Semantic Interoperability Community Of Practice (SICoP)
Semantic Interoperability Conference 2006
9-10 February, 2006

Knowledge Broker Semantic Integration Prototype

Ray Piasecki
BAE Systems Inc.
E&IS Fellow
Technical Director, CFT/TCS Knowledge Systems
(858) 592-5104
ray.piasecki@baesystems.com
Topics

- Research Performed
- Semantic Integration Architecture
- Research Observations
- Prototype Demonstration
Knowledge Broker Prototype

1. Created a Reference Architecture Model that mapped KR, EII and UIM technologies across enterprise platforms and information tiers.
   - Captured core architectural constructs and key technologies.
2. Ran several focused technology studies.
   - Examined functional capability and breadth of several ontology platforms and analytics engines to study elements of the reference architecture.
3. Developed a large scale prototype to simulate an information enterprise.
   - Enterprise knowledge portal with advanced search and link analysis tools.
   - Federated Geospatial data systems.
   - Unstructured data stores.
4. Analyzed phase-1 results and re-factored into spiral-2 prototype.
   - Examined the practicality of knowledge representation platforms as evident in existing academic and COTS products.
   - Evaluated the robustness and stability of supporting standards.
Technical Studies

– Platform Analysis
  – Ontology: SnoBase, Jena2, TM4J
  – EII: MetaMatrix, LiquidData
  – Analytics: Gate, UIMA SDK

– Integration Analysis
  – Semantic Integration: Used the semantic layer as an integration construct. Mapping a domain vocabulary across the structured and unstructured data sets. (KR to EII connection)
  – Structured Data Systems: Federated queries against enterprise metamodel for access to data stores
  – Unstructured Data Systems: Performed an entity extraction and classification on data sets and posted annotations to populate ontological instances

– Ontology Analysis
  – Applied a suite of ontology's built for individual communities of interest (COI) rather then a complex and overly generic cross domain vocabulary
  – Experimented with model size to determine the practicality of reasoning against varied model complexities
  – Compared the end-user “application capability” enabled through a hierarchical taxonomy provided in a topic map against deeper semantic relationships facilitated in OWL/RDF
Knowledge Broker Systems Integration Prototype
Spiral-1 Components

Semantic layer applied as the systems integration construct. Applications searched against the semantic model.

Federated Access to distributed data systems mapped behind a domain focused semantic layer. Classes in the ontology are mapped to elements in the metadata catalog.

Documents and sites ingested. Entities were extracted from unstructured data sets and posted to populate model instances.

Knowledge Representation Platform
Ontology Engine
COI Ontology
COI Ontology
COI Ontology

Federated Data Access & Integration

Domain Knowledge Base

Knowledge Portal

Mission Content Repository
GeoSpatial Search
Information Alerts
Link Analysis
Semantic Search

Data Ingest & Entity Extraction

BAE SYSTEMS COPYRIGHT PROTECTED
Technical Observations

- **Platform technical issues were faced with even pragmatic prototyping**
  - Encountered memory and performance issues persisting OWL statements when trying to create a model robust enough to do some practical level of reasoning.
  - A modest owl model can be 500K statements. Our tests run with 90K OWL statements hosted on the Jena platform took hours to return a query.

- **Established a semantically rich interface to support discovery of enterprise data.**
  - A semantic layer was developed for a domain specific community of interest not as a unifying ontology for cross domain unification or as an enterprise-wide common vocabulary.
  - A universal semantic layer is not seen as practical nor semantically rich.

- **Model Instantiation**
  - Populating the ontological model with all instances of information across a large-scale enterprise does not appear to be feasible. The model quickly becomes impractically large.
  - Instead of implicit model population a query approach was used.
  - Classes in the ontology were mapped against entities in the enterprise metamodel. Attributes of the classes were mapped against metamodel properties. Upon request a class was populated using a query against the metamodel.
  - Complex integration code is required to interface the ontology layer and the back end data integration layer.
  - Evolving COTS products that move to incorporate both traditional EII and Semantic Markup engines in a single platform will significantly reduce integration glue code (i.e Metamatrix match-it).
  - Significant complexity is associated with creating an ontological model abstracting Temporal and Geospatial types

