

A SEMANTIC WIKI ALERTING ENVIRONMENT INCORPORATING CREDIBILITY AND RELIABILITY EVALUATION

STIDS 2010

Brian Ulicny, Chris Matheus^a Mieczyslaw M. (Mitch) Kokar^{a,b} ^aVIStology, Inc. ^{a,b}Northeastern University

WHAT ARE WE TRYING TO DO?

- Provide an up-to-date understanding of current and potential threat by identifying and characterizing the entities involved.
- These include the individuals, groups, locations, activities and events associated with the threat and their interrelationships.
- The threat we are modeling is transnational street gangs operating in the US.
- The current state of the threat is modeled by means of an automatically updated Semantic Wiki representing the state of the group.
- Alerts are automatically sent to relevant parties when the state of the threat changes in significant ways.

HOW ARE GROUPS TRACKED TODAY?

• Civilian/Open Source Technology:

- Alerts:
 - Google News Alerts,
 - Twitter monitors,
 - Cayuga Event Processing (Cornell), RSS/Atom Feeds
- Manual (Semantic) Wikis:
 - MediaWiki (Wikipedia);
 - Semantic MediaWiki
- Military Technology:
 - Alerts:
 - CIDNE, Military Chat
 - Wiki(-like):
 - Intellipedia, TiGR

A WIDE PRESENCE IS SUSPECTED

Because MS-13's organization is so fluid in the United States and members move frequently between here and Central America, no one has a firm grasp on the gang's membership numbers. Law enforcement agencies estimate there are 8,000 to 10,000 members operating in 33 states. Where they are believed to operate:



Significant presence

Lesser degrees of criminal activity



TROY OXFORD/Staff Artist

CONTRAST WITH EXISTING SYSTEM

for where was elvis born?

Map

-

bing

Road

×

Aeria

MEXICO

This Month

Hybrid

Gulf of Mexico

~ Headlines Headline Place Date -Castro a success all along Great American Ball Park 5/8/2010 5/8/2010 Cinematic Titanic: Santa Claus Conquers The Martial Bethlehem AFN honors Elvis-loving Fairbanks minister Anna Fra 1205 Denali Way Fairbank 5/8/2010 Morris Pert: musician and compose 5/8/2010 Arbroath 5/8/2010 Interview: John Flansburgh of They Might Be Giants (Brooklyn, San Francisco Scotlands guiding Shand can be seen at every music Prestwick Airport, Hebrides 5/7/2010 Letters: So glad we can feel safe Memphis.Tennessee.Collic 5/7/2010 5/7/2010 Hues of Deep Purple Perth, Australia Explore Mississippi's Affordable Adventures Clarksdale, Mississippi, Indi 5/6/2010 Can native son Uncle Sam raise a little cash for town' Arlington 5/6/2010 Katrina and all that jazz New Orleans 5/6/2010 Scotland's guiding Shand can be seen at every music Prestwick Airport, Hebrides 5/6/2010 Nobunny does it better Tucson, Arizona 5/6/2010 Selling Uncle Sam Arlington 5/6/2010 12th Street and Delaware The Features 5/5/2010 5/5/2010 Couple put locomotion in local festivals Austin The Features: Reviews of Features at the Maryland F 12th Street and Delaware . 5/5/2010 Find out what these celebrities were called before the England, Tallulah 5/4/2010 Interview: Willie Nelson Austin, Texas, Buckingham 5/4/2010 LATEST NEWS: Ivory Joe Hunter's grave will receive Port Arthur, Kirbyville 5/3/2010 Locals take Beach Blanket stage Beach Blanket Babylon, Sa 5/3/2010 Costello wraps up season, perhaps even series itself Apollo Theater 5/3/2010 I4 4 | ▶ | æ Displaying articles 1 - 100 of 381

Search Enter a place...

FRANCE SPAN Explore Mississippi's Affordable MOROCCO Query identifies documents that contain "elvis" and "born" and a VENEZUELA location. Answers literally all over the BR map. Consensus BOLIVIA answer not obvious from location clusters. Documents are recent news articles.

