Defence Research and Development Canada | Recherche et développement pour la défense Canada

Twitter as a Source for Actionable Intelligence

ICCRTS 18th – 19 July 2013

Bruce Forrester, PhD
Defence Scientist – DRDC Valcartier

DRDC | RDDC

Canada

Social Media as a Source

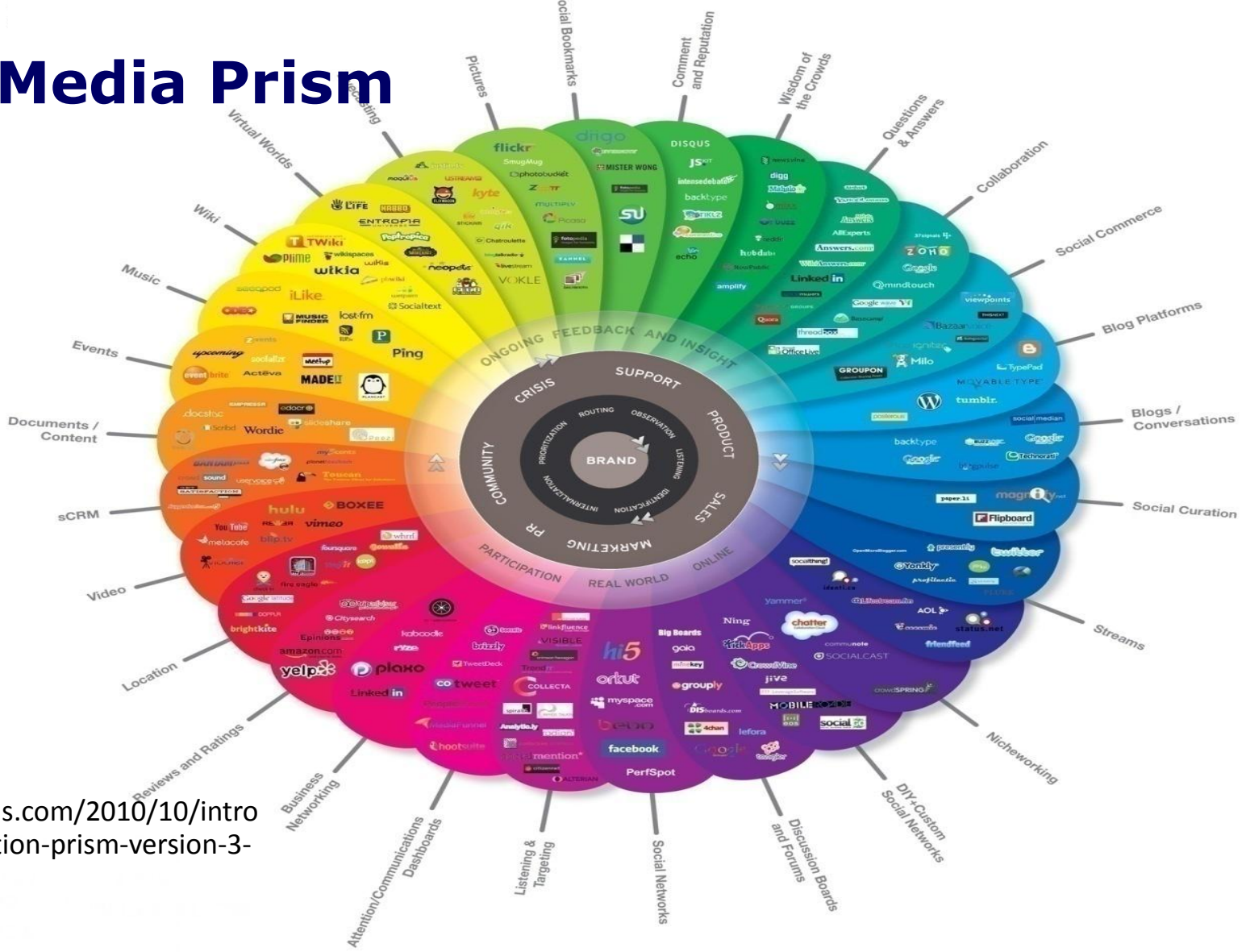
- Lots of post-analysis
- No real time analysis
- Huge increase in use by traditional media
- Arab Sprig was first concrete example of mass use by public for revolution

First conclusive study on the influence of Social Media

- social media played a central role in shaping political debates in the Arab spring.
- a spike in online revolutionary conversations often preceded major events on the ground.
- social media helped spread democratic ideas across international borders.