- **OWL & RDF tests**
  - Although OWL adds ability to specify richer semantic relationships, available platforms and query languages don’t offer a corresponding semantic query capability to leverage that information (i.e. Jena with RDF query).
  - Until OWL-QL arrives one is still operating at the level of extracting an element from a taxonomic hierarchy as opposed to querying for richer semantic constructs such as typed class relationships, cardinality, equality, etc.
Prototype Demonstration
Role Based Login

User login initializes the application windows based on role, task, and mission information context.
Collaborative Knowledge Portal

**Analyst Communities**
- Enterprise Projects
- Enterprise Calendar
- SME List
- Community Ontology
- COI Announcements
- Discussion Groups

**Analyst Tool Palette**
- Knowledge Assistant
- Information Search
- Data Access
- Analysis
- Exploitation

**Analysts Forums**
- Discussion Threads
- Post Messages
- Group Alerts
- Announcements
- Document Repository

**Collaborative Tasks**
- Projects
- Tasks
- Task Status
- Task Description
- Task Flow

**Community Tasks**
- Task
- Project
- Status
- Task Flow

**Combined Joint Task Force**
- Eye On Fallujah
  - US Joint Command has established the Combined Joint Task Force - Eye of Fallujah (CJTF-EOF) with the mission to identify and neutralize the vast IED network operating in the region of Fallujah.
  - The highly decentralized characteristics of the IED cells make them nearly impossible to penetrate.

**Eye On Fallujah**
- Analysis
- IED link analysis
  - Analysis link analysis generated by Knowledge Broker's DIA tool to assess the relationships between Abu Musab al-Zarqawi and Hafiz Mohammed Essed and determine their participation in the IED network...
- Assist Operations Group in planning target acquisition.
- Assist Operations Group in planning target acquisition activities for collection of target information. Coordinate security checks for indigenous personnel. Coordinate with other groups the enemy...
- Night extraction points both prior and rear.
- Collect, Process, Produce and Disseminate Intelligence.
- Collect, Process, Produce and Disseminate intelligence to the appropriate strategic/operational...

**HVT Extraction Update**
- Meeting set on Saturday October 22 at 10:00 AM.
- Post by mission_commander on 10/19/05 3:59 PM in project HVT Extraction

**Combined Joint Task Force Mission**
- JIEDDO Network
- JIEDDO Network
- JIEDDO Network
- JIEDDO Network
- JIEDDO Network

**TerraVista**
- ArcGIS
- TerraVista

**Tool Selector**
- New Project
- Manage Project Templates

**BAE SYSTEMS COPYRIGHT PROTECTED**
Integrated Tool Palette

Start DKA:
- Semantic Search
- Link Analysis
- Event Analysis
- Document Browse

Eye On Fallujah

- US Joint Command has established the Combined Joint Task Force - Eye of Fallujah (CJTF-EOP) with the mission to identify and neutralize the vast IED network operating in the region of Fallujah.
- The highly decentralized characteristics of the IED cells make them nearly impossible to penetrate.

Community Tasks

<table>
<thead>
<tr>
<th>Task</th>
<th>Project</th>
<th>Status</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyze IED link analysis</td>
<td>IED Network Sweep</td>
<td>0%</td>
<td>11/2/05</td>
</tr>
<tr>
<td>Assist Operations Group in planning target acquisition</td>
<td>IED Network Sweep</td>
<td>0%</td>
<td>11/6/05</td>
</tr>
<tr>
<td>Collect, Process, Produce and Disseminate intelligence</td>
<td>IED Network Sweep</td>
<td>0%</td>
<td>11/9/05</td>
</tr>
<tr>
<td>Night Extraction points both pim and sec</td>
<td>HVT Extraction</td>
<td>25%</td>
<td>11/7/05</td>
</tr>
</tbody>
</table>
Domain Knowledge Assistant (DKA)

- Domain Knowledge Maps
- Mission Vocabulary & Taxonomy
- Link Analysis Viewer
- Document Search
- Information Alerts
Start iWeb:
- Reach Back/ Federated Access
- Geospatial data discover, browse and retrieve
Federated Access To GeoSpatial Data Products

Data retrieved from distributed systems and spatially correlated for browsing

Unstructured documents linked to the geospatial data via autonomous information extraction
Questions ?