Agent Semantic Communications Service (ASCS)

Teknowledge

John Li, Allan Terry

November 2004

ASCS Overall Program Summary

The problem:

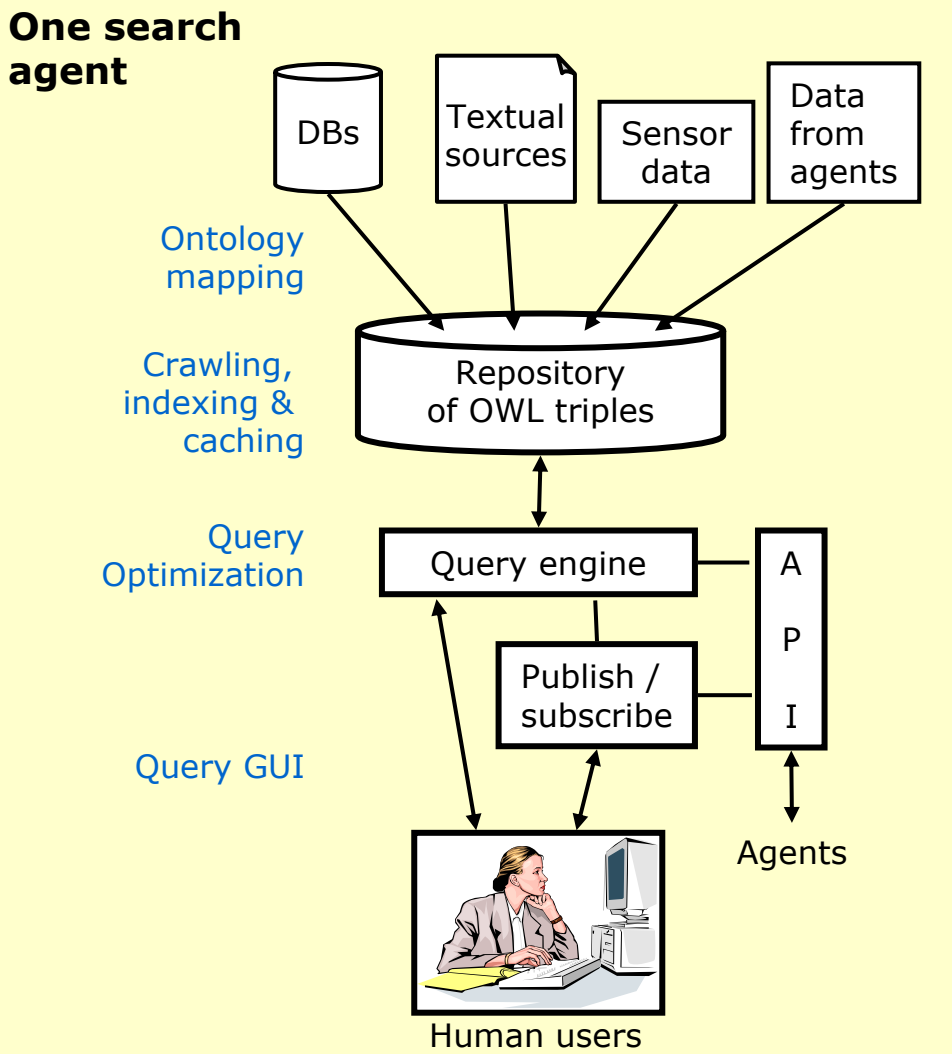
- Leverage semantic markup for integration of heterogeneous data sources

