Counter Intelligence and the Semantic Web

James Hendler
Computer Science Department
University of Maryland
hendler@cs.umd.edu

It is our hypothesis that appropriate use of semantics, and particularly Semantic Web technologies, can have a significant impact on improving the speed of intelligence processing and sharing of intelligence process. For example, analysts within the intelligence community are often tasked as small teams in ad hoc ways to deal with emerging threats (this is sometimes referred to as “surge processing” within parts of the Intelligence Community). For example, immediately following an event a team of analysts may be created to determine who did it, how, and what other intelligence might have been available relating to the event. In other cases, an event may not have occurred, but some potentially interesting intelligence is discovered and a small ad hoc team is formed to explore what may be happening and to issue appropriate warning. In all of these cases, current IT infrastructure is strained because almost by definition the details of the group’s need are not yet in the system of record, the situation was not specifically planned for in IT infrastructuring, and the information sharing needs of the participants were not prearranged. Thus, current practice is for the group to create a small database, a set of notes (often on paper) or other shared information sources that are not accessible to anyone outside the group. Unfortunately, the ad hoc and often unsharable nature of these sources can be a significant detriment as other groups, ad hoc or standing, might be able to use the developing information to “connect the dots” if only it was made available earlier.

We have shown that the tension between the need to develop “certified” information that can be vetted and added to the system of record without fear of compromising an agency’s critical data resources and the need to rapidly create and share information can be mitigated by the use of Semantic Web technologies to allow groups to rapidly create, manage, and maintain local information portals. Such portals can allow greater sharing of information among those with the appropriate need to know (even if not in the original collaborative group), allow the collaborative group to rapidly pool their information for better decision making, and allow the developing information to be integrated with traditional information sources (including message feeds and processed reports) from both within the agency and from other agencies. Further, this information sharing can occur within existing (Web-based) information platforms already available to the intelligence community, thus allowing rapid deployment of portal technology without major new IT infrastructure needs or exotic new technologies.

The University of Maryland’s MINDSWAP research group has developed a technology evaluation project performed at the University of Maryland under a subcontract through CTC Corporation. That work established a base-line “end-to-end” capability demo and validated the Semantic Web Portal concept. We briefly review that work and then describe ongoing research extensions that could significantly enhance the Semantic Web portal technology developed previously.

Previous work:
http://swint.mindswap.org

In a project performed over the past year, we have developed a sample portal technology that was designed to demonstrate the potential of using Semantic Web technology for intelligence analysis. The specific goals of the project were:

- Connecting unstructured inputs to rich metadata
- Bringing in information from "open" sources - provenance
  - tracking who entered the information
  - tracking where the information came from
- Using (multiple) ontologies in OWL
  - To help with annotation
To organize information display
To manage import/export of unstructured and semi-structured data
- Image and subimage annotation and display
- Demonstrating new information portal management tools enabled by the Semantic Web

These goals were demonstrated in the development of a portal technology. An example of this system can be found on the Web at http://swint.mindswap.org which is available on the Web. It should be noted that the information in the portal is all unclassified and some of it is humorously incorrect (for example identifying Osama bin Laden as the kidnapper of Janet Jackson) – this makes it easier to demonstrate the technologies without running afoul of information sharing restrictions.

The Swint portal was able to demonstrate the stated goals using a set of Semantic Web technologies including use of multiple ontologies expressed in the OWL Web ontology language, the use of RDF databases and queries for portal presentation, use of the inferential capabilities enabled by the OWL model theory, and the ability to support all information sharing in the portal using standard web technologies (HTTP protocol, URI-based identification, and XSLT and Perl for presentations).

### Ongoing Work

The current portal focused on showing the feasibility of end-to-end capabilities. We are now working to deepen individual capabilities within the portal and to explore new capabilities that were beyond the scope of the original project. Current work includes:

- Extension to other media: the Swint portal worked primarily based on an image-based markup tool that allowed users to markup images based on the information in OWL ontologies. We content this ontology-based markup can be used for more complex markup including direct markup of language-based web sources, markup of audio and video files, and the markup of more complex photos such as those used by the National Geospatial Agency.

- Geospatial registration of information sources: The Swint portal allows an analysts to markup images based on content, but not to search for this content based on actual location (in various geo-coordinates) or temporal information. Thus, a query such as “show all photographs from location x,y at time T” cannot be built into the current portal. The importance of time and location as queues for intelligence analysis mandates that the tool be extended to include such a capability.

- Multi-view Collaborative work: the Swint model assumes analysts in the group markup information and submit it to the site, and all integration takes place within the web portal. RDF query capabilities, however, should make it possible for the integration of this information to happen more dynamically, so that analysts can rapidly see relevant information created by others, through their own tools, without having to search the site directly. This work will be based on research in distributed RDF query and on applying syndication techniques developed for text-based “blogging” to multimedia sources.

- Extended Provenance information: The Swint site is able to display the direct provenance of any information on the site – that is, who entered the information at what time. However, the site cannot currently track information and display further provenance, such as what sources the analyst used in deriving the information, what other information on the site it is used by, and where the information originally resided (for example, differentiating an open-source photo from one that has been previously included in an intelligence product).

- Information access and control: The Swint site uses system-provided mechanisms for access control. Thus, it is only able to use “role-based” mandatory access. New research on the Semantic Web is starting to explore the use of “rule-based” discretionary access, which could allow analysts finer-grained control over the materials being developed on the site. This would also allow ontologies to specify
default distribution rules (for example, allowing anyone to see certain properties, like the location at which a photo was taken, but limiting access to others, such as the phone number at that location). Site managers would be able to easily override these defaults when information sharing is necessitated, but only with respect to the specific information within their own site.

• Comparison rules: Analyst groups may wish to explore their data dynamically under different “rules” for the information. Thus, for example, a rule such as “all people with the same phone number are the same person,” which is clearly not true in general, might be quite useful for a temporary view of data to see what people are using the same phones. More complex rules, such as “assume any two people with more than three of the following properties are the same even if they have different names” are also desired. The Semantic Web inferencing enabled by the Swint site allows information to be dynamically changed by such rules, but there is no interface to these rules made available to the individual analyst (only the site manager can create these rules, and only by direct entry in an OWL-based rules format). The enhanced site should allow individual analysts to explore the information on the site using different views as needed, but also without imposing such views on other users.

Figure 1: Different views of the Swint portal - a) text search, b) portal presentation, c) image markup, d) management view.