Semantic Wiki for Tactical Intelligence Applications: A Demonstration

Dan Reininger (dan@semandex.net), Jeff Mershon (jdm@semandex.net), Jef Armstrong (jef@semandex.net), Ray Kulberda (ray@semandex.net), Andrew Cohen (andrew@semandex.net), P. Robert Bullard (bob@semandex.net), David Ihrie (dihrie@semandex.net);
Semandex Networks Inc., 5 Independence Way, Suite 309, Princeton, NJ 08540 (609) 681-5382

***Abstract* — This paper demonstrates a *semantic wiki* application that helps tactical users manage data from diverse sources and multiple locations.**

***Index Terms* — semantic wiki, tactical intelligence, ontology, threat characterizations.**

INTRODUCTION

Military forces operating from Forward Operating Bases (FOBs) currently have inadequate means to collect and organize information in ways that can aid in rapid understanding of the evolving conditions and threats in an area. Until recently, only anecdotal evidence existed indicating that lots of structured and unstructured datasets were available "in the wild" but went underexploited by tactical users due to semantic and syntactic incompatibilities. [1].

We have conducted a study to quantitatively profile data sources of relevance to tactical intelligence operations in a counterinsurgency [2]. The viewpoint of our study is from the perspective of tactical ground military intelligence support to operations at the regiment, battalion or company level, particularly semi-independent task forces at these echelons. At this level, the intelligence element of a military organization often serves as the primary data repository and the principal data analysis cell that produces products to support decision-making. While various organizations assign specific information storage and analysis responsibilities to different sub-elements, the intelligence cell typically draws on a broad range of data sets and offers some level of support to virtually the full spectrum of counterinsurgency operations, from civil affairs (CA) and psychological operations (PSYOPS) to kinetic targeting.

The study compiled representative data sources used in theater during combat operations in Iraq and Afghanistan and identifies over 250 sources relevant to tactical operations of conventional and special operations forces engaged in a counterinsurgency. We identified more than 50 formats such as disparate spreadsheets or summarized in text reports that circulate in the field as e-mail attachments. These formats are easy to produce in the field but the information they contain is hard to exploit when it comes time to find quick answers to operational questions.

In this paper we present a demonstration of a *semantic wiki* application that helps tactical users manage these data sets "in the wild". The Semantic Wiki we have developed:

1. Integrates heterogeneous information coming from diverse sources and multiple locations;
2. Uses a flexible ontology that can be evolved by the user community to organize that information in a way that makes it easy for users to capture and understand how each piece of data connects. This makes it possible to analyze information interactions and dependencies;
3. Uses standard web technology such as REST Application Programming Interface to present and extract that information to other tools and systems.

We demonstrate how this semantic wiki application allows non-technical users to integrate and manage data sets in the field and answer contextual analytical questions from its data, without the assistance of specialized IT personnel.

SEMANTIC WIKI IMPLEMENTATION

A semantic wiki is one of the newly emerging Web 3.0 capabilities. Web 1.0 put information on-line by creating and connecting web pages with URLs and HTTP that computers could understand. Web 2.0 enabled people to easily publish information, leading to blogs, social networking and the "traditional" wiki. With the Web 3.0 *semantic wiki*, people and computers both use a common information structure, allowing each to optimize around the things they do best. Computers connect, monitor and process large quantities of data sources and information, while people are much better at observing, interpreting and connecting information. The common structure is a set of web pages representing people, events and other types of *entities*, with links connecting different types of pages according to an *ontology*. The structure of the ontology is accessible to computers and easily understandable by people.

The ontology is defined and maintained by the user community and drives the information organization. When

new information is *collected*, it is categorized and linked into the overall, evolving collection of linked pages (*semantic graph*) according to the structure provided by the ontology. Any type of entity may be represented in the ontology, from the general (person, facility, event, place, network) to the specific (financial withdrawal, graffiti, railroad siding). A new instance of one of these types (Person: John Doe) is created, structured and linked according to ontology. Thus John Doe will have person attributes such as height, gender, or ethnicity rather than event attributes such as type, location and time. The types of linkages that John Doe can have are also appropriate to a Person, such as father-of, employed-by, and similar connections. Compared to other semantic approaches, which utilize a fixed ontology, our approach recognizes and supports the notion that the relevant information structure has to vary over time to stay relevant. We allow this to be done by the community of users in the field to accurately track tactical understanding as situations evolve.

