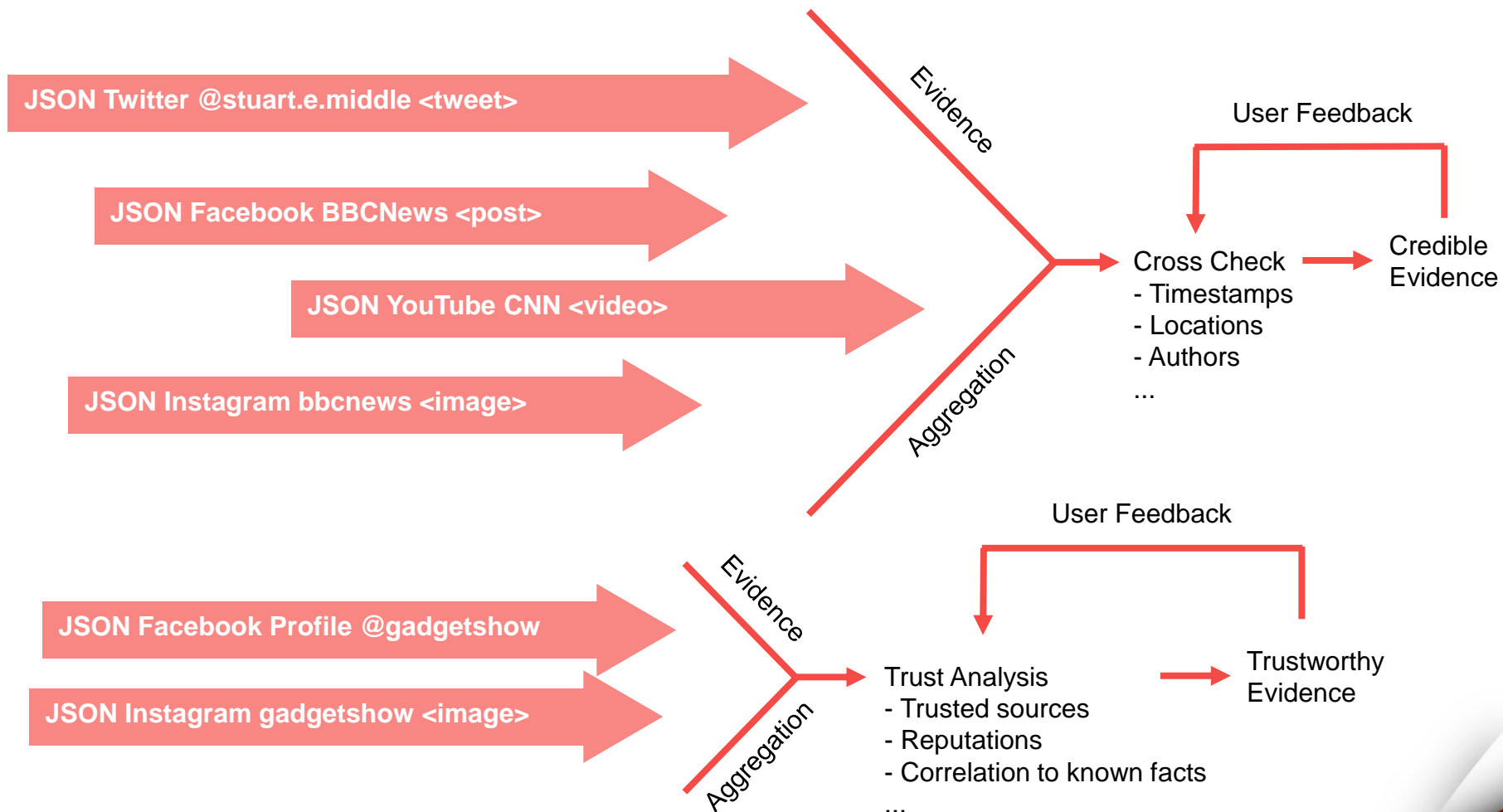
Real-Time Annotation of Social Media





Modality Analysis of Social Media Streams

Incremental Aggregation of Annotated Social Media





Geosemantics and Spatio-Temporal Grounding of Rumours

Geosemantic Classification of Evidence

- What is 'geosemantics'
 - Study of context of spatial data - in our case contextual text relating to mentioned locations
- Our Approach – Geosemantic feature classification
 - Geoparsing [Planet OpenStreetMap database] » **LOC (high precision, native language)**
 - Text + POS + LOC + training set » classifier » **context of how is location is talked about**
 - Classes » timeliness **past | future | present**, situatedness **insitu | remote**, confirmation **confirm | deny**
 - Eyewitness reports » **insitu**
 - Breaking content » **present**
 - Denial of rumours » **deny**
- State of the art – Geosemantic text analysis
 - Text + POS + training set » classifier » **event type**
 - Location text » NLP Grammar » **direction & distance**
 - e.g. trouble spotted 5 miles north of London
 - Location text » sentiment analysis » **good / bad opinion of text**
 - Resilience of approaches across event types and languages an issue

