ONTOLOGY FOR THE INTELLIGENCE COMMUNITY:

TOWARDS EFFECTIVE EXPLOITATION AND INTEGRATION OF INTELLIGENCE RESOURCES

NOVEMBER 28-29, 2007 · COLUMBIA, MD

ABSTRACTS

INVITED TALKS

Toward a Data Integration and Exploitation System that Learns

Todd Hughes (DARPA)

The development of an automated data integration system will entail overcoming a number of significant technical challenges. Fortunately, there are there are many promising technical approaches, many of which have ontologies as essential elements. The efficacy of any technical approach will be determined by the performance of the data integration system, measured against a set of objective criteria in experiments using realistic data sources. This briefing will present some ideas about data integration experiments, which suggest the priority of certain technical challenges and the value of ontologies in an operational data integration system.

TODD HUGHES is a Program Manager for DARPA/IXO (Information Exploitation Office). His research interests include geospatial intelligence, cognitive systems, and data integration. His current focus is the extraction of intelligence from geospatial imagery through neurobiological models of vision and formal semantic models of reasoning. Prior to joining DARPA, Dr. Hughes was Technical Manager at Lockheed Martin Advanced Technology Laboratories, where he oversaw research and development projects in cognitive systems and semantic technology.

How to Keep Track of Absolutely Everything

Werner Ceusters (NCOR)

Referent Tracking (RT) is a new methodology for building digital copies of whatever parts of the world we are interested in. These copies, when running inside a computer, should mimic – down to some determined level of granularity – whatever it is that is happening in reality. RT was originally designed to solve problems in healthcare, but it is applicable also to the work of the intelligence analyst. Key to the paradigm is the assignment of singular globally unique identifiers to particular entities in reality and the systematic reuse of these identifiers in statements about how tracked entities are believed to relate to each other. A carefully implemented RT system, combining observational data from a multiplicity of information sources, is able to track entities and their history over time and is thus an ideal instrument for making deductions, analogies, and predictions about what is and likely will be the case, and also for identifying gaps and areas of uncertainty in our knowledge.

WERNER CEUSTERS studied medicine, neuro-psychiatry, informatics and knowledge engineering in Belgium. Since 1993, he has been involved in numerous national and European research projects in the area of Electronic Health Records, Natural Language Understanding and Ontology. He is currently Professor in the Psychiatry Department of the School of Medicine and Biomedical Sciences, SUNY at Buffalo NY; Director of the Ontology Research Group, New York State Center of Excellence in Bioinformatics and Life Sciences; and coordinator of Bioinformatics for the Health Science Faculties at UB.

Ontology vs. Ontologies – The Intelligence Community Needs to Embrace Both

Steven Robertshaw (Defence Science and Technology Laboratory, UK)

The approach taken in the ontology design process dramatically affects the usability (and re-usability) of computational ontologies. It is argued here that a more rigorous approach to computational ontology

development would benefit everyone. The points raised will be illustrated though personal experiences collected over many years.

STEVEN ROBERTSHAW began his working career in the Royal Air Force as an aeronautical engineer, migrating through surveillance and weapons systems into a black area, where he found himself when the Cold War thawed in the early 1990s. With nowhere else to go, he fell naturally into the study of Ontology, an area from which he has not yet managed to escape. Currently he is working for the UK Ministry of Defence as Principal Scientist in the Informatics team of their research arm: DSTL, where he is focused mainly in the area of Multi-Source Intelligence Fusion.

Intelligence and the Semantic Imperative

David C. Roberts (Chief Technology Officer and Chief Data Architect, CIA)

The discussions of semantic technology can seem to be worlds apart from the day to day working world of intelligence. How can we hope to ever apply semantic technology to intelligence problems, given where we are today?

This talk will focus on the real world of intelligence data. We face semantics issues every day, and every day we have to deal with them. New technology holds out hope that automation can help. But how can we take realistic steps toward these important opportunities, steps that we can apply now to the real world of intelligence data as it exists now? At CIA we have wrestled with these issues and believe that we have a practical bridge to the semantic technology future, producing information products that are practical and in demand today, while also enabling the bright future of semantic processing based on ontologies.

