DAML Agents for Science Data Web

Sanda Mandutianu
Caltech, Jet Propulsion Laboratory
sanda.mandutianu@jpl.nasa.gov

NASA scientists have to extensively search through online mission datasets, reports, papers or other documents. The content of these documents, in the case of images for instance, are very opaque to the traditional searching engines. Looking for a particular dataset is usually done by searching for keywords, which might occur on a web page that refers to a particular dataset. This method is quite inaccurate and non-reliable. If the dataset were associated with a clearly specified description on its content, significantly more adequate retrieval would be possible. We aim to evolve the web

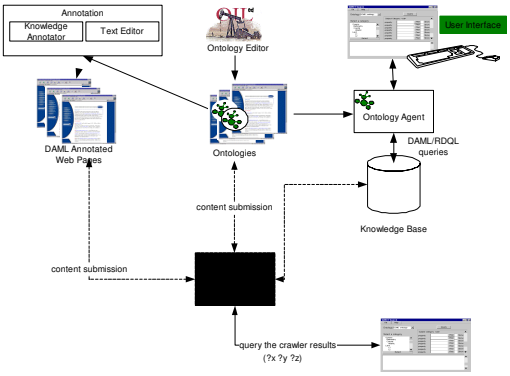

Fig. 1 General Architecture

document repository to a knowledge base. Thus we achieve more flexible and precise knowledge representation and retrieval.

We are exploring the agent technology and its role in the semantic web. We describe an agent approach for science information retrieval based on semantic web concepts:

- RDF/DAML data model;
- Use of ontologies;
- Use of knowledge representation language (DAML+OIL) that support logic inferences;
- Query and inference engine for RDF;
- Agent technology for information retrieval; extract information and share metadata, facilitate query formulation, and evaluate query results.

The system uses metadata collected during the information extraction process to infer additional semantic relationships. This additional information is used to decide what the scope of the search is and to provide more relevant responses to the user.

We are experimenting reasoning based on DAMLJessKB for interpreting DAML documents and adding facts to the knowledge base. We enhance its existing inferential capabilities by applying domain specific rules. The inference process enhances and makes the search much more effective and robust.

The query language is defined in terms of RDF logic. We are experimenting with RDQL –RDF Data Query Language. The queries are formulated using a graphical representation and translated into the internal query language.

We are using the existing NASA Global Change Master Directory (GCMD) as a basis for creating ontologies. GCMD currently uses a vocabulary of 1200 technical terms in the Earth sciences. This list will be expanded to include a thesaurus and incorporated into a larger more general ontology.

The ongoing work is to build the Earth science ontologies, and the inference mechanisms allowing for deductive information retrieval. In this framework, agents will facilitate the information retrieval, the inference processes and the information exchanges.

References

1. DAML, Darpa Agent Markup Language. http://www.daml.org
2. Kopena, J. http://plan.mcs.drexel.edu/projects/legorobots/design/software/DAMLJessKB
3. Lassila,O., and (eds), S.R.R. 1999. Resource Description Framework (RDF) model and syntax specification. W3C recommendation. http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/
4. Global Change Master Directory. http://gcmd.nasa.gov
5. The jena semantic web toolkit RDQL gramma http://www.hpl.hp.com/semweb/rdql-grammar.html