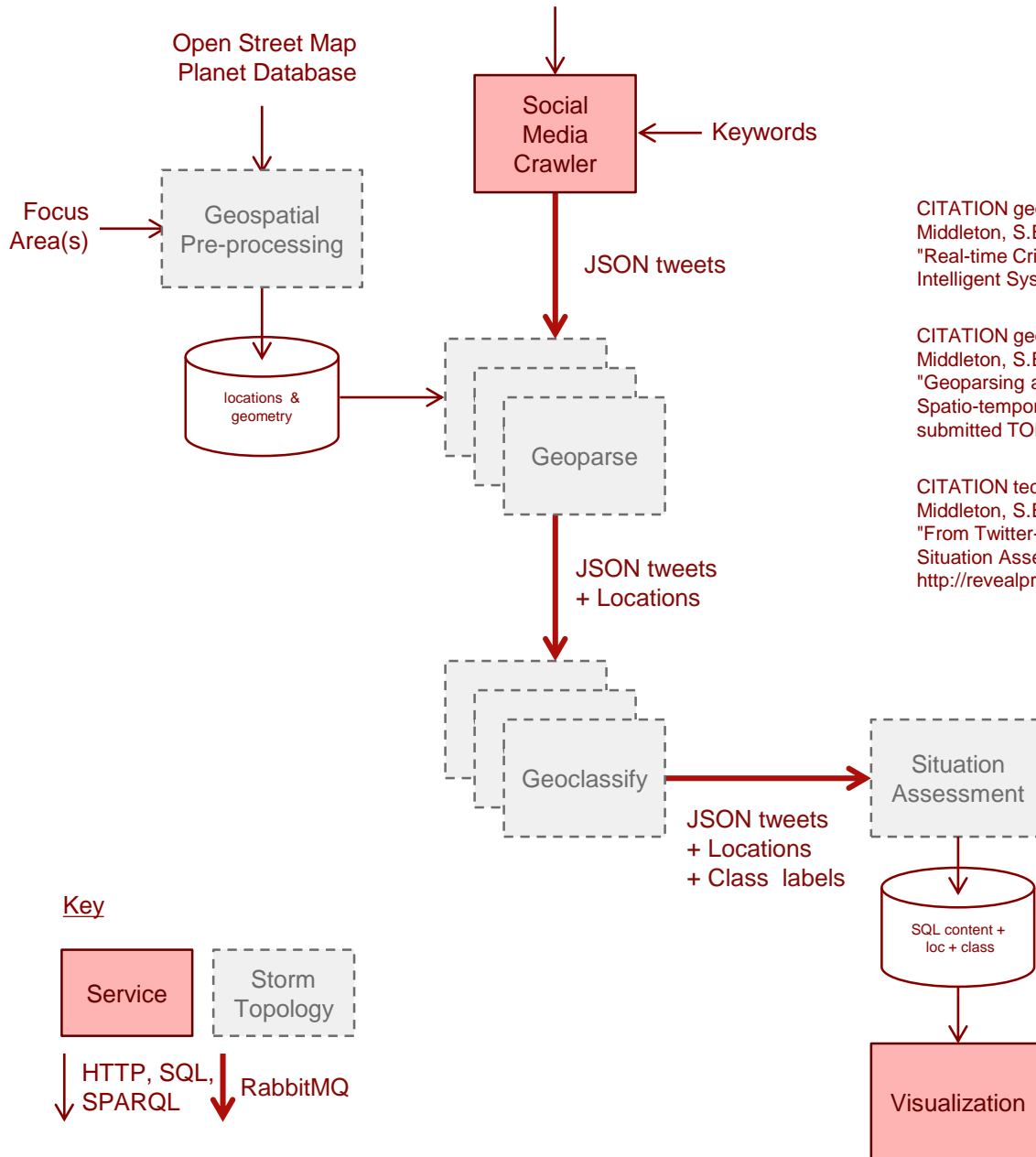


Geose

Twitter Search & Streaming API
You Tube Search API
Instagram Search API

Reveal

f Rumours

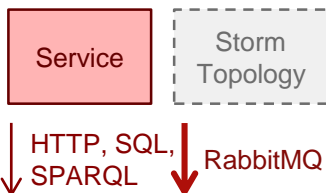


CITATION geoparsing
Middleton, S.E. Middleton, L. Modafferi, S.
"Real-time Crisis Mapping of Natural Disasters using Social Media"
Intelligent Systems, IEEE , vol.29, no.2, pp.9,17, Mar.-Apr. 2014

CITATION geosemantics
Middleton, S.E. Krivcovs, V.
"Geoparsing and Geosemantic Analysis of Social Media for
Spatio-temporal Grounding of Rumours during Breaking News"
submitted TOIS 2015

CITATION tech blog
Middleton, S.E.
"From Twitter-based Crisis Mapping to Large-scale Real-Time
Situation Assessment with Trust and Credibility Analysis", 2014
<http://revealproject.eu/>

Key





Geosemantics and Spatio-Temporal Grounding of Rumours

Case Study - NYSE flooding in 2012 (false rumour)

- Geosemantic filtering of evidence
 - 7000+ tweets in 5 minute analysis window
 - 114 ground truth tweets - WeatherChannel & CNN
 - Geosemantic filter reduced content volume by 95%
 - 100% ground truth recall for CONFIRM class
 - 85% ground truth recall for DENY class

The Weather Channel @weatherchannel Follow

"3 ft of water on floor of the NY Stock Exchange" via @TWCBryan #SuperStorm #Sandy #NYSE