DAVID C. ROBERTS is the Chief Technical Officer of Applications Services in the Central Intelligence Agency. His career spans government and industry. He has worked for software and systems companies, serving in both management and design roles. He was employee #12 at Oracle Corporation, and served as Vice President, Marketing, there. At the CIA, Roberts introduced enterprise architecture and he now leads the Agency's efforts in enterprise data architecture, where he ponders how to apply semantic technology to today's intelligence problems. He has graduate degrees in electrical engineering and computer science, and teaches in computer science at George Washington University.

ACCEPTED PAPERS

Uses of Ontologies in Open-Source Blog Mining

Brian Ulicny, Chris Matheus, Mitch Kokar, Ken Baclawski

The blogosphere provides a novel window into public opinion, but its dynamic nature makes it an elusive medium to analyze and interpret in the aggregate, where it is most informative. We are developing new technology employing ontologies to solve this problem by fusing the signals of the blogosphere and zeroing in on issues that are most likely to migrate offline, enabling analysts to anticipate the threats or opportunities they represent.

Policies for Public Domain Ontologies for the Intelligence Community

Elisa F. Kendall, Jay Jacobs, Deborah L. McGuinness, Stephen Schwab

Numerous RDF vocabularies and OWL, KIF, and other knowledge representation language ontologies have been contributed to the growing body of ontologies available in the public domain over the last ten years. Many of these were created with government-funded research support in the US and EU. Only a small subset is reusable, and fewer are appropriate for use in applications supporting evolving Intelligence Community requirements. This is partly due to decreasing funding available in the US in particular, but also because of lack of well-specified policies for vocabulary management, metadata, and provenance specification. In this paper we will highlight some of the challenges we have faced in developing and attempting to reuse ontologies in support of DARPA and US Department of Defense initiatives, and provide fodder for discussion of requirements for public domain ontologies.

The Use of Ontologies to Support Intelligence Analysis

Richard Lee

In this paper we describe the Metadata Extraction and Tagging Service (METS) system in use at DIA. We briefly describe the purpose and function of the system. We explain why we chose to use OWL and ontologies rather simple XML for the representation of the data it produces. We discuss an experiment we conducted on using ontologies for multi-int data fusion. We describe the OWL ontologies we've developed. We conclude with a list of the ontology and data coordination we hope to do in the future.

Creating a Geospatial and Visual Information Ontology for Analysts

Chumki Basu, Hui Cheng, Christiane Fellbaum

An ontology is a main component of an evolving knowledge base that caters to multiple clients. Consider a scenario where an automated procedure (a computer vision algorithm) used in an analyst tool detects different kinds of "roads" in images, and features in the ontology are used to distinguish a "paved" road from a "dirt road". In another scenario, the ontology enables reasoning about "locations", supporting analysts' geospatial information processing tasks. In this paper, we describe the creation of a multi-use geospatial and visual information ontology, GVIO, building on and integrating with the lexical database, WordNet. To ensure that GVIO can interoperate with other ontologies in useful ways, we inherit as much of the WordNet structure and content as is relevant for the domain of aerial surveillance and link in new content/structure as necessary.

A Multi-INT Semantic Reasoning Framework for Intelligence Analysis Support

Terry Janssen, Herbert Basik, Mike Dean**,** Barry Smith

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. 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.

Achieving Ontology-Assisted Query of Graph Databases

David Silberberg, Wayne Bethea, Dennis Patrone, Paul Frank, David Patrone, John Gersh, Elisabeth Immer

We describe an approach for enabling ontology-assisted queries onto existing schema-based graph database systems without altering the graph query language or the corresponding graph database system. Typical schema-based graph database systems enable analysts to formulate queries using terms from a schema. Our approach enables analysts to formulate queries using terms from a *virtual schema*, which is composed of an ontology, a graph schema, and mappings between them. A software system can then assist the analyst by extracting the predicates and terms from queries, and in conjunction with the ontology and a reasoner, produce a set of corresponding graph queries that contain only terms from the graph schema. These queries are then sent to the graph database for evaluation. This approach enables intelligence analysts to focus on analysis that is more complex while the ontology-assisted query capability performs lower level reasoning. A distinction is maintained between the ontology reasoning and graph query systems to 1) take advantage of the performance of graph query engines while exploiting the semantics of the ontologies, 2) provide multiple analysts with an explicit and consistent semantic model of the graph data, and 3) enable multiple analysts with different semantic models of the data to use their own personal ontologies for analysis.

