Agent Semantic Communications Service (ASCS)

Teknowledge

John Li, Allan Terry

November 2004
The problem:
- Leverage semantic markup for integration of heterogeneous data sources

Technical solution strategy
- Semantic search agents, not the familiar keyword search
- Agents index RDF/OWL triples

Research approach
- Develop search/query engine
- Align terms to allow search across namespaces
- Boost query performance by index and query optimizations
- Broaden queries using simple inference
Approach: Semantic search via distributed agent network

- Each agent commits to an ontology and indexes some information source(s)
- Translation agents bridge differing ontologies
- Search in parallel
- Search control: time outs, loop prevention, number of results
- Can take advantage of new OWL-S and matchmaking services
Ontology mapping / query mapping is critical for semantic search

The ideal way to do it is fully automated, fully provable
  - This is very hard, we need a pragmatic stopgap

LOM: Lexicon-based ontology mapper
  - Simple heuristic method, but useful today
  - Semi-automated: suggests mappings of terms

Lexical methods for matching terms in two ontologies
  - Match whole terms
  - Match words as parts of terms, omitting stop words and term-former characters
  - Match using WordNet synsets
  - Match ontological types in SUMO & MILO using WordNet mappings
  - Extensible, add your own heuristics
Our recent work focuses on exploring the functions and capacity of one search agent

- Crawling and gathering information
- High-performance search via optimizations
- Search broadening with OWL relations
- Authors can publish their information to ASCS to ensure quick indexing
- Subscription of queries and notification

Semantic search engine and LOM are prototype search and translation agents to fit into larger architecture
## Technical Progress: Problems and Resolutions

<table>
<thead>
<tr>
<th>Problems</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical queries are difficult for most users to write</td>
<td>Three GUIs for different classes of users</td>
</tr>
<tr>
<td>Distributed search in open network can be hard to control</td>
<td>Implemented search constraints and load balancing</td>
</tr>
<tr>
<td>Search extended with arbitrary inference rules can be undecidable</td>
<td>Constrained to OWL equivalence, generalization, and inverse relations. Experimented with compiling out inference.</td>
</tr>
<tr>
<td>Automated mapping is very hard</td>
<td>Developed semi-automated heuristic mapping for terms</td>
</tr>
<tr>
<td>Individual search agent performance</td>
<td>Implemented dynamic query optimization. Compute and cache data optimizations</td>
</tr>
</tbody>
</table>
## Technical Progress: Goals

<table>
<thead>
<tr>
<th>Goals</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create semantic search agent</td>
<td>We built two</td>
</tr>
<tr>
<td>Create semantic translation agent</td>
<td>LOM semi-automated utility</td>
</tr>
<tr>
<td></td>
<td>Simple query translation using OWL relations</td>
</tr>
<tr>
<td></td>
<td>Demonstrated hub &amp; spoke semantic integration with SUMO and MILO</td>
</tr>
<tr>
<td>Build foundation for semantic interoperation</td>
<td>Contributed 15 major ontologies to the community</td>
</tr>
<tr>
<td>Characterize the benefit of search that uses a network of agents</td>
<td>Early distributed search and experiments. General version with translation and matchmaking not realized yet</td>
</tr>
<tr>
<td>Demonstrate scalability</td>
<td>Have architecture plausibility experiments. Have systems infrastructure. Not yet tested on massive example.</td>
</tr>
</tbody>
</table>
## Technical Progress: Metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repository size</td>
<td>&gt; 8M triples in one search agent</td>
</tr>
<tr>
<td>Query latency</td>
<td>1-10 sec on standard query suite, 500 MHz PC with 512 MB RAM</td>
</tr>
<tr>
<td></td>
<td>Have kept this constant as repository has grown</td>
</tr>
<tr>
<td>Query latency with inference</td>
<td>Avoided complex inference, used simple OWL relationships</td>
</tr>
<tr>
<td>Ontology mapping precision and recall</td>
<td>Approx 71% and 57% in our experiment</td>
</tr>
</tbody>
</table>
ASCS
Milestones and Accomplishments

GFY01
• DAML version of SUMO upper ontology
• Initial distributed-agent implementation
• Demonstration of ASCS integration with the TekPortal banking product

GFY02
• Public search service, from 10/02 (plucky.teknowledge.com)
• DAML versions of many of our ontologies published

GFY03
• “Power user” GUI to search service (oak)
• Restricted English GUI to search service (ibis)
• Lexical-analysis ontology mapper
ASCS
Milestones and Accomplishments

GFY04

- Authors can publish OWL pages for indexing
- Upgraded from batch to incremental indexing
- Users can subscribe queries
  - Retrieved on user’s schedule or by event (e.g., data changed)
  - Query results emailed to subscriber if results have changed
- Provided Java and Web Services APIs
- OWL versions of SUMO, MILO, and 13 domain ontologies in total
- Submitted crawler, search engine, three GUIs and ontology mapper to SemWebCentral
- Ontology mapper did exceptionally well at I³CON with large ontologies
Availability of results

- Crawler, search engine, three GUIs and ontology mapper are available on SemWebCentral
- 7x24 public search at {plucky | oak | ibis}.teknowledge.com

Papers


Users

- SRI used ASCS as part of its ARDA AQUAINT I project
- Oakland Univ. & Univ of Georgia use ASCS for query refinement
- Many software and ontology downloads from our public site but we rarely hear what they do with them
## ASCS

### Remaining Issues

![DARPA](DARPA.png) ![DOML](DOML.png)

<table>
<thead>
<tr>
<th>Issue</th>
<th>Remediation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Early adopters are needed!</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Our development is done as of this meeting</strong></td>
<td></td>
</tr>
<tr>
<td>Full translation of our ontologies, including axioms</td>
<td>Now possible with SWRL-FOL</td>
</tr>
<tr>
<td>Experimental validation of search with multiple distributed agents</td>
<td>Will need a future project</td>
</tr>
<tr>
<td>Deployment issues for ASCS, like authentication and security</td>
<td>SemWebCentral community and/or another project</td>
</tr>
<tr>
<td>Authors must participate in semantic interoperation</td>
<td>Authors need to provide linkages to their ontologies or adopt standardized ontologies</td>
</tr>
</tbody>
</table>

**Early adopters are needed!**
ASCS Summary

• ASCS semantic search merges data from many sources to return answers (variable bindings), not URLs. Semantic search != Google
• Publish function keeps agent’s repository current, reduces reliance on crawling
• Subscribed queries ensure interested parties are notified about important changes in a timely manner
  – Example: semantically based interest list notification of new information for intelligence analysts
• LOM is a practical, extensible approach to ontology mapping
• An enterprise intranet for dynamic interchange between information producers and consumers would be an ideal early adoption application
• The distributed agent model scales up when the centralized-repository model’s capacity is exceeded