Semantic-Web Technology: Applications at NASA

Naveen Ashish
Research Institute for Advanced Computer Science
NASA Ames Research Center, NS 269/3, Moffett Field CA94035
USA
ashish@email.arc.nasa.gov
http://ic.arc.nasa.gov

Abstract. We provide a description of work at the National Aeronautics and Space Administration (NASA) on building systems based on Semantic-Web concepts and technologies. NASA has been one of the early adopters of Semantic-Web technologies for practical applications. Indeed there are several ongoing (IT) endeavors on building semantics based systems for use in diverse NASA domains ranging from collaborative scientific activity to accident and mishap investigation to enterprise search to scientific information gathering and integration to aviation safety decision support. We provide a brief overview of many applications and ongoing work with the goal of informing the external community of these NASA endeavors.

1 Introduction

We present an overview of several completed and ongoing endeavors at NASA of systems based on semantic web technologies. NASA[1] as an agency has a strategic focus on the four areas of Exploration, Space, Science and Aeronautics. Information Technology(IT) R&D at NASA focuses on the areas on intelligent systems, supercomputing and networking to meet NASA mission needs. Besides, NASA also conducts R&D activities for aviation information management and decision support in collaboration with the Federal Aviation Administration (FAA). The focus of Semantic-Web related work at NASA is not so much on 'fundamental' Semantic-Web issues such as developing languages for ontologies or markup, or developing tools such as for ontology matching, rather the focus is heavily towards applications based on these technologies that can benefit NASA missions. We describe a variety of systems including systems for collaborative knowledge sharing, taxonomy and enterprise search, scientific information gathering, scientific discovery, and aviation information management that have been or are being built to address various NASA (IT) challenges.

[1] http://www.nasa.gov

Dagstuhl Seminar Proceedings 04391
Semantic Interoperability and Integration
http://drops.dagstuhl.de/opus/volltexte/2005/32

2 NASA Semantic Web based Systems

In this section we provide a synopsis of various NASA systems based on Semantic Web technology. The purpose here is not to provide a comprehensive description of any of these systems or discuss the various design and implementation considerations. Rather this is meant to be informational to researchers and practitioners who may be interested in knowing about such application efforts, and then partnering or collaborating with the groups involved.

2.1 SemanticOrganizer

The SemanticOrganizer (SO)[1] is essentially a collaborative knowledge management system for supporting distributed NASA teams. The teams could be working in a variety of domains and tasks ranging from scientific data collection and field activity to accident investigation after mishaps. SO is essentially a customizable, semantically structured repository and provides a common access point for all work products in team tasks. Members of a team can upload, store, and query a wide variety of information in different formats into the repository. As shown in Fig 1 the SO maintains not only data items but also relationships between data items and between data items and people and groups. SO is one of the large Semantic Web applications at NASA having over 500 users and over 45,000 information nodes that are connected by over 150,000 links. There are currently over half a million RDF style triples and over 25 groups. SO allows customization of applications in that concepts for a particular application can be derived from groups of concepts formed and synthesized from more generic concepts. There is a master ontology in the SO system of generic concepts such as ‘person’, ‘team’, ‘image’ etc. and groups of related concepts can be formed (hierarchically) based on different tasks and applications. SO has been fielded in at least 2 major applications at NASA. One is in the scientific domain where SO is being used by the Early Microbial Ecosystems Research Group (EMERG) which is a team of 35 biologists, chemists and engineers spread over 8 locations. In another domain SO has also been fielded as the InvestigationOrganizer (IO) for use in accident and mishap investigations, which have typically involved distributed teams of investigation personnel and engineers operating at different sites.

Valuable lessons have been learned from the experience of fielding this Semantic Web application for the actual uses described above. While a detailed discussion is not possible in this paper, we present a brief summary excerpted from [1]

- The complex network structured storage models present challenges to users who are more used to viewing information with simple or no relationships between various pieces.
- There is a need for both ‘tight’ and ‘loose’ semantics depending on the application and the users.

