A Multi-INT Semantic Reasoning Framework for Intelligence Analysis Support

**Terry Janssen (Lockheed Martin), Herbert Basik (Lockheed Martin),
Mike Dean (BBN), Barry Smith (University at Buffalo)**

Introduction

As is well known, each intelligence agency has developed its own mechanisms for representing data. The resulting stovepipes for transmission pose severe obstacles to the automated integration and management of data, giving rise to problems addressed already in a US Government Executive Order of August 27, 2004, which expressed a mandate to the effect that intelligence agencies must strengthen their mechanisms for the sharing of terrorist information, for example through more widespread and systematic use of XML and similar markup standards.[1]

The volume of available data and the complexity of the National Security environment are increasing so quickly as to overwhelm a finite workforce of analysts. Machines must augment human cognitive capacity in order to achieve the needed level of situational awareness. In what follows, we describe a Lockheed Martin IRAD project to address the problem of integrating the data generated by multiple intelligence agencies. We provide an overview of the project and of the solutions proposed.

Where traditional methods of what is called 'information fusion' have been developed primarily for integration of quantitative data, we focus on qualitative data (pertaining for example to *intention* or *threat*, to *religion* and *family relationships*, or to *relative spatial location*) expressed for example in observation reports.[2] Experience has shown that a combination of semantic technologies is appropriate for capturing such qualitative data. Our goal is to advance the needs of intelligence agents in interpreting very large bodies of such qualitative data by fostering enhanced situational awareness through the application of semantic technology. Little *et al.* describe those aspects of our project which pertain to the use of ontologies to support multi-INT data fusion when enhanced through the consideration of probabilities.[3]

Proposed Solution

The premise of this IRAD project is that a system can be built which allows multi-INT data to be semantically fused and reasoned over by machine.

We are building a common framework which provides services to intelligence analysts in a way that does not impose a common vocabulary across the intelligence community or force substantial harmonization of agency-specific approaches to knowledge representation. To this end we exploit the benefits of modularity in building a common upper-level framework to which agency-specific representations can be mapped according to need. The different modules must be interoperable, in order to allow pooling of data from different intelligence agencies. They must also be of high quality in order to gain gradual common acceptance in ways which bring about network benefits of synchronization in the ways in which data are expressed.

The sort of higher-level ontology-based integrating framework we have in mind is being realized already in the context of the Open Biomedical Ontologies (OBO) Foundry initiative.[4,5] Here a plurality of ontology modules is being created by different community groups using both Web Ontology Language (OWL) and OBO-specific ontology formats against a background of common development principles designed to ensure interoperability. The OBO Foundry family of ontologies is being used in large-scale projects for the integration of qualitative biomedical information, including geospatial information,[6] in ways which provide a precedent for the present IRAD project.[7] They provide a set of shared terminological building blocks which foster reliable pooling of more complex representations created in their terms. One crucial component of the Foundry initiative is the availability of reliable ontology converters. These ensure that the large bodies of biomedical data annotated using ontologies (such as the Gene Ontology[8]) created in the OBO format can be transformed into an OWL Description Logic (DL) format. Similar facilities are available to convert OWL-DL ontologies to the CL format within the framework of our present project.

Semantic Multi-INT Data Integration

Our project hypothesis is that it is possible to create a similar, unified but modular, knowledge space for intelligence-related information integration, comprising both general-purpose open source components (pertaining to geography, religion, transport, etc.) and supplementary special-purpose components provided within various intelligence agencies. The result should provide useful augmentation to human analysts in a way that will help them to achieve the sort of (Level 3) information fusion (situational awareness) which involves integrating characteristics, behavior, political and religious affiliations, locations, etc. of individual entities into higher-level contexts.

Some of the types of IMINT, SIGINT, HUMINT, ELINT, and open source information we will need to integrate are represented in Figure 1. The hypothesis is that even though these different bodies of information are described using different ontologies based on different logical approaches, they can be unified and reasoned over by automated tools given the right sort of computational framework.

