Using Semantic Web Methods for Distributed Learner Modelling

Mike Winter[1], Christopher Brooks[1], Gord McCalla[1], Jim Greer[1] and Peter O'Donovan[2]
University of Saskatchewan
ARIES Laboratory
Saskatoon, SK., Canada S7N 5A9
[1]{mfw127, cab938, mccalla, greer}@cs.usask.ca
[2]peo499@mail.usask.ca

Abstract

Developing a learner model containing an accurate representation of a learner's knowledge is made more difficult in distributed learning environments where the learner uses multiple applications and resources to accomplish learning tasks. To help reduce this difficulty we describe a semantic web approach to representing student models based on distributed student data. We also present a proposal for revising those student models based on arbitrary, web-based learner actions.

1 Introduction

Current online learning is described as often taking place in an 'adaptive learning community' [Gaudioso and Boticatio, 2003] in which online learners use a wide variety of resources to help them perform their problem-solving tasks. These resources include a wide variety of web-pages, instant messaging, online discussion and peer-help tools. In this paper, we present an integrated learner modelling architecture using RDF, RDFS and SOAP that effectively stores and transmits learner information from multiple sources.

The outline of this paper is as follows: Section 2 describes the development of a RDF/RDFS based learner model for a first-year computer science class and the use of the Massive User Modelling System (MUMS) [Brooks *et al.,* 2004] which allows the collection of learner modelling information from diverse application sources for use by our learner models. Section 3 describes how the integrated system is currently deployed for hundreds of students while Section 4 discusses our future goals including the use of information retrieval techniques with MUMS to update our learner models.

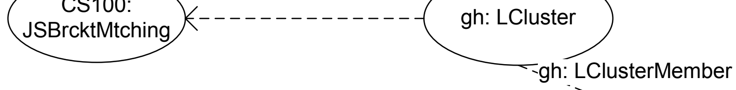
2 Granular Learner Models with RDF and RDFS

The first challenge for a learner modelling system in a distributed learning environment is to effectively attach meaning to the learner data it is receiving. Using RDF to model learners has many advantages for this task. First, RDF is a well-specified semantic data model that can be easily serialized between systems, allowing easy sharing of learner models and learner information between interested components. Second, popular RDF packages such as Jena[1] allow for the easy manipulation of RDF graphs, including reasoning capabilities that allow a modelling component to make inferences regarding learners over multiple ontologies. Finally, RDF is able to refer to an arbitrary number of ontologies within a single graph. This allows a student modelling component to accurately model many different aspects of a learner by combining statements that use different ontologies in the same graph. The learner modelling component that we have developed uses multiple RDF schema (RDFS) ontologies to define the classes and relationships contained in RDF graphs that act as student models. The two main ontologies we have encoded in RDFS to express learner model information are listed below.

1. **Granularity Hierarchies.** To define concept maps for the domains being studied by the learners in our system, we use the granularity hierarchy formalism which is an extended semantic network that defines both specialization and aggregation relationships between topics [McCalla *et al.*, 1992]. In the granularity hierarchy formalism, a **K-Cluster** represents a particular semantic aggregation of topics while an **L-Cluster** represents a particular semantic specialization of a topic. A topic can have more than one K-Cluster and/or L-Cluster relationship. The major advantage granularity hierarchies provide in terms of domain modelling is the ability to represent a domain at multiple levels of detail simultaneously. Currently, a domain map has been developed using this method that completely models the topics within a first-year Computer Science course at the University of Saskatchewan (Figure 1). This domain map contains over five-hundred topic nodes and thousands of granularity hierarchy relationships between them.

2. **Ontology of Learning Outcomes.** For purposes of learner modelling, a concept map is not enough; the knowledge of particular learners must be added to instantiations of the map. Student knowledge of a topic can be represented as an increasing degree of proficiency as detailed by Bloom's taxonomy [Bloom, 1956]. We have developed an

[1] http://jena.sourceforge.net



RDFS version of the taxonomy encompassing its eight levels of knowledge ranging from **Knowledge** (basic recall) to