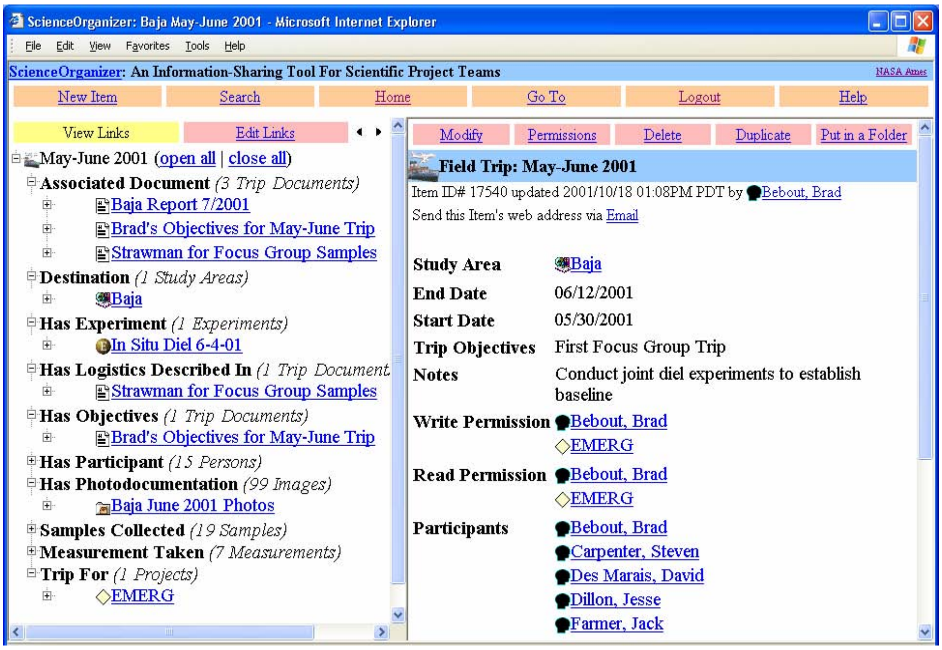

Fig 1. SemanticOrganizer

- Principled ontology evolution (over the course of an application) is difficult to sustain
- Navigating a large semantic network (5000 nodes, 30000-50000 semantic connections) is problematic for end users
- Automated metadata extraction is critical as users or administrators cannot manually annotate or enrich the data coming in so that it may be semantically related and queried.

More information about the SemanticOrganizer may be obtained at http://ic.arc.nasa.gov/sciencedesk/. The points-of-contact(POCs) for the project are Dr. Richard Keller (keller@email.arc.nasa.gov) and Dr. Daniel Berrios (berrios@email.arc.nasa.gov) both at NASA Ames Research Center.

2.2 NASA Taxonomy

The NASA Taxonomy is an effort on building an enterprise-wide taxonomy for NASA. The intended use of the taxonomy is to help NASA personnel – scientists and engineers find information, through the use of intelligent search, browsing, and navigation systems that utilize the taxonomy. The NASA Taxonomy development, funded out of the NASA CIO office, was led by Taxonomy Strategies Inc.[2] A top-down approach to taxonomy development was followed. Subject matter experts from various areas such as unmanned space mission development, mission technology development, engineering configuration management and product data management systems

[2] http://www.taxonomystrategies.com

were extensively consulted. Also, input was sought from managers of IT systems and project content for manned missions. The taxonomy development was done in accordance with industry best practices such as hierarchical granularity, polyhierarchy, mapping aliases, existing standards, and modularity. The taxonomy comprises of different chunks or "facets" reflecting organization and division as NASA such as based on discipline (various NASA scientific and engineering disciplines), NASA locations, NASA organizations etc. An intelligent browsing and navigation capability for NASA enterprise information using the taxonomy was also developed. This capability was developed based on the Seamark semantic navigation system from Siderean Inc.[3] A snapshot of the browsing and navigation capability in action is shown in Fig 2.

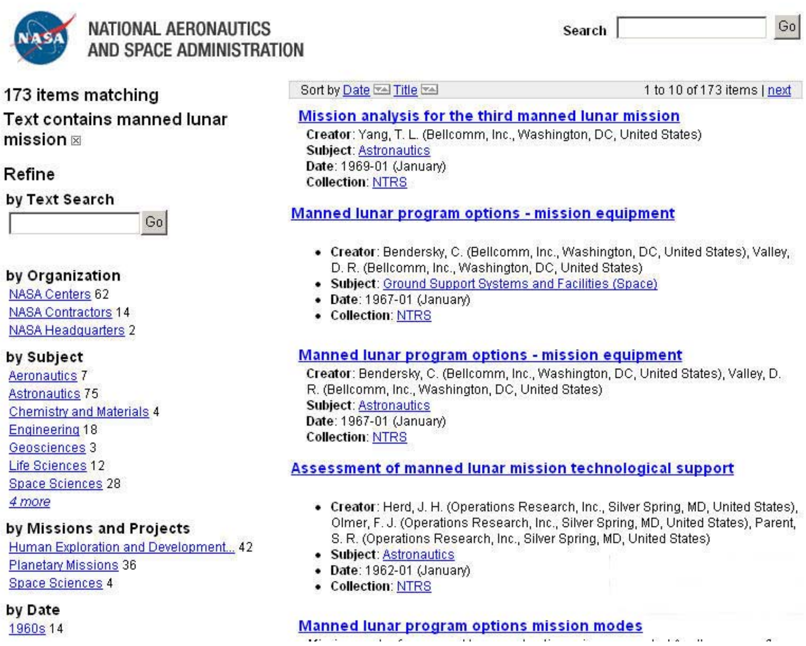

Fig 2. Search, browse and navigate

More information is available at http://nasataxonomy.jpl.nasa.gov/ . The POC for NASA Taxonomy is Ms. Jayne Dutra (Jayne.E.Dutra@jpl.nasa.gov) at JPL.