RETWEETS: 789 FAVORITES: 36

6:42 PM - 29 Oct 2012

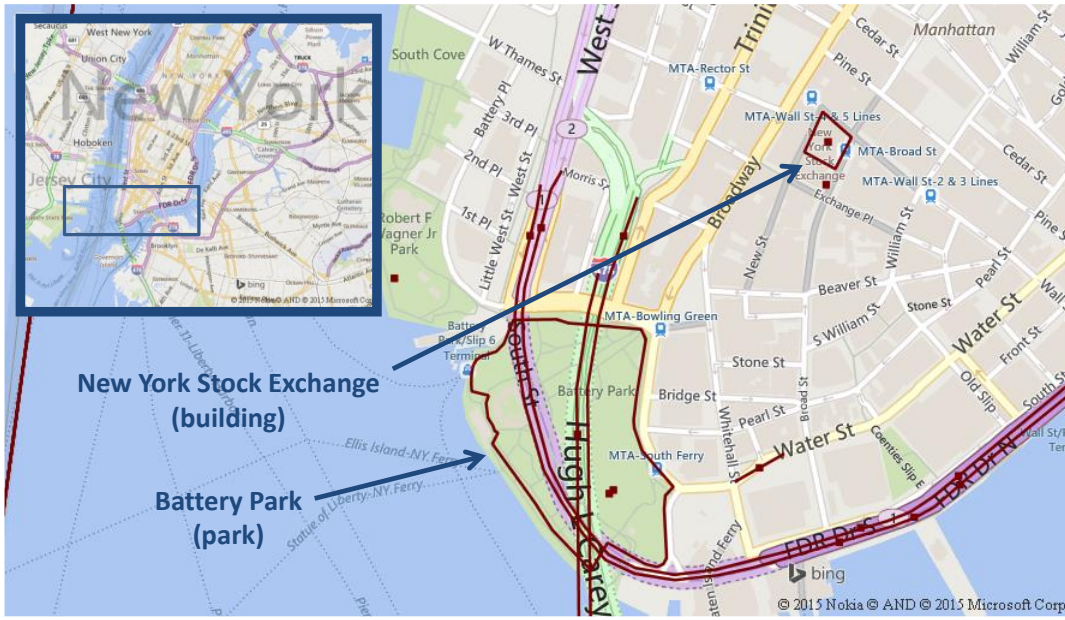
Reply to @weatherchannel @TWCBryan

CNY HIV Care Network @CNY_Connections · 29 Oct 2012
@weatherchannel @TWCBryan 3ft or 3 inches???? Any water on the floor of the NYSE would be horrible that might mean all lower manhattan

Jake Perganson @pergy33 · 29 Oct 2012
@weatherchannel @twcbryan @weatherchannel: "3 ft of water on floor of the