Start Over

Search

Mississippi

Adventures

CUBA

HONDURAS Caribbean See

PANAMA

ECUADOR

PERU

HAIT

COLOMBIA

WHAT IS NEW IN OUR APPROACH?

- Automatic population of Semantic Wiki
 - Using Entity Extraction and Formal Reasoning
- Cross-document alert generation based on semantic knowledge base
- Generation of alerts based on dynamically updated model of group (not just watchlist)
 - E.g. Alert me if there is a 10% increase in arrests of gang members in a specific city, week over week.
- Information Evaluation (per STANAG 2022)
- A successful implementation will allow analysts to interact with a dynamically updated model and receive alerts when significant changes occur.

ANTICIPATED BENEFITS

- Timeliness of alerts to increase operator's productivity
- Automatic analysis of large quantities of data (much redundant) to improve operator's awareness
- Semantically normalized information (entities/ relations/events) to improve quality of operator's decisions, relevance reasoning
- Focused, customizable filtering/monitoring to make the approach useful for various types of operations
- Evaluation of information for reliability/credibility to provide higher operator's trust in the system
- Visual interactive information exploration (maps, timelines/tracks, charts) to provide system usability

SEMANTIC WIKI ALERTING ENVIRONMENT (SWAE) OVERVIEW



10/28/10 VIStology, STIDS 2010, GMU

PROBLEM DOMAIN: STREET GANGS

Street gangs are analogous to terrorist organizations

loose organizations with hierarchical membership

uniformed military

- narcotics operations are often used for funding
- the threat is organized in dispersed cells
- the local population must often be won over to provide information against the threat
- Wealth of dynamic, online information for various sources (Twitter, MySpace, news sources, etc.)
- Mara Salvatrucha (MS-13):
 - Started in El Salvador in the 1980s US: >15K members, >115 "cliques", 33+ states
 - Foreign: Canada, Guatemala, Honduras, Mexico and El Salvador
 - -----



SWAE PHASE I PROTOTYPE



DATA SOURCES



topix

28 gang members arrested in New Orleans area 🗗

Twenty-eight individuals with ties to the MS-13 and Latin Kings transnational violent street gangs are facing criminal charges or deportation following a three-day joint multi-agency enforcement action in the New Orleans area. Comment?

28 Immigrants With Ties To Gangs Arrest ed 🖾

Twenty-eight people with ties to the MS-13 and Latin Kings gangs were arrest ed and several will be deported after a three-day operation, immigration officials said.



RSS 2 RDF

```
<?xml version="1.0" encoding="utf-8"?>
```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:content="http://purl.org/rss/1.0/ modules/content/" xmlns:r="http://backend.userland.com/rss2" xmlns="http:// purl.org/rss/1.0/">

<channel rdf:about="http://www.topix.com/search/article?q=%22ms-13%22+OR+%22mara
+salvatrucha%22&x=0&y=0">

<dc:title>Search for ""ms-13" OR "mara salvatrucha"" </dc:title>

<topix:rsslink xmlns:topix="http://www.topix.com/partners/rsscomment/" xmlns:georss="http://www.georss.org/georss">http://www.topix.com/rss/search/ article.xml?q=%22ms-13%22+OR+%22mara+salvatrucha%22&x=0&y=0</topix:rsslink>

<dc:description>News continually updated from thousands of sources across the
web</dc:description>

<items>

<rdf:Seq>

<rdf:li rdf:resource="http://www.connectionnewspapers.com/article.asp? article=341435&paper=59&cat=104"/>

<rdf:li rdf:resource="http://www.mysanantonio.com/news/local_news/ suspected_ms-13_gangster_busted_near_natalia_99664094.html"/>

<rdf:li rdf:resource="http://www.charlotteobserver.com/2010/07/28/1586317/ ms-13-gang-member-sentenced-to.html"/>

. . .