[3] http://www.siderean.com/

2.3 SWEET

The Semantic Web for Earth and Environmental Technologies (SWEET)[2]is an effort aimed at making the discovery and access of NASA's Earth Science data products over the Web more effective and intuitive. The focus has primarily been on developing a variety of ontologies for earth system science information sources, earth system science data and earth system science subject domains of interest. A search tool that utilizes these ontologies has also been developed. The ontologies are written in OWL and are decomposed into earth realms (the various "spheres" i.e., atmosphere, ocean, solid earth etc. of the earth realm), substances (particles, chemical compounds etc.), living elements (plants and animal species), physical properties, unites, numerical entities, temporal entities and relationships, spatial entities and relationships, natural phenomena and processes, and human activities. The Postgres ORDBMS has been used for ontology storage given its support for object-oriented features as well as geospatial query processing capabilities. A search tool that utilizes these ontologies and also using Latent Semantic Analysis (LSA) to uncover relationships between (search) terms is being developed.

Information about SWEET is available at http://sweet.jpl.nasa.gov and the POC is Dr. Robert Raskin (raskin@seastar.jpl.nasa.gov) at JPL.

2.4 NASA Discovery Systems Project

In addition to completed or ongoing R&D projects, we also describe an envisioned new NASA research program which is strongly related to semantic interoperation amongst other things. The NASA Discovery Systems initiative is a planned 5 year effort that seeks to revolutionize scientific activity. The scientific discovery process at NASA (and indeed in general) is highly data driven. Scientists and engineers have a significant need to understand the vast data sources that are being created through various NASA technology and projects. The current process to integrate and analyze data is labor intensive and requires expert knowledge about data formats and archives. Current discovery and analysis tools are fragmented and mainly support a single person working on small, clean data sets in restricted domains. The Discovery Systems project will develop and demonstrate technologies to handle the details and provide ubiquitous and seamless access to and integration of increasingly massive and diverse information from distributed sources. New technology that generates explanatory, exploratory, and predictive models, makes these tools easier to use, and integrates them in interactive, exploratory environments that let scientists and engineers formulate and solve increasingly complex interdisciplinary problems is sought. Clearly the importance of semantic technologies cannot be underscored The following are the identified thrust areas that have been identified for pursuit in realizing the Discovery Systems vision:

- Collaborative exploratory environments and knowledge sharingMachine assisted model discovery and refinement
- Machine integration of data based on content
- Distributed data search, access and analysisMore information about the Discovery Systems initiative can be obtained at http://postdoc.arc.nasa.gov/ds-planning/public. The POC for Discovery Systems is Dr. Barney Pell (pell@email.arc.nasa.gov) at NASA Ames Research Center.

2.5 SWIM

System Wide Information Management (SWIM) is an initiative[3] aimed at providing information integration capabilities for systems in the National Airspace System (NAS). The NAS is a complex network of distributed and interconnected information and data sources, simulations, and client programs and end users. There are currently numerous decision support system (DSS) tools that are used to aid in various tasks such as maintaining aircraft separation, gate assignments at airports, take-off and landing scheduling etc. These DSS tools access a wide variety of information from different sources such as airport information, weather information, radar tracking information, information from the ATCs, airport towers etc. Connections to all such information sources are currently stove-piped. SWIM aims to provide an integrated pool of all information which then becomes a single point of access that any DSS tool or intelligent agent in the NAS can tap into.

NASA currently has an ongoing effort to realize the vision of such an integrated information pool based on semantic web technologies. Indeed, for a start, there are already initiatives in place on using standard XML based languages for marking up aviation related data such as the Aeronautical Information Exchange Model (AIXM)[4] and the NAS Information Exchange Language (NIXL). Also ontologies are expected to play a critical role in providing semantic interoperation of such sources. The current focus of SWIM activity is on developing an information sharing and integration prototype for NAS information sources based on ontologies, markup languages and distributed (information) agents.

The POCs for SWIM are Dr. Naveen Ashish (ashish@email.arc.nasa.gov) and Mr. Andre Goforth (agoforth@mail.arc.nasa.gov) at NASA Ames Research Center.

[4] http://www.eurocontrol.int/ais/aixm/conceptual.htm

Conclusions

Clearly Semantic Web based technologies are and will continue to play an important role for several NASA IT projects to serve its mission needs. This report, while no means intended to be comprehensive, provides an overview of such efforts. The external community is encouraged to seek collaborations and relationships with the above or other Semantic Web based initiatives at NASA.

My thanks to Barney Pell, Rich Keller, David Maluf, and Dan Berrios at NASA Ames, Rob Raskin and Jayne Dutra at JPL, Everett Cary at Emergent Space Technologies, Gary Davis at NASA GSFC and Brad Allen at Siderean Systems for their input for this presentation.

References

1. Keller, R., et al. *SemanticOrganizer: A Customizable Semantic Repository for Distributed NASA Project Teams*. in *3rd International Semantic Web Conference*. 2004. Hiroshima, Japan.
2. Raskin, R. *Semantic Web for Earth and Envionmental Terminology (SWEET)*. in *Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data at ISWC 2003*. 2003. Sanibel Island, FL: JPL.
3. Ashish, N. and A. Goforth. *Intelligent Information Fusion in the Aviation Domain: A Semantic-Web Based Approach*. in *AIAA Intelligent Systems Technical Conference*. 2004. Chicago, IL.