OMG!
I JUST GOT
BORN!

Social Media Prism



http://www.briansolis.com/2010/10/introducing-the-conversation-prism-version-3-0/

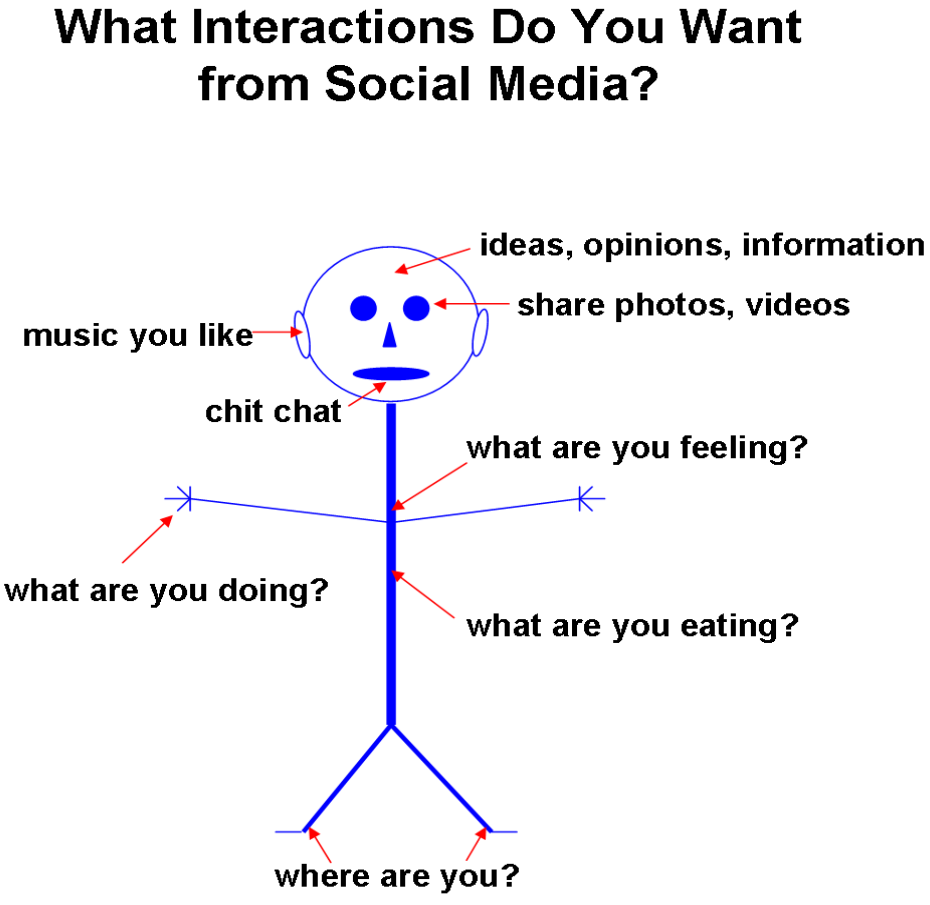
Intelligence Exploitation of Social Media

Example of an area for concentration

Phenomenon	INT Product	Military/INT impact
What is happening in remote areas (where there are few other sources avail)?	➢ Response to targeted RFI ➢ Alert service ➢ Threat assessment ➢ Information bulletin	➢ Current up-to-the-minute SA of a particular area ➢ Enables operational planning ➢ Tactical threat assessment