NY Stock Exchange" via @TWCBryan #SuperStorm #Sandy #NYSE. #false

Snowman @cooljoe813 · 29 Oct 2012
@weatherchannel

Hillary Lake, KATU @hillarylake · 29 Oct 2012
@weatherchannel @twcbryan on TV your CNBC reporter Scott somebody said you all have NOT confirmed reports of NYSE flooding. #whichisit?

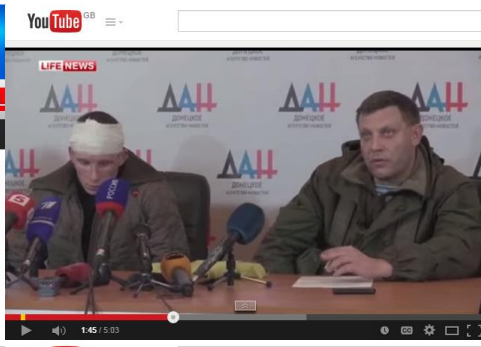
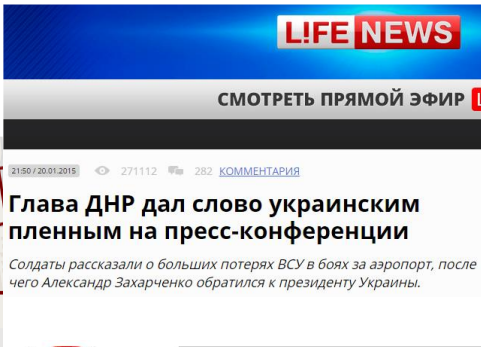
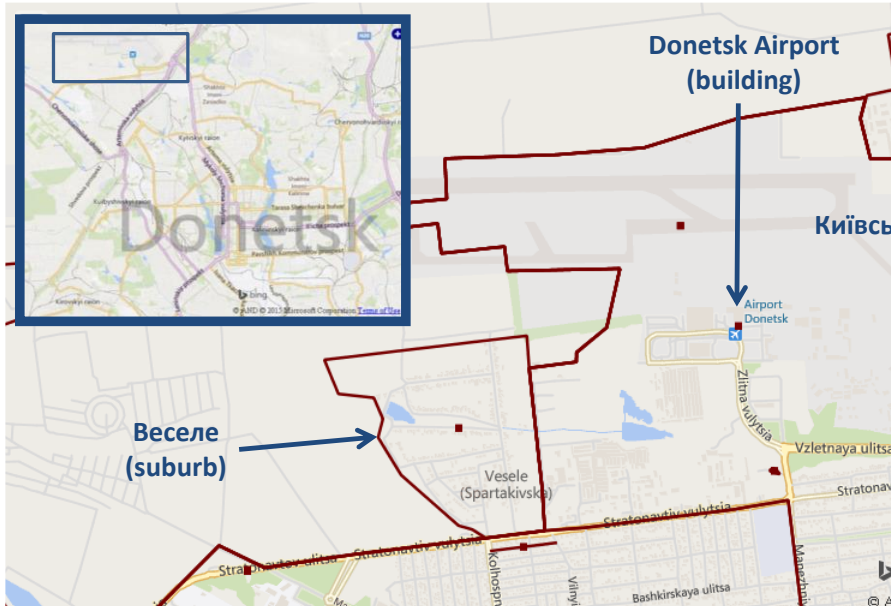