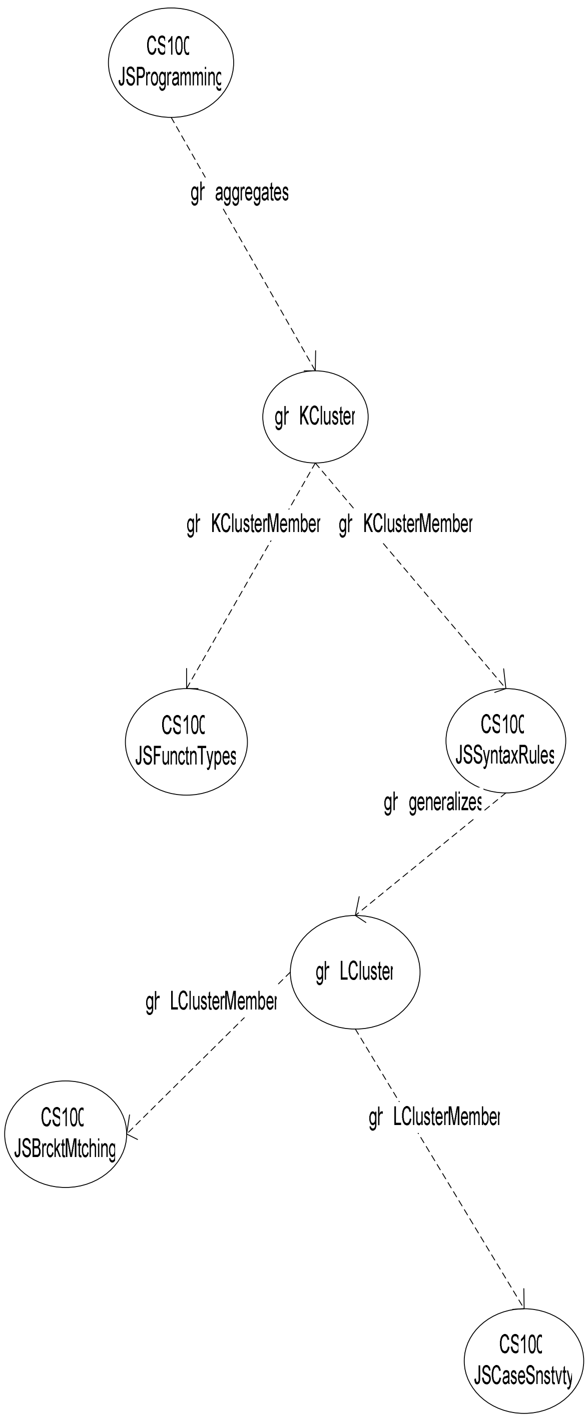


Figure 1. Section of a Granular Learner Model

Evaluation (assess and contrast). To build learner models of real learners, quizzes have been developed for an online version of the first-year Computer Science course modeled above. Answers to questions are annotated with the topic(s) that they cover and the Bloom's level of knowledge that will be demonstrated if they are answered correctly.

2.1 MUMS: Collecting Distributed Learner Modelling Data

MUMS is an event system designed to collect and distribute notifications of user actions from the applications where they happen to interested third parties [Brooks *et al.*, 2003]. Applications that generate events are called **Producers**, applications that receive events are called **Consumers** or **Modellers,** and the application that routes messages from Producers to Consumers is called a **Broker. Filters** act as intermediaries between the Broker and Consumers, and provide miscellaneous security, routing and reasoning services. Consumers send event subscription requests (or queries for archival user information) to the Broker and then receive user events as they happen in real-time.

There are three design principles underlying the MUMS system: interoperability, extensibility and scalability. Interoperability is necessary because of the diversity of applications that are involved in generating and consuming user events. Interoperability is ensured in MUMS in two ways. The first is by an implementation in of the Web-Services Events (WS-E) [Catania *et al.*, 1985] specification in WSDL, with a SOAP binding. This specification details the type and format of messages that should be sent between components in a web-service based event system. The use of web-service standards enables a language and system-neutral transport protocol for event messages. The second way interoperability is ensured in the MUMS system is by the use of RDF statements as the payload of each notification. RDFS or OWL schemas provide the ontologies for the RDF payloads and allow the various applications on the MUMS network to have a shared semantic understanding of user events. Extensibility is an important feature of the MUMS network because Modellers and Producers are volatile and may be online or offline at any given moment. Extensibility is provided for by the subscription mechanism of the MUMS system. Consumers send subscription requests in RDQL to the broker regarding arbitrary events (usually involving certain users or groups of users) and receive events generated by applications across the MUMS network that match the subscription request. Another mechanism ensuring extensibility is the lightweight API that easily allows components to talk to the MUMS network. Scalability is ensured by the distributed nature of the network and the clustering capabilities inherent in the router design.