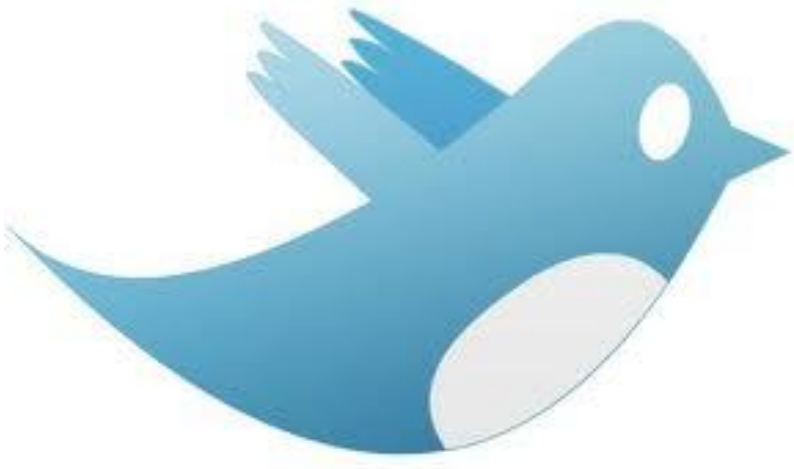
The Studies

- Twitter data
- Philo and Metho issues
- User traits
- Influence
- Roles
- Validity
- Content
- Prediction



Twitter Studies

- Over 40 studies researched
- Large majority are business oriented
- Authors come from a mathematical or computer science background



- However, do provide relevant methods and tools for intelligence purposes

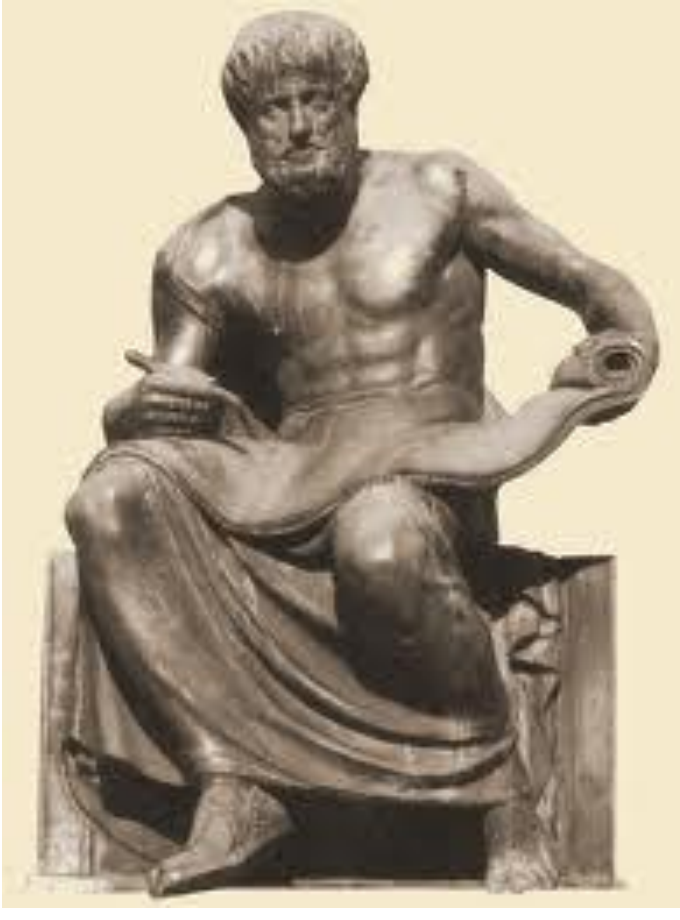
What makes Twitter different?

Traditional OSINT Sources	Twitter as a Sources
Well-defined use of acronyms	High use of text and chat acronyms
Use of most common language dialect	Local dialects and special use of words that have different generalized meanings
Edited	Not-edited
Written by professional authors	Written by anyone and everyone
Use of proper grammar, spelling and punctuation	Anything goes
Minimal use of sarcasm, street language, profanity	Anything goes