The Semantic Wiki is implemented using our commercial semantic engine, Tango. Like other semantic technologies, such as Twine [3], Zemanta [4] and Noovo [5], Tango is built on top of a relational database, and not an RDF store. The Tango meta-model may be thought of as being conceptually closer to an object/UML orientation.

Much of Tango — including the UI — is controlled through the schema. When it starts up, a schema, which is stored in a custom XML format, is read in from disk. A UML representation of the loaded schema is generated to disk. The schema can be updated while Tango is running, and the schema changes persisted to disk. These dynamically introduced schema changes are properly reapplied if Tango is ever restarted.

We recently added an OWL counterpart to the UML generator, and an OWL file is also generated at start up. The OWL and UML representations can also be generated on demand while the application is running. This is important because a goal of ours is to support dynamic lists of concept instances within the ontology. Users define temporal, spatial and semantic constraints for set membership into these lists and group them to create threat characterizations, and we want to be able to capture that user knowledge and make it available to other services via OWL. Instance data can be returned in a variety of formats, including KML, our own TORI XML [6] structure, and JSON, and we recently introduced support for RDF.

In an effort to keep the ontology OWL-DL compliant, certain features of our meta-model are not currently exported, including relationship certainty, and evidentiary associations.

The main driver for our support of OWL/RDF is to facilitate re-use of data by other emergent analytical tools and systems that can deal with OWL/RDF structured data [7]. As the integration and interoperability efforts with other systems continue within our on-going projects we expect to receive feedback on the ontology and its structure, and identify future user requirements from the program transitions we will be doing next year.

DEMONSTRATION

The demonstration is based on some of the current capabilities we have developed under the ONR Large Tactical Sensor Networks project in support of Marine Corps Intelligence needs [8]. This Semantic Wiki Implementation for Marines (SWIM) combines Marine Corps customized data connections, ontology and threat models with our commercial semantic engine, Tango. We use the ontology and threat models provided by tactical intelligence users with current counterinsurgency operations experience to detect and issue indications and warning alerts on enemy threats in progress based on those previous observations. The demonstration uses representative but unclassified IMINT, SIGINT and HUMINT data. IMINT data includes suspicious event data from processed UAV video with focus on activities of vehicles possibly involved in threat activities. HUMINT and SIGINT data includes representative formats and entity types from tactical and national data sets. As this data is collected, reports are created and the data is presented to software applications and analysts who semantically link it based on the ontology. Fig. 1 shows the customized SWIM data processing pyramid, with the raw data at the lower level and the tactical intelligence analyst interacting with the semantic wiki on top.

From a capabilities perspective, our demonstration focuses on three important specialized types of concepts supported within the Semantic Wiki: Smart Lists, Characterizations and Semantic Widgets:

Smart Lists — A Smart List is a set of pages that match any criteria, such as new people entering a controlled area (HUMINT), calls from a monitored phone (SIGINT) during a certain time of day, or vehicles behaving erratically in the vicinity of an operation (IMINT). The Semantic wiki keeps every list dynamically up-to-date and can be combined with alerts for a powerful mechanism to monitor virtually any change to data relevant to the mission.