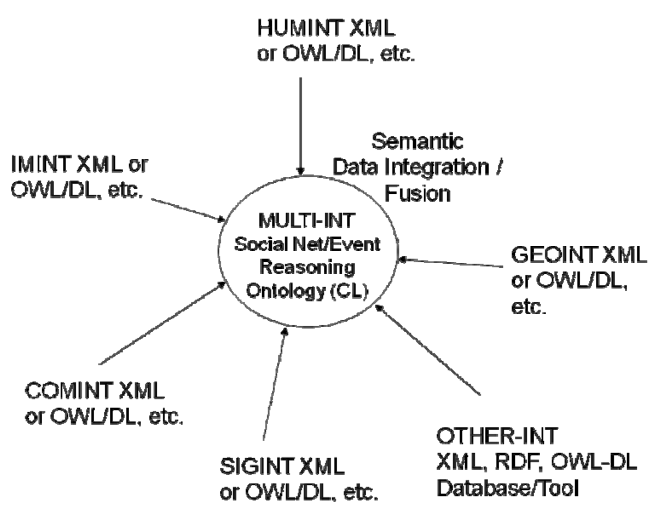

Figure 1: Varieties of multi-INT Information

Two levels: the level of classes and the level of instances

Our project rests on a distinction between two levels of entities and of corresponding information that is in some ways analogous to the distinction between T boxes (for terminology) and A boxes (for assertions) used in the DL community. On the higher level are classes or types (and corresponding generic information); on the lower level are instances or particular entities (and corresponding specific information). Classes or types are for example *person*, *settlement*, *plan*, *train*, *aircraft*. Instances are this particular person John or the particular plan that was agreed on by John and four other persons at such and such a time and place. Names of classes or types are used to annotate instance-level data, as for example when a report on image data uses GARCON-F markup to capture the fact that a *business jet* of a certain *type* is *stationary* at a certain *airport.* Ontologies as we conceive them provide the resources to capture generic information in a shareable form which makes associated data themselves shareable and algorithmically tractable.

In the domain of intelligence our information comes from a number of varied sources, many of which will produce either:

1) similar information about the same instances,
2) similar information about distinct instances (which may accordingly be confused),
3) differing information on the same instance (this can produce conflicting information, e.g., concerning the spatial or temporal location of an event),
4) differing information about different instances (there may be two separate but related items being tracked in different ways).

Where we are reasoning about how instance-level data fit together to form a common operating picture, there is inevitably uncertainty. Intelligence reports are noisy and information is incomplete. There are active attempts by adversaries at deception. This will mean that all of the mentioned alternatives will generate knowledge problems, for example because we sometimes believe that two instances are identical when they are in fact distinct. In practical terms it means that combining probabilistic reasoning with semantic technology is an important enabling capability for multi-INT fusion. And facilities for probabilistic reasoning will accordingly be an essential component of our project.

The Need for High-Expressivity Ontology Languages

Intelligence agencies have developed INT-specific terminologies for describing qualitative data which define terms and relationships in semantically similar but not identical ways. Many of the most advanced of these models have been represented in the OWL-DL format. OWL-DL is a W3C standard with many attractive algorithmic properties. Unfortunately it is a low-expressivity language, which means that it faces considerable difficulties when used to express complex qualitative information especially in areas where time and change are involved.[9] For this reason our project will draw on the resources not only of OWL-DL but also on the more expressive language of Common Logic (CL), a proposed ISO standard.[10] We will draw specifically on the resources of CL with Well-Founded Semantics,[11] which has not only nice computational properties when used in support of reasoning over large bodies of data, but also the sort of high expressivity we need to represent complex real-world situations. CL provides the expressivity we need to describe things that are changing/evolving over time, for example military and paramilitary organizations, family and tribal groups, which gain and lose members, change their locations, become allied or alienated.[12] Well-Founded Semantics provides fast and efficient query answering capabilities even when addressing large data collections comprehending numbers of entries in the 10s of millions. Like OWL-DL, Common Logic is XML compliant. At the same time CL is marked by a high degree of syntactic flexibility and thus individual CL systems may use a non-XML syntax; these are however in every case mappable to a fully XML-compliant syntax.

Our system will be useful only if it can be executed in responses to real query needs of intelligence analysts in response to real-time changes in real-world environments, and thus it has to be computationally quite nimble. Given that the use of OWL-DL is becoming more widespread it should work well also with OWL-DL resources.

Another key facet of the ontologies of interest to the intelligence community is the ability to express relationships between people, and to construct representations of social networks over populations of individuals. Our project will expand the models of social networks currently in use by adding dynamic spatial and temporal relationships which will be fully integrated within the larger modular framework.