Philosophical issues

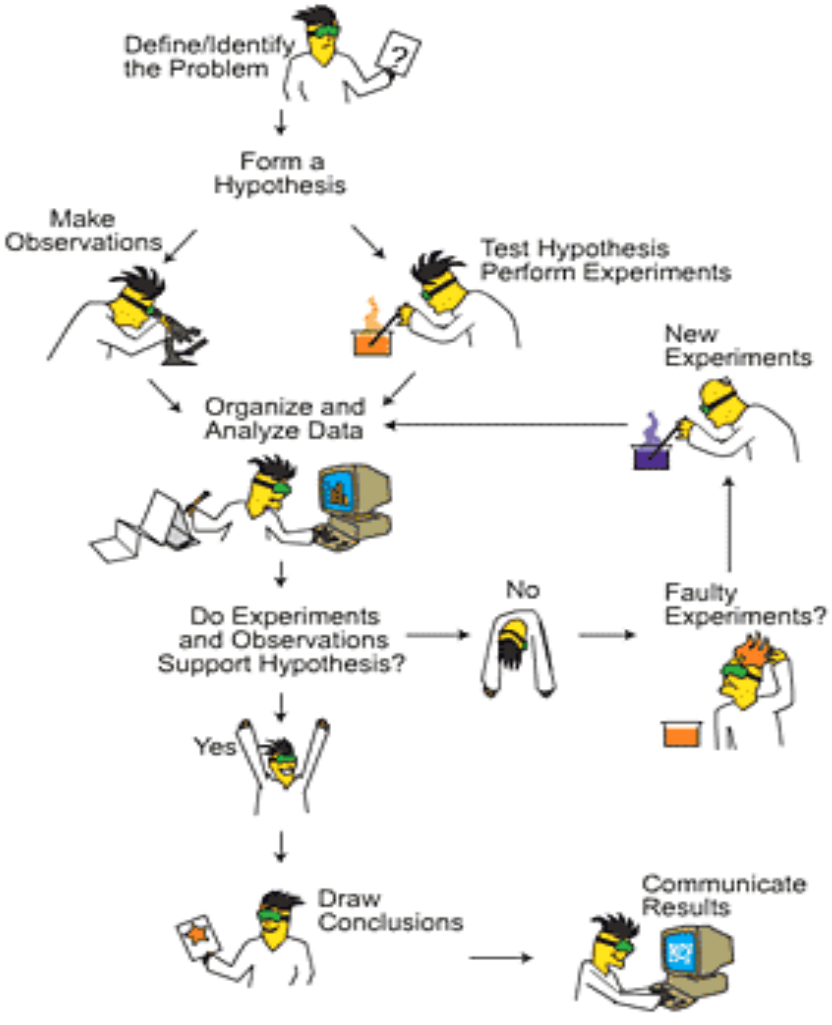
- A change in how research is conducted:

1. Statistical significance without meaning
2. Data cleansing – removing outliers
3. Danger lies in aggregation



Methodological issues

- Twitter data collection
  - REST
  - Search
  - Streaming
- Target population
- No standard methods
- Sites are in constant Beta
- Number of users varies



Studies of users' traits

- information sharing, information seeking, and friendship-wise relationships
- people with similar life outlooks and interests tend to "hang out" together, to talk about headline news and respond to fresh news

Studies on influence

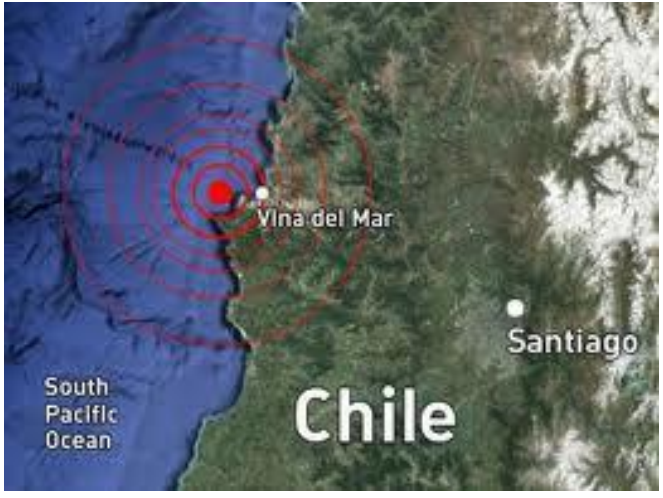
- **Indegree** influence - the size of the user's audience.
- **Retweet** - how well a user produces content that has pass along value.
- **Mention** - the ability of a user to engage others in a conversation.