ENTITY/RELATION EXTRACTION

• Information to extract from text:

- Information source
- URLs of cited information
- Locations of events (where)
- Times of events (when)
- Types of events (what)
- Participants in events (who)
- Extraction Software used:
 - OpenCalais
 - BaseVISor (RDF matching, Regex)
 - (UIMA)

OPENCALAIS PROCESSOR

<rdf:Description rdf:about=

http://d.opencalais.com/dochash-1/6d3695ba-2142-3679-b8db-5e206844c924/Instance/40>

<oc:detection> <![CDATA[[in a news release.</p> The arrest follows]the May 28 arrest in Santa Cruz of
[X] [, another [Gang Y], or [VariantName V], member]]]></biddetection>

<oc:docId rdf:resource="http://d.opencalais.com/dochash-1/6d3695ba-2142-3679-b8db-5e206844c924"/>

<oc:exact>the May 28 arrest in Santa Cruz of [X]</br>

<oc:length>55</b:length>

<oc:offset>1071</b:offset>

<!---this incident URI is what the InstanceInfo is about->

<oc:subject rdf:resource="http://d.opencalais.com/genericHasher-1/f6e310ea-f54a-3aee-bc99-7293eea20f44"/>

<rdf:type rdf:resource="http://s.opencalais.com/1/type/sys/InstanceInfo"/>

</rdf:Description>

<rdf:Description rdf:about="http://d.opencalais.com/genericHasher-1/f6e310ea-f54a-3aeebc99-7293eea20f44">

<rdf:type rdf:resource="http://s.opencalais.com/1/type/em/r/Arrest"/>

 $<\!\!\mathrm{c:person\ rdf:resource="http://d.opencalais.com/pershash-1/1b1289ef-845f-31a9-a640-b6724dbe61e1"/>}$

<c:date>2010-05-28</c:date>

<c:datestring>May 28</c:datestring> </rdf:Description>

• • • •

OpenCalais, Geonames plus BaseVISor

- Leverages OpenCalais' free web service
 - Results returned as RDF based on OpenCalais ontologies
 - Ontologies not specific to street gangs
 - Results not always correct or complete so requires additional analytic processing
 - Only based on local contexts (not document wide)

SEMANTIC ANALYSIS



BASEVISOR SEMANTIC ANALYSIS

• Augments OpenCalais output

- Adds data types to RDF
- Corrects misidentifications (Mara Salvatrucha not a person)
- Time and location inferencing based on Global Document Context
 - Provide who, what, when, where for ALL events of interest
 - Infer specific geolocation (lat/long) using Geonames and Global Document context. (San Francisco source, "Santa Cruz" -> Santa Cruz, CA)

• Ontological Reasoning

- Insert initial facts AND inferred facts into RDF data store
- Based on Gang Ontology and rules
- E.g. If John is a member of Latin Disciples and Latin Disciples is a gang, then John is a GangMember (Ontology)
- If John joined MS-13 (i.e. there is a joining and John is the Agent and MS-13 is the Theme/Object), and MS-13 is a gang, then John is a GangMember (Rule)



RDF DATA STORE

- Add RDF results stored in time-dependent context with an RDF Data Store
- Currently using OpenSesame
 - Free, open source Sesame-based RDF Store from openRDF.org
 - Implements a query language very close to SPARQL 1.0
 - Java based API integrates well with BaseVISor

SPARQL MODEL UPDATES AND ALERTS

```
#"Arrests, Count"
```

SPARQL_ARRESTS_COUNT = PREFIX <u>rdf</u>: <http://www.w3.org/ 1999/02/22-rdf-syntax-ns#> SELECT DISTINCT ?r ?n ?d WHERE { ?r rdf:type <http://s.<u>opencalais.com</u>/1/type/ <u>em</u>/r/Arrest> . ?r <http://s.<u>opencalais.com</u>/1/pred/ person> ?n . ?r <http://s.<u>opencalais.com</u>/1/pred/ date> ?d }

#"Recent MS-13 Trial, Count"

SPARQL_RECENT_MS13_TRIAL_COUNT = PREFIX <u>rdf</u>: <http://
www.w3.org/1999/02/22-rdf-syntax-ns#> SELECT
DISTINCT ?r ?v ?n FROM <http://vistology.com/swae/
2010/week31> WHERE { ?r rdf:type <http://
s.opencalais.com/1/type/em/r/Trial> . ?r <http://
s.opencalais.com/1/pred/person> ?v . ?v <http://
localhost/default#memberOf> <http://localhost/
default#MS-13> OPTIONAL {?v <http://s.opencalais.com/
1/pred/commonname> ?n} FILTER (!regex(?n,
"salvatrucha", "i")) } LIMIT 5000