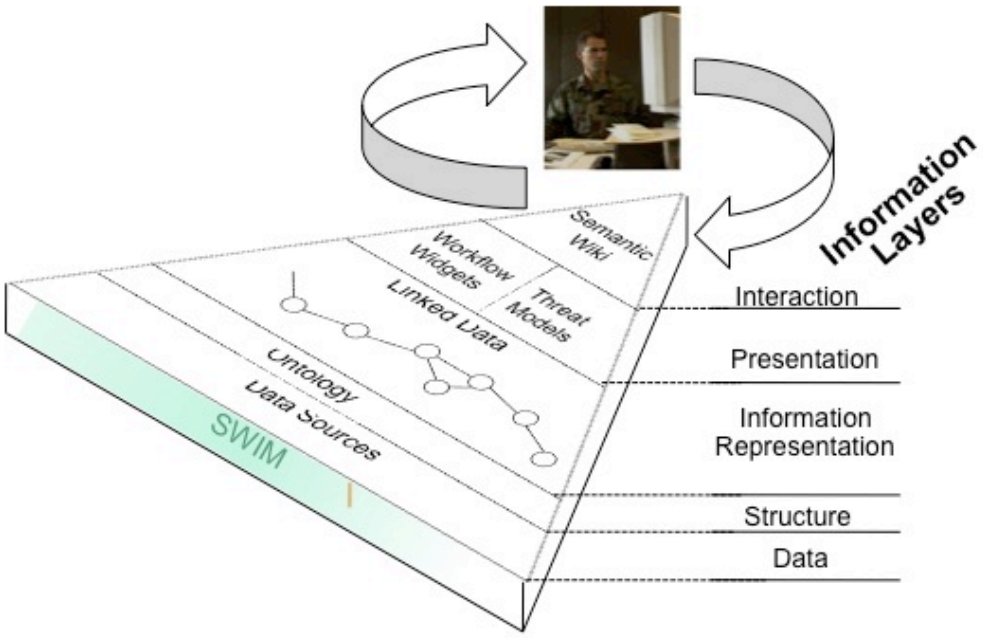

Fig. 1. Semantic Wiki for Tactical Intelligence Applications — Data Processing Pyramid

Characterizations — User-definable characterizations are the method tactical intelligence analyst can use to ask specific operational questions and determine if the information to answer them is available. A simple characterization might be used to mark as suspicious anyone who contacts a person on a watch list (represented as a Smart List). More complex characterizations can provide alert "clues" based on threat models of enemy Tactics, Techniques and Procedures (TTPs), or link specific devices to events based on complex associations.

One example of a characterization is the question: "Is this individual still at this location?", based on the operational need to verify information before a raid. Specific indicators could include SIGINT clues such as tipoff phone calls or a sudden absence of phone calls; IMINT might show vehicles leaving an area or people scattering through a field. HUMINT indications might involve an enemy operative seen buying food in a different town than expected. And, the characterization can combine these into both logical and temporal patterns: a flurry of calls followed by silence, with vehicles seen leaving an area shortly thereafter is a much stronger indicator than any of those detectable features in isolation.

A second example involves operational questions around whether an informant can be trusted. A call from a known bad guy may or may not be suspicious, since most informants associate with unsavory characters. However, a call from an unknown phone originating in the vicinity of a facility where suspicious activities have been observed represents a much more suspicious pattern.

We demonstrate specific examples of how characterizations help answer contextual questions such as: "Are these events a threat precursor, based on known tactics and trends?"

Semantic Widgets — The Semantic Wiki dashboard is home to widgets: mini-applications that let a tactical analyst perform common tasks and provide fast access to information. Because all the data on the Semantic Wiki is conformant to the ontology, the output of one widget can be linked to be the input to another, allowing users to create analytical pipes that capture best practices and serve to maintain knowledge continuity across rotations.

ACKNOWLEDGMENT

This material is based upon work supported by the Office of Naval Research under Contract No. N00014-07-C-0218 and DARPA under contract LM TT0705405 and 5R-44LM008474-03 (NIH). The views and findings expressed here do not necessarily reflect the views of these organizations.

REFERENCES

[1] Todd Hughes, "Toward Semantic Integration of Data in the Wild", Invited Talk, Ontology for the Intelligence Community (OIC-2007), November 28-29, 2007, Columbia, Maryland.

[2] P. Robert Bullard, "Sources of Tactical Data: A Study and Quantitative Profile", CUI project report prepared by Semandex for DARPA IXO, May 20, 2008.

[3] Twine is a product of Radar Networks, San Francisco,CA., United States, www.twine.com

[4] Zemanta is a product of Zemanta Ltd., London, UK, www.zemanta.com

[5] Noovo is a product of Noovo, LLC, Palo Alto, CA, United States, www.noovo.com

[6] *Tango Representational State Transfer (REST) API Guide*, Semandex Networks Inc., October, 2008.

[7] S. Stoutenburg, et.al., "Ontologies for Rapid Integration of Heterogeneous Data for Command, Control and Intelligence" In Proceedings of Ontology for the Intelligence Community (OIC-2007), Editor: Kathleen Stewart Hornsby, November 28-29, 2007, Columbia Maryland. http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-299/Proceedings.pdf

[8] Martin Kruger, Large Tactical Sensors Networks, BAA 07-026, Industry Brief, May 17, 2007. http://www.onr.navy.mil/02/baa/docs/07-026_07_026_industry_briefing.pdf