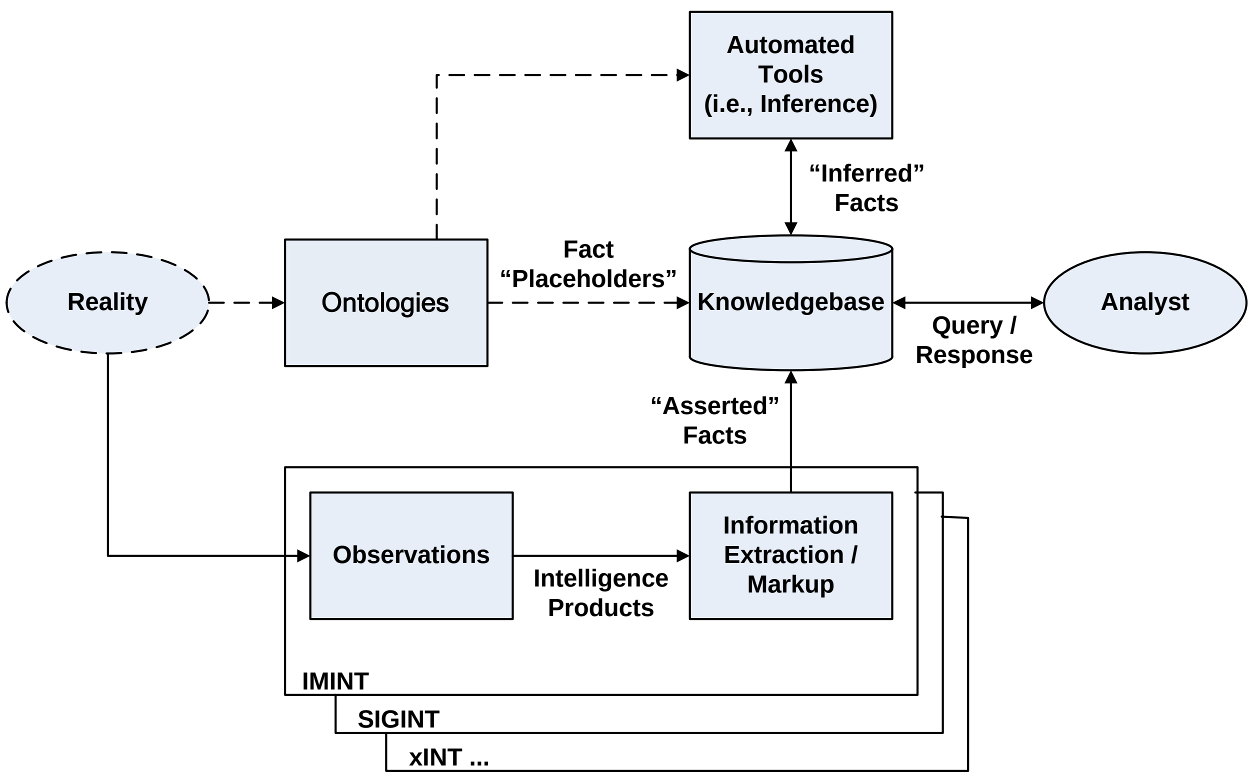

Figure 2: An Outline of the IRAD Architecture

Conclusions

In the foregoing we have described only the basic outlines of the project. In addition we are realizing a number of additional components, including image annotation, data import and results visualization. Our major focus is to construct the engineering required to take this into production (Figure 2), and to bring our pilot testing on artificial data to the level where the approach can be thoroughly tested by information analysts on large bodies of real-world data.

Acknowledgments

We are grateful to the Lockheed Coporation, and also to the following for crucial contributions to this work: Jim Pridgen, Tony Chen, Jessica Hartnett, Kirsten Gheen, Sarah Taylor, William Andersen, Ryan Kohl, Mitch Kokar, Kathryn Blackmond Laskey, and Eric Little.

References

---

[1] http://usinfo.state.gov/xarchives/display.html?p=washfile-english&y=2004&m=August&x=20040827173509adynned0.6143305

[2] Maureen Donnelly, "Relative Places", *Applied Ontology*, 2005. 1 (1).

[3] Eric Little, Kathryn B. Laskey, Terry Janssen, "Ontologies and Probabilities: Working Together for Effective Multi-INT Fusion" Ontologies for the Intelligence Community, Columbia MD, November 2007.

[4] http://obofoundry.org/.

[5] Smith B, Ashburner M, Rosse C, *et al.* "The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration", *Nature Biotechnology* (in press).

[6] http://darwin.nerc-oxford.ac.uk/gc_wiki/index.php/EnvO_Project.

[7] Alan Ruttenberg, "Harnessing the Semantic Web to Answer Scientific Questions: A HCLS IG Demo", http://www.w3.org/2007/Talks/www2007-AnsweringScientificQuestions-Ruttenberg.pdf

[8] http://www.geneontology.org/.

[9] Leo Obrst, "Ontologies for Semantically Interoperable Systems", *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, November 2003, 366-369.

[10] The Common Logic Standard, http://philebus.tamu.edu/cl/.

[11] Weidong Chen, Terrance Swift and David S Warren, "Efficient Top-Down Computation of Queries Under the Well-Founded Semantics", *Journal of Logic Programming*, 1993, 12, 1-199

[12] K. Hornsby and S. Cole, "Modeling moving geospatial objects from an event-based perspective", *Transactions in GIS*, 11(4): 2007, 225-243.