ALERT GENERATION: EMAIL

Subject: Daily MS-13 Summary Report

Date: Mon, 2 Aug 2010 14:50:42 -0400 (EDT)

From: ms13watcher@vistology.com

To: osimakoff@vistology.com

Recently identified MS-13 members:

James Viadero Hector Retana Edel Hernandez-Martinez

• • •

Recent MS-13 Arrests:

Edel Hernandez-Martinez

locname="Cincinnati"

url="http://www.journal-news.com/news/crime/second-man-arraigned-in-casa-tequilaslayings-819307.html"

Oscar Montoya

date="2010-07-27"

locname="Centreville" age="30" gender="M"

url="http://www.examiner.com/x-5919-Norfolk-Crime-Examiner~y2010m7d28-Immigrationand-Customs-Enforcement-arrests-87-criminal-aliens-across-Virginia?cid=channel-rss-News"

Recent MS-13 Convictions:

• .

Alejandro Umana date="2010-06-29"

locname="Cincinnati" age="25" gender="M"

charge="December 2007 murders of two brothers"

url="http://www.topix.com/rss/search/article.xml?q=%22ms-13%22+OR+%22mara+salvatrucha%22&x=0&y=0"

SEMANTIC MEDIAWIKI

- Semantic MediaWiki (SMW) is an extension of <u>MediaWiki</u> the wiki application best known for powering Wikipedia that
- Helps to search, organise, tag, browse, evaluate, and share wiki content.
- While traditional wikis contain only text which computers can neither understand nor evaluate,
- SMW adds *semantic annotations* that allow a wiki to function as a collaborative database.
- In-line triples. Subject is page topic. [Predicate:Object]
- Semantic MediaWiki was first released in 2005, and currently has over ten developers, and is in use on hundreds of sites.
- In addition, a large number of <u>related extensions</u> have been created that extend the ability to edit, display and browse through the data stored by SMW.

SMW PAGE GENERATION: GEOSPATIAL OVERLAY



10/28/10 VIStology, STIDS 2010, GMIU

SMW PAGE GENERATION: INDIVIDUAL

| | age discussion | edit history | | | New App Can Identify I
Submitted |
|-------------------|------------------|------------------|----------|----------------------|-------------------------------------|
| | ngmar Gu | andique | | | |
| SMW | Age(s) | 27 | | | |
| A | ternate Name(s) | Ingmar Guandique | | | |
| ation | Name(s) | Ingmar Guandique | | | |
| Nain page | Gender | Male | | | |
| current events | | | | | |
| lecent changes C | ategory: Gang Me | mbers | | | |
| Random page | | | | | |
| F | acts about Ingm | ar Guandique 🕦 | | | |
| ch | AlternateName | | | Ingmar Guandique + 🔍 | |
| | | | B:age | 27 + 🔍 | |
| Go Search | | | B:gender | Male + 🔍 | |
| ox | | | B:name | Ingmar Guandique 🕂 🔍 | |
| /hat links here | | | | | |
| elated changes | | | | | |
| pecial pages | | | | | |
| rintable version | | | | | |
| Permanent link | | | | | |
| Browse properties | | | | | |