Geosemantics and Spatio-Temporal Grounding of Rumours

Case Study - Donetsk Airport 2015 (conflicting claims)

- Spatio-temporal mapping of evidence
 - 300,000 Tweets, YouTube & Instagram posts over 24 hours of Ukraine Crisis for 20th Jan 2015
 - 4 ground truth YouTube videos used by LifeNews reports on 20th Jan 2015
 - Focus: Donesk Airport cluster
 - Ground truth URI's ranked 10,14 & 28 out of 30





Knowledge-based Approach to Trust and Credibility Modelling

Definitions

- What is 'credibility' and 'trust' ?
 - Trust and credibility are not well defined – below is our interpretation
 - **Credibility** - consistency with other content (e.g. similar reports) and contextual information (e.g. local geography)
 - **Trust** - subjective assessment of likelihood of content being false

- **A credible news report might still be false!**



Knowledge-based Approach to Trust and Credibility Modelling

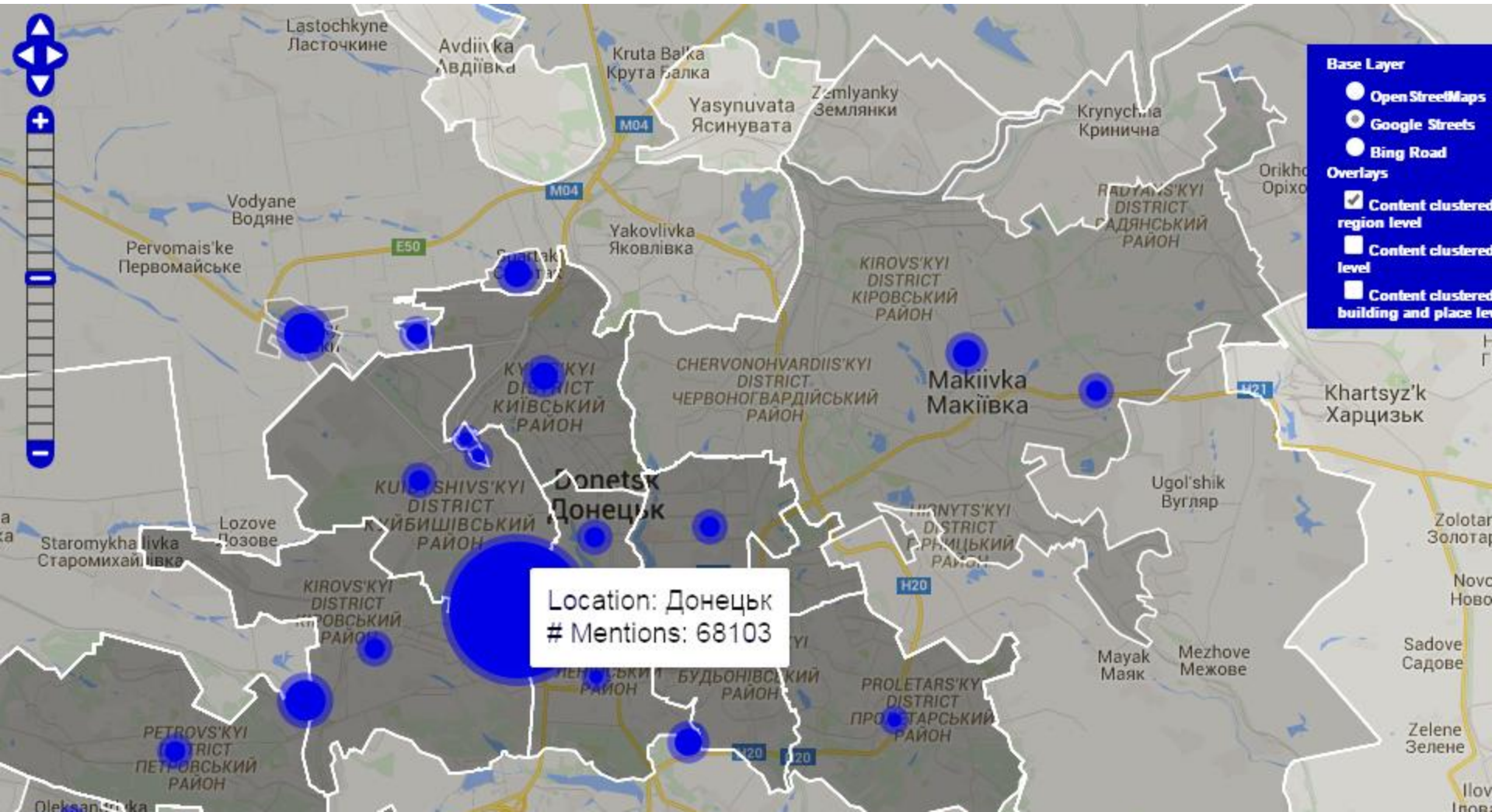
Approach

- Our Approach – Knowledge-based Trust Modelling
 - Journalist already have personal sets of trusted sources they have come to rely upon
 - Knowledge-based approach
 - Journalist assert **a-priori known facts** (e.g. trusted sources, known locations)
 - Evidence from streams asserted incrementally into a **triple store** (i.e. usekm + owlim)
 - **Simple inference » classify evidence** » interactive exploration of evidence with journalist
 - OWL classes & individuals, owl:Restriction, owl:intersectionOf, SPARQL, GeoSPARQL ...
 - **Not a black box** - End users explore the evidence and we help them make a verification decision
 - Scalable approach able to represent different viewpoints of Journalists
- State of the Art – Trust and Credibility Modelling
 - Unsupervised learning (e.g. Bayesian Network, Damper Shafer) » **trust prediction without explanation**
 - Supervised reputation models » **trust prediction with explanation**
 - Heuristics & activity metrics » **trust prediction with explanation**



