U.S. Army Research, Development and Engineering Command

Accelerating Exploitation of Low-grade Intelligence Through Semantic Text Processing of Social Media

ARL

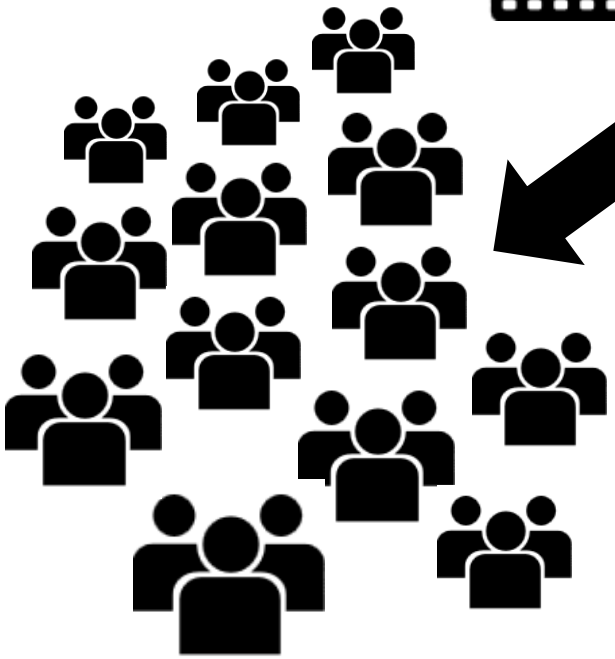
TECHNOLOGY DRIVEN. WARFIGHTER FOCUSED.

Dr. Sue E. Kase & John Dumer
U.S. Army Research Laboratory
Computational Information Sciences Directorate
Tactical Information Fusion Branch
18th ICCRTS, June 2013

SOCIAL MEDIA CONTENT

Increasing usage of mobile devices

Facilitates information sharing from anywhere at any time



Social Media

Unprecedented volumes of real-time information contributed by large communities of social media users

Is social media potentially valuable to the intelligence community?

How can social media be searched for information relevant to national security?

WHAT IF NO TRADITIONAL SOURCES OF INTELLIGENCE AVAILABLE

Constrained resource allocation environment

HUMINT

Spot Reports

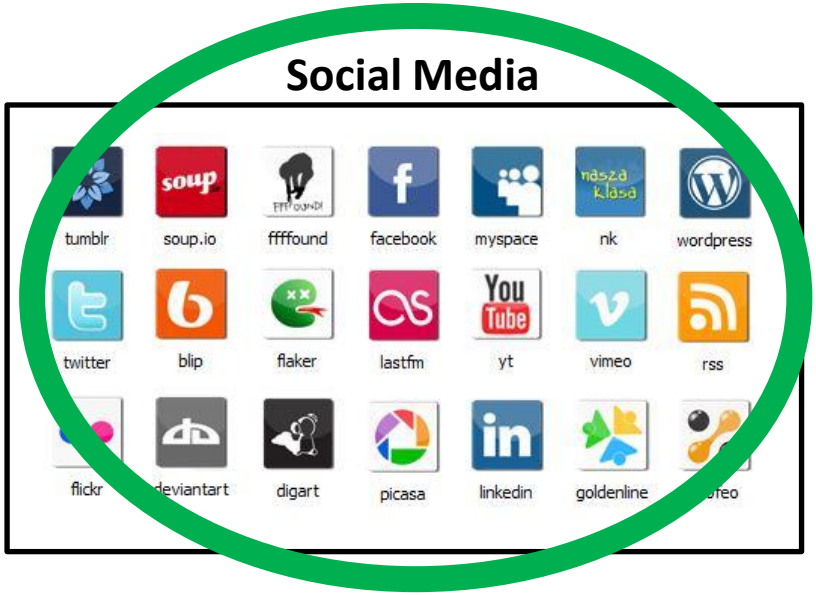
Tactical Reports

Challenges and risks with direct intelligence data collection

Intelligence collection relying on the eyes and ears of the general population

Citizen-authored information

Social Media



A source of 'low-grade' intelligence information if *search and extraction technologies* are developed

FRAME-BASED SEMANTIC SEARCH OF SOCIAL MEDIA

CHALLENGE

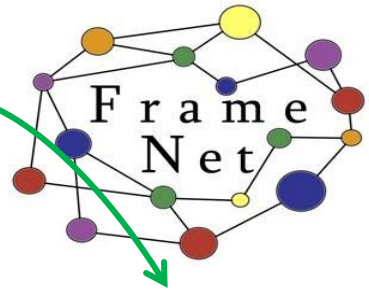
Social media content is dynamic, 'unstructured' containing hashtags, urls, emoticons, replies and mentions, and little useful information