SWAE BENEFIT EXAMPLE



 Documents containing "MS-13", "arrest", and City

- Semantically Identified MS-13 Arrests
- Actual MS-13 Arrests

| NATO STANAG 2022 (JC3IEDM, US Army HUMINT) | | | | |
|---|--|-----|--|--|
| Reliability (Source) | Credibility (Reported Information) | | | |
| A: Completely reliable . It refers to a tried and trusted source which can be depended upon with confidence. | 1 : Confirmed by Other Sources . It can be stated
with certainty that the reported information
originates from another source than the already
existing information on the same subject. (JC3IEDM:
3 Independent Sources) | | | |
| B: Usually reliable . It refers to a source which
has been successfully used in the past but for
which there is still some element of doubt in
particular cases. | 2: Probably True . The independence of the source
of any item of information cannot be guaranteed, but
from the quantity and quality of previous reports, its
likelihood is nevertheless regarded as sufficiently
established. (JC3IEDM: 2 Independent Sources) | | | |
| C: Fairly reliable . It refers to a source which
has occasionally been used in the past and upon
which some degree of confidence can be based. | 3: Possibly True . Despite there being insufficient
confirmation to establish any higher degree of
likelihood, a freshly reported item of information that
does not conflict with previously reported behaviour
pattern of target. (1) | | | |
| D: Not usually reliable . It refers to a source
which has been used in the past but has proved
more often than not unreliable. (JC3IEDM: The
probability of producing erroneous
information is high (>30%).) | 4: Doubtful . An item of information which tends to
conflict with the previously reported or established
behaviour pattern of an intelligence target. | GMU | | |
| E: Unreliable . It refers to a source which has been used in the past and has proved unworthy of any confidence. | 5: Improbable . An item of information that
positively contradicts previously reported information
or conflicts with the established behaviour pattern of
an intelligence target in a marked degree. | 5 | | |
| F: Reliability cannot be judged . It refers to a source which has not been used in the past | 6: Truth of information cannot be judged. | | | |



STANAG 2022 EXTENSION

BASEVISOR EVALUATION REASONING

BaseVISor performs "information evaluation" based on NATO STANAG 2022 credibility/reliability metrics

Same event; different sources \rightarrow Credibility \uparrow

Same who; what; when; different where

-> Credibility Ψ

- Social Media; no Mainstream source → Reliability Cannot Be Judged
- Mainstream (US) media -> Usually Reliable

| Dec 10, 2008 Rod Blagojevich, a Democrat, was arrested
Tuesday on federal corruption charges. Illinois Gov. Rod
Blagojevich returned to work Wednesday, (CNN) |
|--|
| Oct 22, 2010 A twice-elected Democrat, <i>Blagojevich</i> , 53, was
arrested in December 2008 on charges that he tried to link
official actions by his office to (Bloomberg.com) |
| Jan 29, 2009 Mr. Blagojevich, who was arrested Dec. 9 on corruption charges, (NYTimes) |
| Dec 11, 2008 Blagojevich, who was arrested Tuesday on
corruption charges (ChicagoDefender.com) |
| <pre><no date="">Illinois Governor Rod Blagojevich (D) was arrested today on corruption charges. The (WhoPlaysIn.com)</no></pre> |
| Who: Rod Blagojevich owl:sameAs Blagojevich C
sameAs Illinois Governor Rod Blajojevich (D) |
| When: Today owl:sameAs Dec. 9 owl:sameAs Tuesday |
| Where: Location? |

What: Arrested...on federal corruption charges 27 owl:sameAs arrested ... on charges that he tried to link owl:sameAs corruption charges

SWAE PHASE I CONCLUSIONS

• Implemented an automated SWAE process that:

- Collects and assembles large amounts of streaming data
- Uses semantic web technologies to provide understanding of content, including classifications and relationships/links
- Automatically generates alerts about critical events in real time
- Distributes alerts to users (currently via email, but easy to extend)
- Allows user to specify alert conditions and to view the collected data within the SMW
- Keeps/updates models of situation and detects when data deviates from in model
- Represents assembled information in Semantic Wiki pages for distributed collaborative assessment
- Leveraged and integrated existing algorithms and software
 - OpenCalais, Geonames, SMW, SPARQL, BaseVISor
- BaseVISor-like reasoning on RDF/OWL graphs and ontology is critical to the approach

FUTURE RESEARCH

- Alert Language and Engine
- Semantic Wikis
- RDF Data Stores
- Data Sources
- Improved Extraction Algorithms
- Enhanced Alerting and Customization
- Reliability/Credibility/Uncertainty Reasoning
 - Can SNA metrics be used as indicators of Reliablity?
 - How to infer source independence (e.g. on Twitter)?
 - How to maintain source track record?
- Evaluation and Stress Testing

QUESTIONS?

