DAML Tools for Intelligent Information Annotation, Sharing and Retrieval

UMBC
Johns Hopkins University Applied Physics Lab
MIT Sloan School

May 2004
Our Research Agenda

• We have three themes, each lead by one team member and with participation from all
  – Semantic web & agents (UMBC)
    Tim Finin, Anupam Joshi, Yun Peng
  – Semantic web & information retrieval (JHU)
    Jim Mayfield
  – Semantic web & rules (MIT)
    Benjamin Grosof

• Our approach is largely experimental -- building prototypes & tools rather than proving theorems
Theme #1

OWL will enable agents in open, dynamic environments (e.g., agentcities) to share knowledge and manage security, privacy, trust and commitments.

- Working in multiagent systems
- Supporting pervasive systems
- Supporting trust and security in web services
Travel Agent Game in Agentcities

Motivation
- Market dynamics
- Auction theory (TAC)

Features
- Open Market Framework
- Auction Services

Technologies
- FIPA (JADE, April Agent Platform)
- Semantic Web (RDF, OWL)
- Web (SOAP, WSDL, DAML-S)
- Internet (Java Web Start)

Ontologies
- http://taga.umbc.edu/ontologies/
  - travel.owl – travel concepts
  - fipaowl.owl – FIPA content lang.
  - auction.owl – auction services
  - tagaql.owl – query language

Motivation
- Market dynamics
- Auction theory (TAC)
- Semantic web
- Agent collaboration (FIPA & Agentcities)

Features
- Open Market Framework
- Auction Services
  - OWL message content
  - OWL Ontologies
  - Global Agent Community

Features
- Direct Buy
  - Report Direct Buy Transactions
  - Bid
  - CFP
  - Report Auction Transactions
  - Proposal
  - Report Travel Package

Ontologies
- http://taga.umbc.edu/ontologies/
  - travel.owl – travel concepts
  - fipaowl.owl – FIPA content lang.
  - auction.owl – auction services
  - tagaql.owl – query language

Multimedia
- Owl for negotiation
  - Owl for contract enforcement
  - Owl for representing and reasoning
  - Owl for service descriptions
  - Owl as a content language
  - Owl for modeling trust

http://taga.sourceforge.net/
OWL in Pervasive Computing

OWL ontologies for modeling & reasoning about contexts.

OWL for privacy policies to control sharing of user info.

OWL ontologies for privacy policies to control sharing of user info.

http://pervasive.semanticweb.org/
Rei Policy Applications: past, present & future

- Coordinating access in *supply chain management system*
- Authorization policies in a *pervasive computing environment*
- Policies for team formation, collaboration, information flow in *multi-agent systems*
- Security in *semantic web services*
- Privacy and trust on the *Internet*
- Privacy in a *pervasive computing environment*
Theme #2

OWL will be integrated with other knowledge representation paradigms for real world reasoning, e.g. rule based systems and bayesian belief networks

– F-OWL reasoner
– Bayes OWL
– OWL and rules
F-OWL Reasoner

• F-OWL is an OWL reasoner implemented in XSB’s Flora-2 system.


• Features:
  – Supports RDF and OWL-Full
  – Supports RDF/N-Triple query
  – Supports Dynamic Import
  – Provides a Java API
  – Tested with the RDF and OWL test cases, See [http://www.w3.org/2003/08/owl-systems/test-results-out.html](http://www.w3.org/2003/08/owl-systems/test-results-out.html)
Bayes Owl

- Probabilistic extension for uncertainty in ontologies:
  - Extend OWL for probabilistic annotation
  - Translate OWL ontology to Bayesian network (OWL-BN)
  - Probabilistic mappings between individual OWL-BNs
  ⇒ Treat ontology reasoning as Bayesian inference

- Plan for remainder of 2004
  - Complete translation rules for RDF constructors and formalize translation rules based on OWL semantics
  - Complete construction procedure for conditional probability tables
  - Complete translation algorithm, OWL → OWL-BN
  - Preliminary investigation on probabilistic concept mapping between OWL-BNs
MIT rules work

1. How to combine rules with ontologies
2. Uses of rules for services, especially in e-commerce, e.g. for
   • E-contracts
   • Financial reporting and information
   • Trust policies
3. Analysis of business case, value, strategy, and requirements for rules and services
New MIT Research Results

• **Trust Policies in Finance using Rules**
  Application scenarios; analysis of business case, value, strategy, and requirements

• **Representing Process Handbook Ontologies using Rules** – including OO-style default inheritance
  Early prototype; new design improves scaleability; analysis of business case, value, and strategy (e.g., to use Legacy OO process framework knowledge)

• **Refining SweetDeal E-Contracting approach** (rules+ontologies)
  Concepts, paper, presentations

• **SWRL V0.5, V0.6** – co-authored; co-led Joint Committee rules effort
  Adds horn rules and RuleML syntax to OWL, in tightly integrated fashion.
  Reports; W3C member submission.