SOLUTION

DECISIVE ANALYTICS

CONTOUR SHAPING KNOWLEDGE

Distill social media for relevant information using frame-based semantic modeling



Semantic search capability part of DAC CONTOUR suite

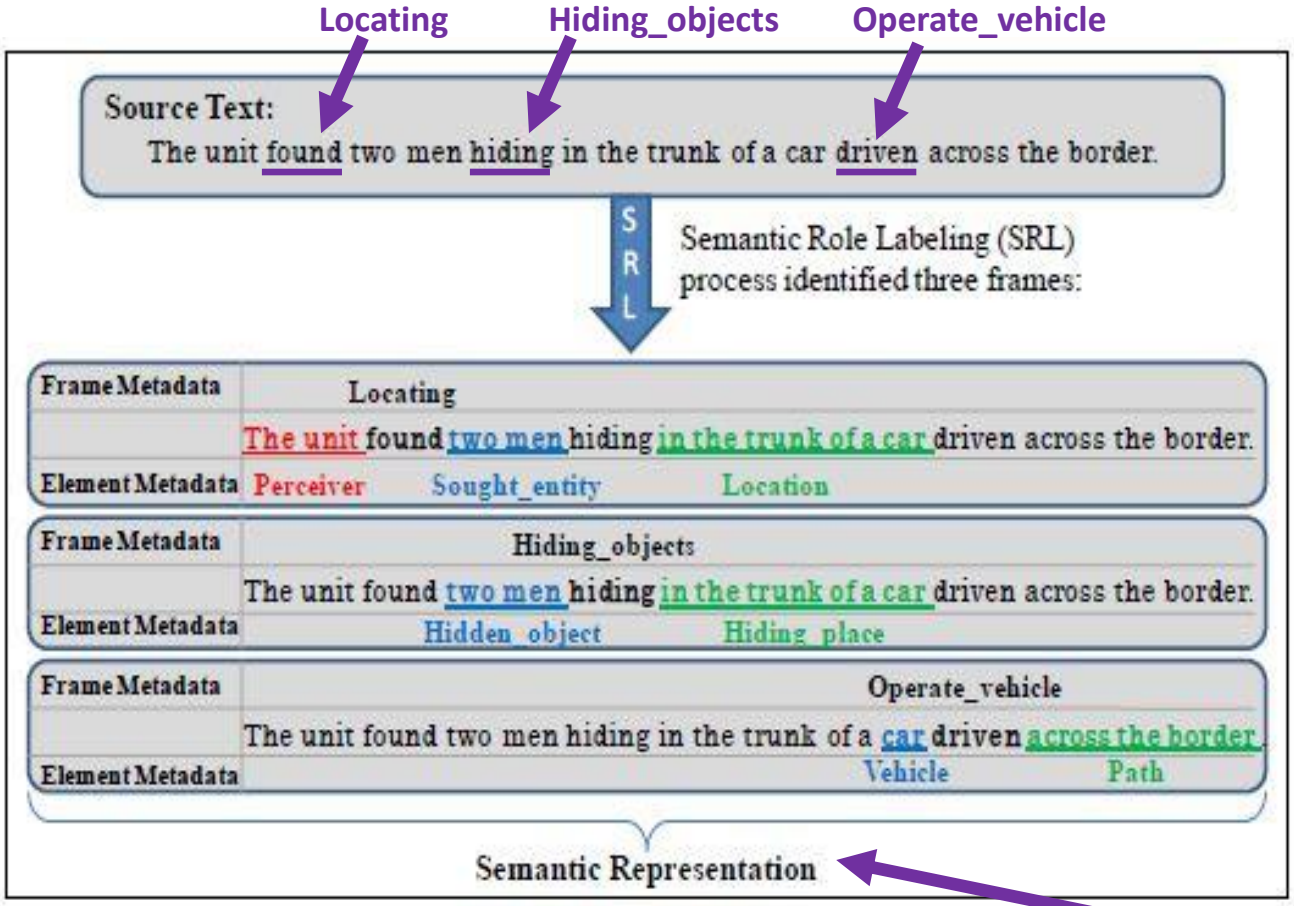
Import unstructured text from variety of sources, map to ontology of frames

Construct queries using frame (Concept) and frame element (Role) filtering widgets

FrameNet-based SRL

Frame Semantics theory (Fillmore & Colleagues, 1976)
Semantic Role Labeling (SRL) (Gildea & Jurafsky, 2002)

Berkeley FrameNet lexical database
https://framenet.icsi.berkeley.edu/fndrupal/



Frame: *a schematic representation of a situation involving various participants, props, and other conceptual roles*

3 frames in target sentence
- **Locating ('found')**
- **Hiding_objects ('hiding')**
- **Operate_vehicle ('driven')**

Locating Frame has Frame Elements
- **Perceiver**
- **Sought_entity**
- **Location**

