Semantic Web and Target-Centric Intelligence: Building Flexible Systems that Foster Collaboration

Aaron Manes
University of Maryland, College Park
8400 Baltimore Ave.
College Park, MD 20740
1.301.314.6627
awmannes@comcast.net

Jennifer Golbeck
University of Maryland, College Park
8400 Baltimore Ave
College Park, MD 20740
1.301.314.6609
golbeck@cs.umd.edu

James Hendler
University of Maryland, College Park
8400 Baltimore Ave
College Park, MD 20740
1.301.314.6611
hendler@cs.umd.edu

ABSTRACT
With the rise of the netwar paradigm new tools are needed to support intelligence collection and analysis. The Semantic Web uses information online in which data is defined in machine-readable terms, allows for the creation of flexible, adaptable knowledge bases that can be used collaboratively. This paper discusses how the Semantic Web facilitates research on terrorist organizations, particularly how a variety of useful features – such as network visualization and data attribution – can be used.

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]

General Terms
Design, Human Factors.

Keywords
Semantic Web, Terrorism, Intelligence Analysis, Web Portals

1. INTRODUCTION AND BACKGROUND

As intelligence requirements shift from the traditional hierarchical model to the target-centric approach, intelligence analysts and other intelligence community stakeholders require new information technology tools that foster collaboration. This paper is a brief discussion of an effort to use the Semantic Web to build the equivalent of an online white board that allows information sharing while also incorporating the computers' ability to rapidly search, organize, and graph data.

Traditionally the intelligence cycle has been a process with carefully defined steps in which separate communities focused on their part of the process and left the other aspects to other parts of the intelligence community. This assembly line approach has proven inadequate to fighting netwar – that is conflict with the complex, dynamic, non-linear networks that present the greatest intelligence challenges in the information age. Real-time situations involving numerous teams from different parts of the intelligence community monitoring and responding to events in disparate locations require new tools to facilitate collaboration. [1]

In a sense, target-centered analysis is not new, but it was not formally incorporated by the intelligence community. Individuals and small teams created ad hoc databases to track issues. But these were designed around the immediate needs of the investigator or investigating team. Consequently these databases, often paper files, are not readily intelligible to outsiders. Renowned CIA case officer, Robert Baer, who devoted a substantial part of his career to finding out who was behind the 1983 truck-bombing of the U.S. Embassy in Beirut, describes running a half-dozen Lebanese agents who gathered rumors, public records, political membership lists, old newspaper articles and photos in his book *See No Evil*. He would combine this information with information from the CIA database as well as transcripts from wiretaps. Then, Baer writes:

I would spend hours poring over the take, making connections between people, eliminating false leads, adding to my matrices. My makeshift charts started to look like the wiring diagram for a Boeing-747 cockpit. [2]

These charts are familiar to any researcher, but pity the investigator who inherits such a file. Unique abbreviations, cryptic notes, and assumptions about the information and the relationships charted characterize these charts. Far more useful in the netwar paradigm is an online database that shows who had entered data, with notes about how and why they

came to their conclusions, thereby providing a window into the thinking of the investigative team. Such a database would also transcend the limits of a paper system. Where only a few people can view a folder or chart at once, this online database could be accessed by multiple people from multiple locations. This would facilitate teams made up of members operating from diverse locations, it would also allow for easier collaboration between different teams. It would also allow investigators of one situation better access to the data of a related investigation. This would facilitate the intuitive processes - the hunches - that help investigators see patterns. A team looking at a new incident might find something useful in the false leads from an earlier stalled investigation.

2. INTELLIGENCE ANALYSIS ON THE SEMANTIC WEB

The Semantic Web is a useful medium to build these new tools. The best, brief definition of the Semantic Web is: "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."[3] Defining concepts for the computer means that the computer can then recognize specific instances of the concept and relate them to other concepts the computer. This allows the computer to process and manipulate data more elegantly and with greater ease. This has some inherent advantages to the intelligence analyst. It also eases the implementation of useful applications.

2.1 A Terrorism Portal

The website http://profilesinterror.mindswap.org is a terrorism web portal that serves as an access point to a Semantic Web terrorism database. It has numerous, useful attributes. An account-holder can enter new online data that is immediately incorporated in the database. The nature of the Semantic Web means that new instances linked to existing information automatically become live links, allowing for easy navigation. Finally, the Semantic web enables easy migration of data – both absorbing information from existing databases and exporting into other Semantic websites.

The site has several functions to facilitate an analyst's work:

- *Data Attribution*: Information is tagged by who entered the data. This simple step is essential to breaking down the oft-cited bureaucratic barriers to intelligence sharing. By seeing the specific person who provided the data, the analyst can quickly reach that individual for further background information.
- *Network Visualization:* Spontaneously generated graphs of social networks indicating different kinds of links between individuals and also graphs showing links between individuals and events. New information is immediately incorporated into these graphs. (Fig. 1)
- *Probability Visualization*: Information that is uncertain can be given a probability rating. The network graphs take this into account and create a broken line to indicate the uncertainty. (Fig. 1)
- *Comments Section*: Any visitor to the site can post notes. This parallels a chatroom and would allow analysts in different locations to discuss data online, and facilitate incorporating the discussion into the database. (Fig. 1)
- *News Reader*: Pulls in RSS feeds from relevant sites and highlights known instances and links to them.