Communicator roles

- Idea starter - An individual who starts a conversation meme
- Amplifier - An individual who collates multiple thoughts and shares ideas and opinions
- Curator - An individual who use a broader context to define ideas
- Commentator - An individual who detail and refine ideas
- Viewer - An individual who takes passive interest in the conversation

Studies on Validity of content

- Retweet is an indicator of value
- reliability of tweeted information Chilean earthquake (Feb 27, 2012) comparing rumors and news
- collaborative filtering affect differentiated news from rumors

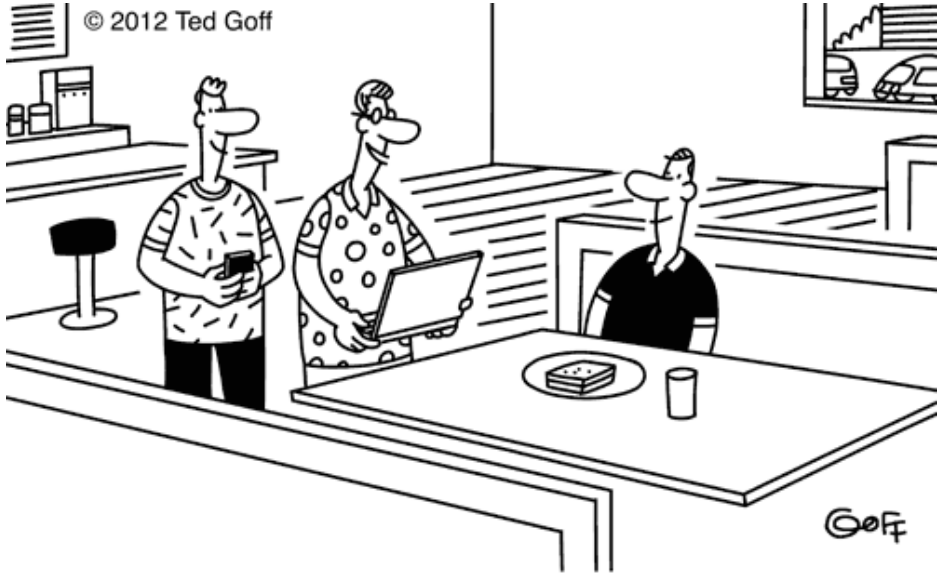


Studies on Prediction

- Much research on box office revenues and elections
- Prediction limited to human related events
- Social media had some effect but not accurate due to current methods
- Short life cycle is more accurate

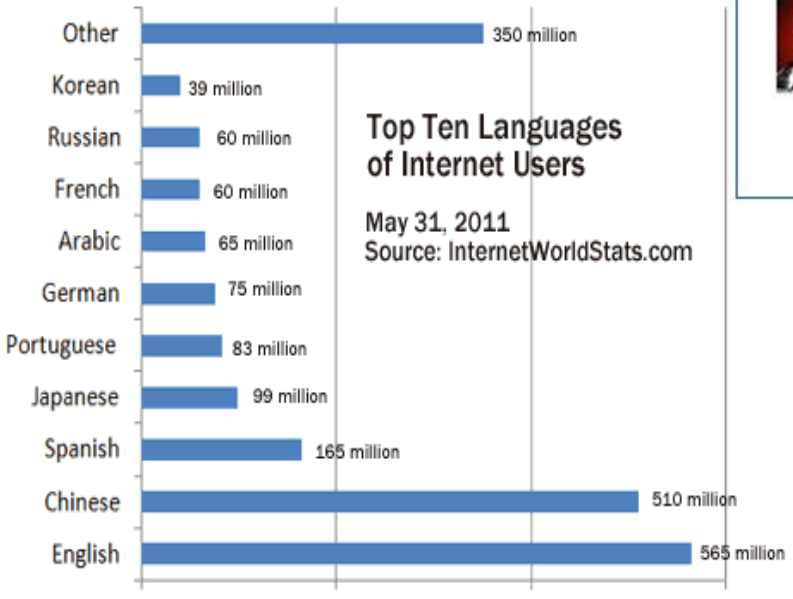
Studies on Prediction 2

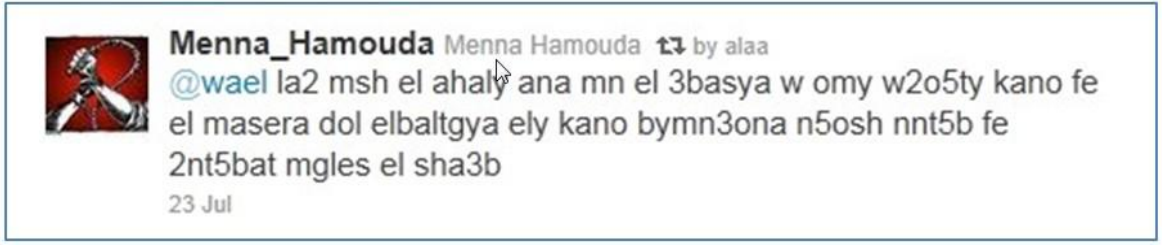
- All post-hoc studies on elections – no prediction
- Methodological problems
- Better solutions now
- Mood correlated to DJIA but shifted 3-4 day later



"Twitter and Facebook can't predict the election, but they did predict what you're going to have for lunch: a tuna salad sandwich. You're having the wrong sandwich."