Ontologies and Probabilities: Working Together for Effective Multi-INT Fusion

Eric Little, Kathryn B. Laskey, Terry Janssen

This paper will address the question of how formal ontologies can best be combined with probability theory to provide theoretically sound and practically useful semantic technology for multi-INT fusion. We will investigate theoretical concerns associated with the connections between logics associated

with formal ontology (e.g., description logic, common logic, first-order logic) and those of probabilistic mathematics. The goal is to provide a high-level discussion of the issues involved with combining ontologies and probabilistic systems as a basis for dialog between these two communities, and to identify a broadly construed research agenda for their mutual development and interaction.

Ontologies for Rapid Integration of Heterogeneous Data for Command, Control, & Intelligence

S. Stoutenburg, L. Obrst, D. McCandless, D. Nichols, P. Franklin, M. Prausa, R. Sward

Increasingly Command and Control (C2) systems require the ability to respond to rapidly changing environments and intelligence. C2 systems must be agile, able to integrate new sources of information rapidly for enhanced situational awareness and response to real-time events. Data from varied sources across the world must be integrated and transformed into knowledge that can be leveraged. Machine-to-machine capabilities are also increasingly necessary to accomplish mission goals. To this end, we developed ontologies and rules to address emerging mission needs. We have found that ontologies and rules offer a powerful tool for rapid enterprise integration. With these, we were able to integrate new sources of data within hours, instead of weeks or months as with traditional software development methods.

Ontology-Driven Imagery Analysis

Troy Self, Dave Kolas, Mike Dean

This paper presents a new paradigm for imagery analysis where imagery is annotated using terms defined in ontologies, enabling more powerful querying and exploitation of the analysis results. The ontology terms represent the concepts and relationships necessary to effectively describe the objects and activities within a domain of interest. A platform for viewing and editing imagery annotations is described along with a specialized semantic knowledge base capable of efficiently querying the information using semantic, spatial, and temporal qualifiers. The Ontologies used for representing the annotations and domain of interest are also described.

Ontological Support for Bayesian Evidence Management

Michael N. Huhns, Marco G. Valtorta

This paper describes our work on an integrated system that can assist analysts in exploring hypotheses using Bayesian analysis of evidence from a variety of sources. The hypothesis exploration is aided by an ontology that represents domain knowledge, events, and causality for Bayesian reasoning, as well as models of information sources for evidential reasoning. We are validating the approach via a tool, Magellan, that uses Bayesian models for an analyst's prior and tacit knowledge about how evidence can be used to evaluate hypotheses.

Geospatial Ontology Trade Study

James Ressler, Mike Dean

This short paper summarizes a survey of ontologies relevant to geospatial intelligence. 45 geospatial and temporal ontologies, in 11 categories, were assessed against 3 use cases: annotation, qualitative reasoning, and information integration. Specific recommendations and more general conclusions are provided. The paper presents an illustration of a feature with several different ontology representations.

A Pragmatic Foundation for Defining a Rich Semantic Model of **Track**

Rick Hayes-Roth, Curtis Blais

Many defense, homeland security, and commercial security objectives require continuous tracking of mobile entities. The systems that perform these functions produce information products called *tracks*. A track associates observations with the mobile entity and typically includes position, velocity, and other simi8lar attributes. Military systems have sophisticated tracking and track fusion processes, but lack uniformity in syntactic and semantic content preventing effective sharing of the information. In other domains of interest, such as seagoing surface ships, dangerous cargo and persons of interest, tracking systems are less mature and have marginal performances. It is now essential that we be able to share

information across different tracking systems working in relating domains. In this paper, we describe the Rich Semantic *Track* model as a foundation for sharing world state information across multiple systems. The model exhibits a belief and evidentiary structure that has not been emphasized in previous track models for broad application.

Toward Automated Provability-Based Semantic Interoperability between Ontologies for the Intelligence Community

Andrew Shilliday, Joshua Taylor, Selmer Bringsjord, Konstantine Arkoudas

The need for interoperability is dire: Knowledge representation systems employ ontologies that use disparate formalisms to describe related domains; to be truly useful to the intelligence community, they must meaningfully share information. Ongoing research strives toward the holy grail of complete interoperability, but has been hindered by techniques that are specialized for particular ontologies, and that lack the expressivity needed to describe complex ontological relationships. We describe *provability-based semantic interoperability* (PBSI), a means to surmount these hindrances; *translation graphs*, one of our key formalisms for describing the complex relationships among arbitrary ontologies; and ways in which these techniques might be automated.