There are also several useful features that are in development:

- *Graph Functionality*: The graphs are being improved to allow the user to define the specific queries being graphed. Also, the nodes of the network graph will be live links, facilitating the analyst's ability to navigate the database.
- *Calendars and Maps*: Locations and dates can be encoded to be Semantic Web readable. Just as the social network graphs are spontaneously generated and change as new information is added, maps and calendars can be incorporated into the database allowing visualization of the sequence and location of events. Nodes indicated on these maps and graphs will also be navigable.
- *Reasoning*: The Semantic Web enables the inclusion of rules so that the database can infer certain relationships from given data. For example by defining parent-child and sibling relationships the database will be able to show other familial relationships. Also, individuals who shared an address at the same time can be understood as housemates.
- *Automated Instance Creation*: Newsfeeds can be automatically marked up and, using rules and existing instances, new instances can be created.
- *Names Processing*: The multiple schools of transliteration between Latin and non-Latin alphabets has created tremendous problems for intelligence analysts. By creating a basic database of common variations on the spelling of common names and using rules the database can avoid double-entering people based on variations of the spelling of their name.

This is, by no means, the complete list of possible features nor are these capabilities unique to the Semantic Web. Other programs exist to do some of these operations. However, the unique flexibility of the Semantic web makes it relatively easy to incorporate these and other features onto the Semantic Web database. By including IT specialists in the intelligence stakeholder community, new features can readily be incorporated to address specific problems.

2.2 Building a Terrorism Ontology

A key feature of the Semantic Web is the need for an ontology – that is a way of organizing data and establishing relationships between concepts. The ontology is what defines the data for the Semantic Web database.

Using the Semantic Web's capabilities to monitor terrorist activity and model terrorist networks requires an ontology crafted to express complicated and sometimes contradictory information in an accessible and intuitive way. Obvious classes of items to define are people, which have evident properties such as place of birth, height, and nationality. Similarly, terrorist attacks have dates, targets, and numbers of casualties. But an actual attack is only the final link in a long chain of events including reconnaissance and planning, and assembling components and personnel. Major attacks often require years of planning and the individuals involved in the attack are frequently bound by a complex web of relationships that extend for years and even decades. Terrorists and terrorist organizations engage in a wide variety of activities that reflect the variety of human endeavor. An ontology that reflects this must maintain a balance between being sufficiently comprehensive while not being overly specific.[4]

A successful terrorist ontology will not be the end of the process, but instead the platform for effective analysis. A basic terrorism ontology could be shared across intelligence agencies which would facilitate information sharing. But teams working on specific issues could create classes, properties, and rules tailored to the specifics of the issue. A team investigating a terrorist organization's attack plans might require in-depth explorations of the characteristics and production of explosives. A team investigating an international drug trafficking operation might be particularly in the history of the ships involved and create classes and properties to address these issues. However, if these two teams found their investigations intersecting they would easily be able to merge their separate databases and incorporate specific tools and sub-ontologies.

3. CONCLUSION

This paper has focused on how the Semantic Web can be applied to researching terrorism, but its functions could be adapted for analysts examining innumerable other issues both in the public and private sector. The Semantic Web will allow many intelligence community stakeholders to examine and sort the same data – using a variety of computing tools. This both facilitates human intuition and the exchange of ideas, which is at the core of successful intelligence, while bringing into play the data processing strengths of the computer. These tools can help the analysts structure the data, a crucial tool in the current information saturated environment. Bringing numerous different perspectives and tools to bear on a problem is essential in the complex netwars that challenge modern intelligence agencies.

4. ACKNOWLEDGEMENTS

This work, conducted at the Maryland Information and Network Dynamics Laboratory Semantic Web Agents Project, was funded by Fujitsu Laboratories of America -- College Park, Lockheed Martin Advanced Technology Laboratory, NTT Corp., Kevric Corp., SAIC, the National Science Foundation, the National Geospatial-Intelligence Agency, DARPA, US Army Research Laboratory, NIST, and other DoD sources.

5. REFERENCES

[1] Clark, Robert, *Intelligence Analysis: A Target-Centric Approach*, CQ Press, Washington, DC: 2004.

[2] Baer, Robert, *See No Evil*, Crown Publishers, New York: 2002.

[3] Berners-Lee, Tim, Hendler, James, and Lassila, Ora, "The Semantic Web," *Scientific American*, May 2001.

[4] Mannes, Aaron and Golbeck, Jen, "Building a Terrorism Ontology," Mindswap Technical Report: http://profilesinterror.mindswap.org

Fig. 1. The Semantic Web portal page for Osama bin Laden. This page features the social network visualizations, how the comments page can be used to discuss the data, and a few cases of probability visualization where data is not confirmed.