• **RuleML V0.85** – co-authored; co-led RuleML Initiative
  Adds improved syntax support for OWL, RDF, and object-oriented -- in coordination with development of SWRL
MIT Plan: next 6 months

• Continue the focus areas listed earlier

• New phase effort on tools for rules
  – New prototyping of SweetRules for RuleML/SWRL translation/interoperability, inferencing & authoring/testing
  – Support OWL, XSB, Jess, other systems
  – Pluggable architecture to integrate tools from multiple groups in and out of DAML
  – Open source on SemWebCentral
Theme #3

OWL will enable IR systems to work with documents which contain knowledge encoded in semantic markup.

- Swoogle indexing and retrieval system for semantic web documents
- Swangling semantic web documents to prepare them for other conventional IR systems like Google.
Why use IR techniques?

• To retrieve over the structured & unstructured parts of a semantic web document
• To prepare for the appearance of text documents with embedded SW markup
• To harness web infrastructure such as Google
• To take advantage of the unique and useful properties of IR techniques
  – e.g., ranking matches, computing the similarity between two documents, relevance feedback, etc.
http://swoogle.umbc.edu/

Swoogle is a crawler based search and retrieval system for semantic web documents.
Harnessing Google

• Google started indexing RDF documents some time in late 2003
• Can we take advantage of this?
• We’ve developed techniques to get some structured data to be indexed by Google
• And then later retrieved
• Technique: give Google enhanced documents with additional annotations containing Swangle Terms™
Swangle defined

swan·gle

Pronunciation: ‘swa[ng]-g&l
Function: transitive verb
Inflected Forms: swan·gled; swan·gling /-g(&-)li[ng]/
Etymology: Postmodern English, from C++ mangle,
Date: 20th century

1: to convert an RDF triple into one or more IR indexing terms
2: to process a document or query so that its content bearing markup will be indexed by an IR system

Synonym: see tblify

- swan·gler /-g(&-)l&r/ noun
Swangling

• Swangling converts a triple into seven word-like terms
  – One for each non-empty subset of the triple components with
    missing elements replaced by a “don’t care” URI
  – Terms generated by a hashing function (e.g., MD5)

• Swangling an RDF document means adding triples for
  swangle terms.
  – Triples might be explicit ones or added by inference
  – Resulting RDF document can be indexed and retrieved via
    conventional search engines like Google

• Allows one to search for a semantic web document
  with a triple that claims “Osama bin Laden is located
  at X”
A Swangled Triple

<s:SwangledTriple>
  <s:swangledText>N656WNTZ36KQ5PX6RFUGVKQ63A</s:swangledText>
  <rdfs:comment>Swangled text for
    [http://www.xfront.com/owl/ontologies/camera/#Camera,
     http://www.w3.org/2000/01/rdf-schema#subClassOf,
     http://www.xfront.com/owl/ontologies/camera/#PurchaseableItem]
  </rdfs:comment>
  <s:swangledText>M6IMWPWIH4YQI4IMGZYBGPYKEI</s:swangledText>
  <s:swangledText>HO2H3FOPAEM53AQIZ6YVPFQ2XI</s:swangledText>
  <s:swangledText>2AQEUJOYPMXWKHZENIJS6PQ6M</s:swangledText>
  <s:swangledText>IIVQROAYRH6GGRZDFXKEEB4PY</s:swangledText>
  <s:swangledText>75Q5Z3BYAKRPLZDLFNS5KKTMOY</s:swangledText>
  <s:swangledText>2FQ2YI7SNJ7OMXOXIDEEEE2WOZU</s:swangledText>
</s:SwangledTriple>
Summary: 2004 Deliverables

- Software tools on semwebcentral.org
  - Swangling tools, F-owl, Rei, SweetDeal, BBN tools, …

- OWL ontologies
  - For FIPA standards
  - For security and trust
  - For swangling
  - SOUPA: Standard Ontologies for Ubiquitous and Pervasive Applications

- Hosted services
  - Swoogle

- Contributions to standards
  - RuleML, SWRL, SWSI, …