3 The Integrated Modelling Network

The learner modelling component we have developed is a consumer on the MUMS network, with a separate instantiation of the domain model for each learner. As the MUMS events are encoded in RDF, they are simple for the modelling component to either add directly on to the existing RDF learner model or to use them as the input for inference. Currently, the learner modelling system is deployed for a first-year Computer Science course at the University of Saskatchewan that has around three-hundred students enrolled, thirty-five of them through an online version of the course. The learner modelling system takes as inputs the answers to quizzes in the online course, as mentioned previously, as well as the learners' other actions in the online course, such as a reading a lesson or working with an interactive program. In addition, all of the students' activities on an online class discussion board are sent over the MUMS network and received by the learner modelling component. A MUMS-enabled web proxy is also available for use in research studies. While we are just starting to build the learner models for the first time, the combination of our RDF-based user messaging system and our RDF learner models has been effective in combining distributed sources of learner information into coherent and accessible learner models.

4 Next Steps: Using Information Retrieval Techniques for Student Model Updating

Once the events about a student's behaviour have been transmitted to a student modelling system by the MUMS network there still exists the difficult problem of determining what relevance those events have in relation to its understanding of the learner's knowledge and plans. One already-implemented approach to translating events was discussed in the last section where the answers to quiz questions have pre-determined mappings to learner knowledge assessments. However, the MUMS network is able to transmit information from any arbitrary application, including ones where learner actions are not pre-analyzed. The remainder of this section will detail a proposed general approach to translating events involving a learner's interaction with text-based resources, such as web pages and message board postings, to appropriate learner model revisions.

Assuming a learner model like that discussed in Section 2, a learner's reading of a web page will have to be translated into an update of the model's understanding of the learner's domain knowledge. The way in which a textual resource view has to be interpreted in terms of the learner's knowledge gain can be further decomposed into two separate problems: determining the topic(s) of the textual resource and determining the amount of knowledge the learner has gained from the resource. One way in which the topic of the web page can be determined is by associating a representative piece of text with each knowledge node in the domain model and then using an appropriate information retrieval technique such as vector scoring to determine which topic the web page is most likely to be about. Determining the knowledge gain of the learner resulting from his/her viewing the text resource is trickier because the gain would vary based on the attention the learner paid to the resource, the quality of the resource, and other factors which would not generally be known to the student modelling system. To gather relevant data for this task, a study is being designed that will use MUMS events from the online course and the associated online discussion board that are hooked into the MUMS systems as producers, generating events each time a learner interacts with them. In addition, any other web-based resources that the users of the study access will generate events on the MUMS network through the MUMS-enabled web proxy. The tests in the online course are divided into pre and post-lesson components, and the resulting change in the learner's knowledge as discovered by the tests can then be correlated with the web resources they have viewed.

Acknowledgments

This work was partially supported by the Government of Canada's NSERC LORNET NCE grant.

References

[Gaudioso and Boticatio, 2003] Gaudioso, Elena and Boticatio Jesus, G. Towards Web-based Adaptive Learning Communities. *Proceedings of the 2003 International Conference on Artificial Intelligence in Education.* Sydney, Aus., IOS Press.

[Brooks et al., 2004] Brooks, Christopher, Winter, Mike, Greer, Jim and McCalla, Gord. The Massive User Modelling System (MUMS). *Proceedings of Intelligent Tutoring Systems 2004.* 2004, To appear.

[Catania et al., 1985] Catania, Nicholas, Kumar, Pankaj, Murray, Bryan, Pourhedari, Homayoun, Vambenepe, William and Wurster, Klaus. Web Services Events (WS-Events) Specification v. 2.0. Hewlett-Packard Company.

http://devresource.hp.com/drc/specifications/wsmf/WS-Events.pdf

[McCalla et al., 1992] McCalla, G. I., Greer, J. E., Barrie, B., and Pospisil, P. Granularity Hierarchies, International Journal of Computers and Mathematics with Applications (Special Issue on Semantic Networks), Vol. 23, pp. 363-375, 1992.

[Bloom, 1956] Bloom, B.S. (Ed.) Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain. New York; Toronto: Longmans, Green, 1956.