Studies Concerning Language





ur = You are

l8tr = Later

Studies of Other Indicators

- Volume
- Temporal change
- Sentiment
- Geolocation
- Deception

Conclusions

- There is a large base of users from which data and information can be acquired.
- Users can be identified for the most part to a fair degree of granularity.
- We can determine the specific roles played by users within a topic.
- There are methods that can be used to determine the most influential users.
- We can determine where an idea started, and how and by whom it propagated throughout the network.
- Ordinary users can have a major influence on the spread of news and information. These individuals can easily be identified.

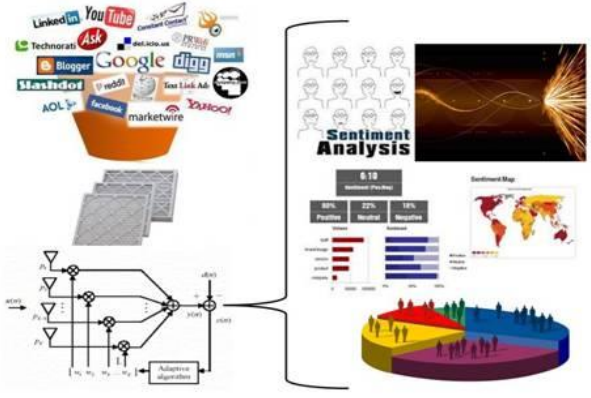
Conclusions 2

- There is a wealth of information that can be gleamed from tweets through the included references to links and URLs.
- The results of prediction using twitter are varied but promising and are likely to improve as methods improve.
- A rough degree of sentiment can be calculated.
- Language and culture are and will continue to be significant barriers to understanding content on a deep level.
- A geolocation can be found or inferred in many cases.
- There are several methods that could be used to help detect deception and rumors.

Implications

- Early stages of research
- Need cultural framework
- Promising for INT work
- ARP work to follow

Monitoring, Analysis of Social Media for Behaviour Estimation



Partners: CDI / TBD, Army G2 / TBD, SOFCOM / TBD, Navy / TBD, NCR, NLP, Thales, U of Montreal – RALI Lab

Project Manager: Bruce Forrester

Delivery by: DRDC Valcartier, DRDC Toronto

Linkages: NATO ET.BY, CRTI, Square Dance

Industry: NCERC Industry Research Chair

Start – End: Apr 2013 – Mar 2018

	FTE	Contract	Total Cost
DRDC	1380	1200	3580
CF	TBD	500	500

Objective: To understand:
✓What are the intelligence questions that can be answered through analysis of social media?
✓What kinds of indicators have meaning within this new context?
✓How does one handle translation, interpretation of cultural idiosyncrasies, and the use of natural language contained in social media?
✓What automated tools and algorithms can be used for monitoring, analysing and behaviour estimation with the big data of social media?

Science and Technology:
Human Science: Socio-cultural modelling and frameworks; sentiment, emotional and trend analysis; human domain analysis tools; language translation.
Computer Science: Knowledge-based systems technologies; machine learning; estimation analytics; knowledge discovery and data search; Big Data; context-dependent, case-based and self-improving automated reasoning.

Output and Deliverables: The deliverables will consist of interim reports and publications, proof-of-concept prototype, algorithms, and final report.
✓Intelligence Community SM requirements report
✓SOTA report for monitoring, filtering and data capture
✓SOTA report for advanced analysis of SM data/information
✓SOTA report for estimation using SM data/information
✓Proof-of-concept prototype (final version including all algorithms)

Desired Outcomes: Understand strategic, operational and tactical intelligence uses of social media (SM) in order to contribute to the DND OSINT Capability. Development of prototypes in the form of automated tools and algorithms for monitoring, capturing, and analysis of data collected from SM for Intelligence production and behaviour estimation.

bruce.forrester@drdc-rddc.gc.ca