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

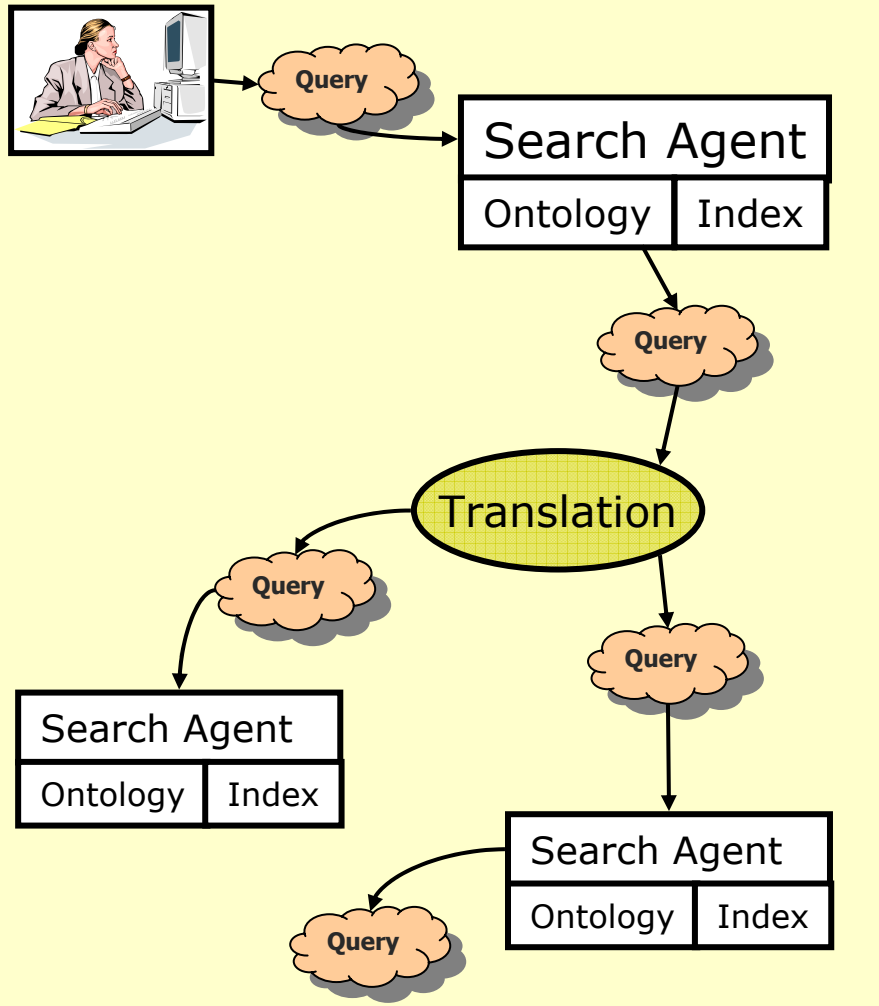


TEKNOWLEDGE

ASCS
Technical Problem and Approach 1

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**



ASCS
Technical Problem and Approach 2

- 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

TEKNOWLEDGE

ASCS
Technical Problem and Approach 3

- 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

TEKNOWLEDGE

Technical Progress: Problems and Resolutions

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

Technical Progress: Goals

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

TEKNOWLEDGE

Technical Progress: Metrics

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

TEKNOWLEDGE

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

TEKNOWLEDGE

ASCS
Milestones and Accomplishments

GFY04

- Authors can publish OWL pages for indexing
- Upgraded from batch to incremental indexing
- Users can subscribe queries
  - Retried 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 I^3CON with large ontologies

TEKNOWLEDGE

ASCS Transition/Handoff

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

- Li, J., Pease, A., Barbee, C. 2001. "Performance of Semantic Search." Teknowledge Technical report, June 15, 2001.
- Li, J., Pease, A. 2002. "Building the DAML Semantic Search Services." Teknowledge Technical report, June, 2002.
- Li, J. 2004. "LOM: A Lexicon-based Ontology Mapping Tool" In *Proceeding of the Performance Metrics for Intelligent Systems (PerMIS '04)*. Information Interpretation and Integration Conference (I^3CON), Gaithersburg, MD.

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

TEKNOWLEDGE

ASCS
Remaining Issues

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

Early adopters are needed!

TEKNOWLEDGE

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