Knowledge-based Approach to Trust and Credibility Modelling

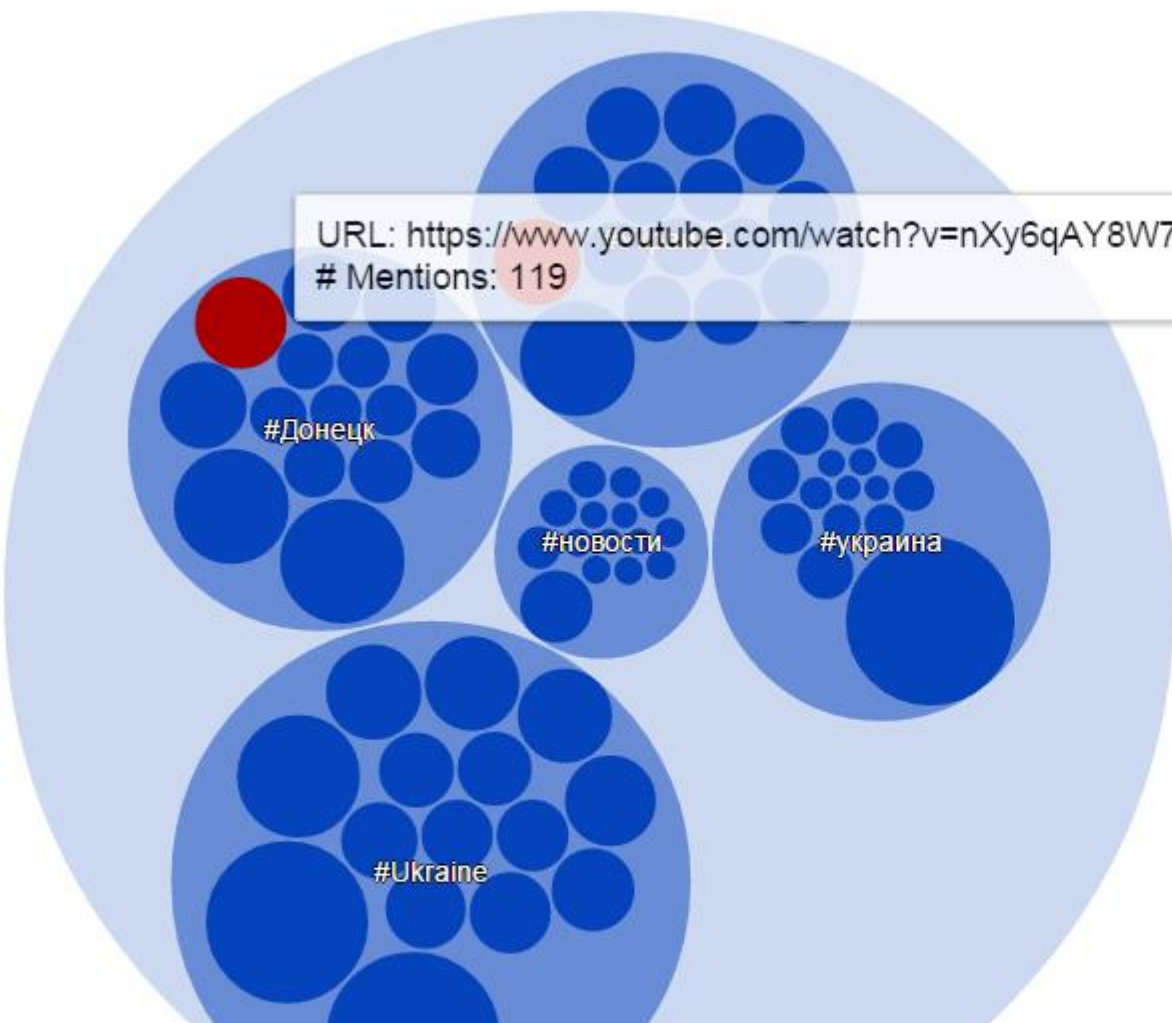
Early Results - Work in Progress





Knowledge-based Approach to Trust and Credibility Modelling

Early Results - Work in Progress



Tag Clusters (ranked by mentions) 5 ▾

URLs per Cluster (ranked by mentions) 15 ▾

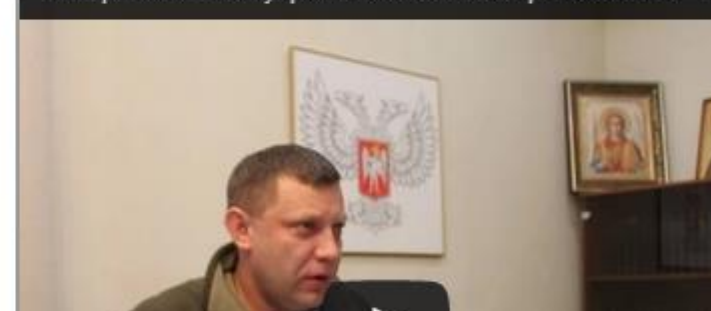
Legend

- Cluster of circles = All URLs mentioned with this TAG
- Circle = URL mentioned by a content item

Detailed URL Info

 URL: <https://www.youtube.com/watch?v=nXy6qAY8W70>
 Mentions count: 119
 URL Snippet:

