Representation and Distribution of Geospatial Knowledge

R. M. Malyankar
Department of Computer Science & Engineering
Arizona State University, Tempe, AZ 85287.
Email: rmm@acm.org
URL: http://www.eas.asu.edu/~gcss/

Introduction

Our project is concerned mainly with knowledge representation and the markup of information for web-related uses, and potential applications of such markup. In particular, we are developing taxonomies of certain subsets of maritime information, using these taxonomies to develop markup vocabularies for these subsets, and developing sample applications (e.g., for information retrieval) using this markup. We focused in the first instance on XML markup representation and applications, while allowing for a future transition to semantic web technologies. The project idea pre-dated OWL (Web Ontology Language) and even DAML (DARPA Agent Markup Language) – at the time the project commenced, RDF was the most recent Web technology related to knowledge representation. Our primary government partner was the U.S.Coast Guard; we have also discussed project products with other federal agencies.

Management Structure and Collaboration Over Time

The project has progressed from taxonomy creation to markup language design and the development of demonstration applications and support tools. Collaboration between the academic and government sides evolved in accordance with this progression, turning out to consist of three phases so far:

- In the initial phase, we worked with the Coast Guard to identify a suitable application domain area which was also within their areas of interest and consistent with specific initiatives which were being planned within their organization. It turned out that they were interested in waterway information transfer at that time, and XML markup of maritime information supported needs in that initiative. This phase largely preceded the submission of the proposal for the project described here.
- In the second phase, we built demonstration products (demonstration ontologies, schemas) for the government partner, working mainly on academic research problems rather than artifacts intended for the direct use of the government partner.
- In the third phase, we are worked on development of software artifacts for the government partner, supported and guided by them.

The management structure evolved in tandem with the stages described above. On the academic side, the project was structured as a small academic research project, with two principal investigators and graduate and undergraduate student assistants. The government side initially designated a liaison who was involved with their waterways initiative as our main point of contact, who helped with obtaining source material and documents and identifying suitable canonical sources of information which we needed during the taxonomy construction phase of the project. As the project entered its third phase, a formal statement of work with deliverables was drawn up covering the artifacts to be delivered to the government partner, and the government liaison was formally appointed as the government's technical representative, consistent with their usual contract management procedures. One principal investigator worked on liaison between the academic and government partner throughout the project. The primary means of interaction was e-mail, accompanied by the occasional telephone conversations, and a few face-to-face meetings. In addition, we interacted at a lower level of intensity with entities other than the Coast Guard who were interested in markup of maritime information; these interactions consisted mostly of discussing markup suitable for the subset of maritime information of interest to them.

Research Value and Accomplishments

The core scientific research area of this project is knowledge representation. Within this area, this project focused on ontological engineering – building domain computational ontologies from source

material, and ontology mapping and merging for different computational ontologies created from different sources. The secondary scientific research areas were integrated information retrieval and vocabulary generation for markup languages and the semantic web. Investigating these research problems in the context of marine transportation system (MTS) information allowed the academic partner to work in a real-life domain with substantial problems – large-scale, time-varying, unstructured, multi-form (text and non-textual), and irregular data, thus grounding the research in real-life issues. For the government, development of the models and schemas for maritime information should be useful in information transfer in the MTS and as a basis for markup standard development.

Concerning accomplishments, the primary goals of this project were knowledge representation and definition of the markup language by construction from first principles (ontologies). Most of the applications and tools which we developed were intended to demonstrate the possibilities of marked up maritime information. Certain supporting tools were also developed, one being a registry for models and XML schemas. This schema management system and registry can be used as a repository for XML schemas; it also possesses certain other functionality useful for schema designers and application programmers. The distinguishing features of this system, as compared to other registry and repository efforts, are its ability to contain information about ontologies as well as XML schemas, and the application of certain techniques based on ontology mapping to the problem of detecting overlaps and conflict between XML schemas. (Our poster in this conference contains more information about this tool.) Given the large number of stakeholders in the MTS, many of whom already have their own independently constructed information models, creating and managing a new single information model or XML schema covering all the diverse sources of data would require a large investment of time and resources and be an extremely complex task for logistical and technical reasons. The registry provides a single repository and search system for these multifarious schemas, and will in the future provide means of restructuring schemas in logical ways. It is widely applicable - the General Accounting Office has identified the proliferation of overlapping and incompatible vocabularies and structures as one of the most important problems in the adoption of XML in the Federal government, and mentions XML registries (planned and in various stages of implementation) as part of the solution to this problem.

Challenges and Barriers

Most of the challenges encountered during this project were related to time and personnel management and scheduling, especially matching work requirements to the university's academic semester schedules, working around examinations and university breaks, and managing turnover caused by the floating nature of the student population. A second significant issue was the 'hardening' of software, e.g., producing secure and robust code – developing robust software, which can be used with real data and in real-world situations involves much more than design and coding, and requires experienced software programmers. Other issues we sometimes encountered were vocabulary matching, getting academics up to speed on the government domain, and making sure academics and government agreed on terminology – we believe these particular problems were mitigated on this project by the academic investigator's prior exposure to the domain, gained on a previous collaboration. Still, liaison and reporting took more time than originally anticipated. Lastly, in general it is likely that resource availability (especially personnel time) on the government side might be difficult to manage, but we believe we managed to avoid this problem by working on a goal which happened to fit well with an existing agency initiative.

Acknowledgments

Pete Smullen, Ernest Hunt, Mark Lewandowski, Anvith Baddam, Prajakta Nivargi, and Pradnya Dharia deserve special mention for their contributions. Jay Spalding and Kathy Shea (USCG) filled the roles of liaison and technical representative. The work described in this report was partially supported by the United States Coast Guard and partially by the National Science Foundation under grant EIA-9983267. The opinions and views expressed herein are those of the author and do not represent official views of the National Science Foundation or the United States Coast Guard.