Frame Element : *frame-specific defined semantic role that is the basic unit of a frame*

Comparison with WordNet and ontologies

Multiple annotated examples (~20) of each LU.
Examples taken from naturalistic corpora.
LU linked to semantic frame similar to thesaurus.
All combinatorial possibilities of the LU
Analysis proceeds frame by frame
Network of relations between frames

Builds a semantic model of the text

FrameNet-based
QUERY BUILDING INTERFACE

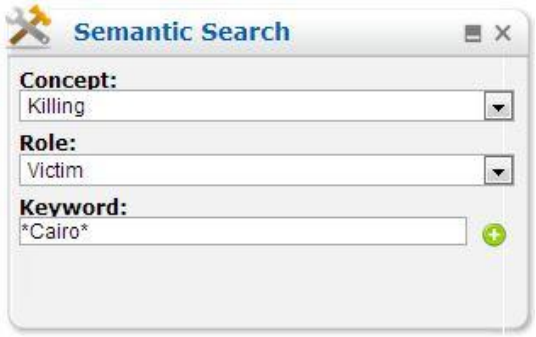
FrameNet-based SRL → **Semantic model of the text** → **Concept + Role filtering widgets**

Mapping between Frame Concepts + Frame Element Roles and the target text

Select Frame from dropdown menu in Concept widget

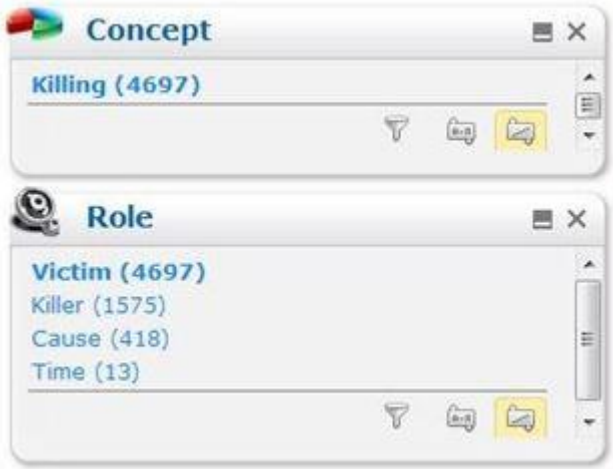
Then Select Frame Element from dropdown menu in Role widget

Alternative Example: ***Using the Semantic Search widget***



Killing Frame with Victim Frame Element selected Keyword *Cairo* entered for targeted search

Example: ***Killing Frame selected Frame Elements of Killing Frame displayed in Role widget***



Then Select Frame Element

USE CASE:
SEMANTIC SEARCH OF TWEETS

Traditional sources of intelligence information are not currently available

Use CONTOUR's FrameNet-based semantic search on social media

Demonstrate how CONTOUR's query interface can be used to identify low-grade intelligence information

DECISIVE ANALYTICS

CONTOUR
SHAPING KNOWLEDGE

Twitter: streaming real-time data, immense scale

Twitter users: act as social sensors, post in anticipation of events, report breaking news

Twitter messages: little textual information, informally written, embedded non-text characters

Tweet content types: personal updates, calls for participation, warnings about threatening situations

Tweets from Arab Spring protest

Data set: 7.3 million tweets
Collected between Feb 01 – 19, 2011

NS CTA INARC collaboration:
Blender Cross-source Information Extraction Laboratory
City University of New York (CUNY)
Dr. Heng Ji

SEMANTIC SEARCH OF TWEETS

[DEMO or MOVIE]

An analyst receives an external report about a protest occurring in Egypt possibly near a Forward Operating Base (FOB)

The analyst's chain of command needs to know what is going on in that area; there is no time to collect intel

A stream of tweets is captured and imported into CONTOUR's semantic search system

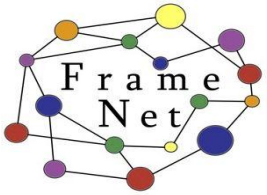
ACKNOWLEDGMENTS

SBIR Phase II
Decisive Analytics Corporation

http://www.dac.us/Capabilities/Intelligent_Products

ARL NS CTA
BLENDER Lab, CUNY
Dr. Heng Ji

http://nlp.cs.qc.cuny.edu/



FrameNet Project

https://framenet.icsi.berkeley.edu/fndrupal/

Utilize social media to gather information about events occurring or events that may occur within a geographic region such as an area of operations

Develop new text analytics technologies such as CONTOUR's Frame-based semantic search to process social media content that is unstructured and primarily non-informative

Use Case demonstrates how an analyst can search a large stream of tweets with frame-based queries constructed from Concepts, Roles, and Keywords

Collecting low-grade information from social media - when added together - produces actionable intelligence such as determining how an event may affect missions