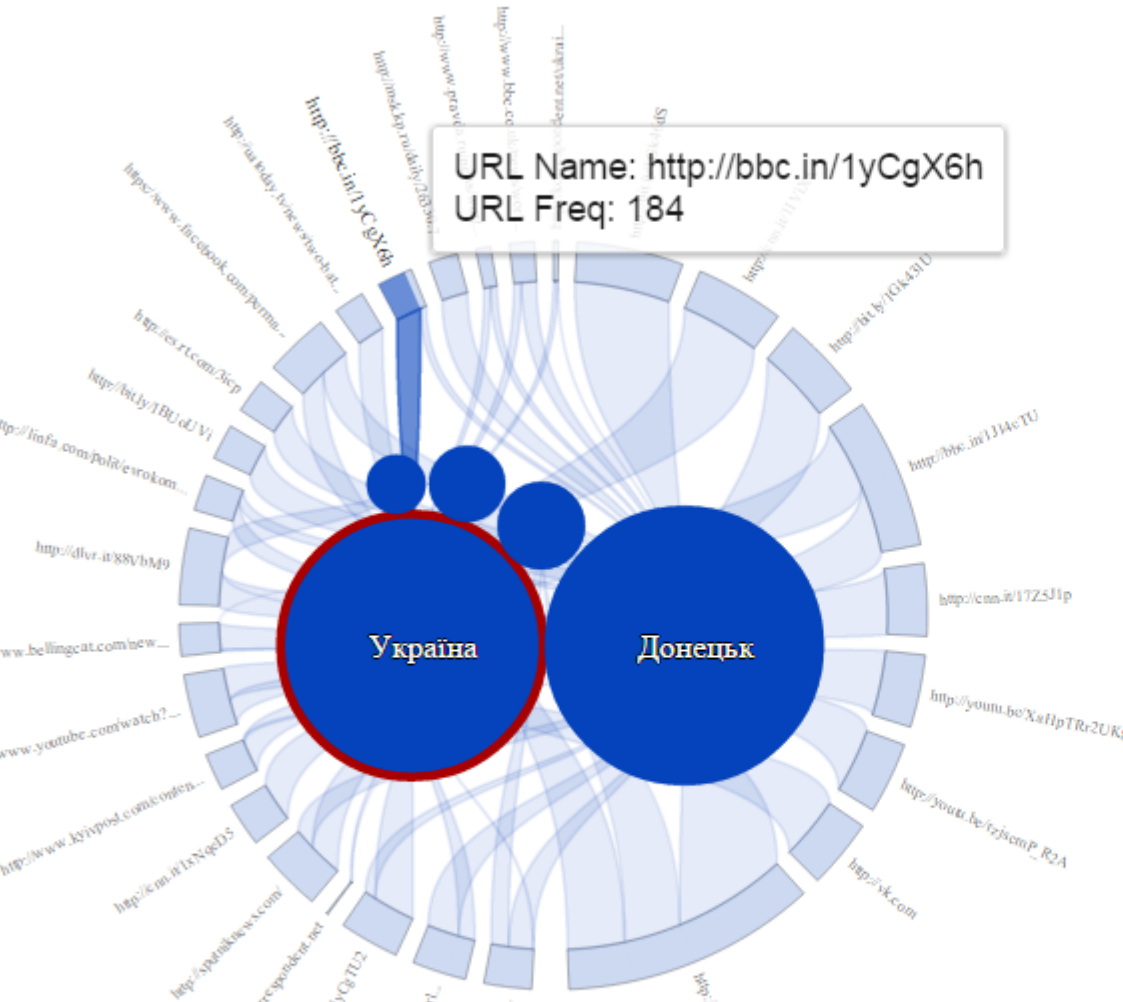
Захарченко: "Все украинские попытки провалились"





Knowledge-based Approach to Trust and Credibility Modelling

Early Results - Work in Progress



Top N URLs ranked by # mentions

Website filter for URLs

Legend

- Circle = Location
- Chord = URL

Detailed URL Info

URL: <http://bbc.in/1yCgX6h>

Mentions: 184

URL Snippet:

BBC News Sport More

NEWS

World

Europe

Ukraine conflict: Battles rage



Future Work

Roadmap for REVEAL Trust and Credibility Analysis

- REVEAL project runs until Sept 2016
- Ethnographic studies with Journalists
 - Crawl content in parallel to journalists searching User Generated Content (UGC)
 - Record ground truth by observing Journalists verifying news for real & explaining decisions
 - Empirical analysis - compare automated decisions with ground truth
- MediaEval 2015 verification challenge
 - Competition verifying images using a common Twitter dataset (10 different news events)
- Users trials 2015 - 2016

Many thanks for your attention!

Any questions?

Stuart E. Middleton

University of Southampton IT Innovation Centre

email: sem@it-innovation.soton.ac.uk

web: www.it-innovation.soton.ac.uk

twitter: [@stuart_e_middle](https://twitter.com/stuart_e_middle), [@IT_Innov](https://twitter.com/IT_Innov), [@RevealEU](https://twitter.com/RevealEU)

