Melis, E., McLaren, B.M., & Solomon, S. (2008). Towards Accessing Disparate Educational Data in a Single, Unified Manner. In P. Dillenbourg and M. Specht (Eds.), *Proceedings of the Third European Conference on Technology Enhanced Learning (EC-TEL 2008)*, Lecture Notes in Computer Science 5192 (pp. 280-283). Berlin: Springer.

Towards Accessing Disparate Educational Data in a Single, Unified Manner

Erica Melis, Bruce M. McLaren, Silvana Solomon

German Research Institute for Artificial Intelligence,
66123 Saarbrücken, Germany, melis@dfki.de

Abstract. Educational researchers need to exchange and compare their learner-interaction data in order to benefit the learning science community as a whole. In order to support this, we propose accessing data in different repositories via a *mediator component* that maps generic queries to the specific format of a target repository. This approach is supported by a common ontology, and we illustrate the beginnings of such an ontology. We are in the early stages of developing this concept but show its promise by discussing how it can be applied to repositories of disparate educational data, such as collaborative learning interactions and cognitive tutor data.

1 Introduction

A key problem for educational researchers today is sharing and exchanging their learner-interaction data. In order to compare results across studies and across educational systems, it is important to have share data across the studies and systems. We investigated how we can access educational data from different repositories that rely on a variety of perspectives and scenarios, including technology-enhanced learning in lab and classroom experiments, inquiry learning, collaborative learning, classroom learning, and one-on-one tutoring.

Achieving a common access to log data from different learning environments holds several potential advantages for educational researchers and educational technologists. First, it would allow researchers to share and exchange data freely between their systems in theoretically neutral fashion, enabling more direct comparison between approaches and methodologies. Second, it would help to develop community-wide standards and a common format for educational data. This development of standards builds upon previous efforts of the EU Kaleidoscope community [1]. Third, a natural outgrowth of joint access could be the development of shared analysis tools, such as learning curve and social network analyses.

What are the problems to overcome? A key issue is determining how to connect educational data from different perspectives and at different levels of granularity, taking a cue from principled 'knowledge analyses' that have been done by prominent researchers in cognitive science and artificial intelligence [2, 3]. For instance, an intelligent tutor collects data at the cognitive level, while a collaborative learning system collects data at the social interaction level and each has different requirements

Melis, E., McLaren, B.M., & Solomon, S. (2008). Towards Accessing Disparate Educational Data in a Single, Unified Manner. In P. Dillenbourg and M. Specht (Eds.), *Proceedings of the Third European Conference on Technology Enhanced Learning (EC-TEL 2008)*, Lecture Notes in Computer Science 5192 (pp. 280-283). Berlin: Springer.

for data storage, format, and analysis. Moreover, although our main interest focuses on student actions, other information must be logged, standardized, and correlated in particular, contextual information. Contextual information includes data about how the educational software responds to student actions and data that are obtained during system use, such as from questionnaires.

There are two principled ways to achieve this goal: (1) coalesce or translate data from different educational data repositories into a single common repository or (2) access the data in different repositories via a mediator that maps generic queries to the specific format of the target repository. The second approach has the advantage of (1) allowing data sources to remain in their original form, avoiding constant translating and copying of data to a central store, and (2) accessing data through a web service.

Necessary steps to support this approach include a formalization and implementation of the ontologies of the different repositories and the development of a common 'umbrella' ontology into which the separate ontologies can be mapped. We are in the early stages of developing this concept but demonstrate its promise by showing how it can be applied to repositories of disparate educational data, such as collaborative learning interactions and cognitive tutor data.

2 Access to Distributed Log Data Repositories

Our mediator approach is based on past work reported in [4]. The mediator architecture allows an application to retrieve objects or data from heterogeneous repositories. A "mediator component" accepts queries formulated in a uniform query language, translates them into repository-specific queries, and passes them to the corresponding repository (see Fig. 1). A 'wrapper' is used with each repository, containing the specification of the ontology of the repositories knowledge (as an OWL definition) and the mapping to the terms of a common ontology. The wrapper translates queries from the common language/ontology into the language of the repository using the mapping. For the translation of queries, we use an ontology-based query-rewriting method. It queries a repository according to the specific commands of the repository; it transfers the query results of the repositories (e.g., URIs) to the application it serves.

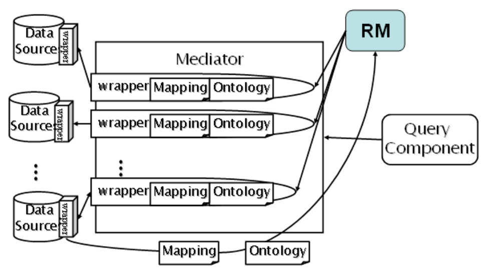

Fig. 1: The mediator architecture

The mediator approach leaves us with the questions "How do we use the mediator technique to query user log data?" and "How do we translate the log data ontologies?" In this paper, we concentrate on the second question, because it must be answered before the implementation of the mediator. The steps towards the translation include (1) a formalization of the ontologies of the repositories, (2) the development of a common ontology, and (3) the development of the mappings.

Melis, E., McLaren, B.M., & Solomon, S. (2008). Towards Accessing Disparate Educational Data in a Single, Unified Manner. In P. Dillenbourg and M. Specht (Eds.), *Proceedings of the Third European Conference on Technology Enhanced Learning (EC-TEL 2008)*, Lecture Notes in Computer Science 5192 (pp. 280-283). Berlin: Springer.

The transformations that the mediator requires to work with log data ontologies involves a complex ontology which needs to describe not only learning objects but also the UserLogActions and Events. It also requires that we rewrite all the mapping instructions in the XML-based ontology mapping language.

3 Ontologies for the Different Log Data Repositories

We built ontologies for five repositories/tools by analysing the tools' logged data and, when provided by the log data/system owner, some schema specifications for this data (DTD, XSD, databases, etc.). We used OWL (Web Ontology Language) language for the representation of ontologies. OWL is designed for use by computational applications but at the same time is human readable. It was developed to augment the facilities for expressing semantics provided by XML, RDF, and RDF-S. Since OWL is based on XML, it can be easily exchanged between different types of computers using different operating systems and application languages. We modelled the ontologies with the help of Protégé[1] [7]. The Protégé-Frames editor enables users to build and populate ontologies that are frame-based, in accordance with the Open Knowledge Base Connectivity protocol (OKBC). In this model, an ontology consists of a set of classes organized into a subsumption hierarchy to represent a domain's salient concepts, a set of slots associated with classes to describe their properties and relationships, and a set of instances of those classes – individual exemplars of the concepts that hold specific values for their properties.

The systems/formats for which we built the log data ontologies span the gamut from collaborative learning technologies to inquiry learning systems to intelligent tutoring systems. The specific systems we evaluated and created ontologies for are: Digalo[2], ActiveMath [6], the PSLC DataShop [5], GSIC Valladolid [6], and a Demonstrator from Grenoble [7]. After analysing the schemas and log file samples provided by the owners of these various systems, we built an ontology for each data format with Protégé (The ontologies can be downloaded from http://www.noe-kaleidoscope.org/group/datarep/links_page.html). Our next step was to analyse the requirements of the five ontologies and map them to a single, common ontology.

4 Common Ontology

When we refer to the *common* ontology we mean common for the group of repositories whose data/ontology could (somehow) map onto the shared ontology. The goal of the common ontology is to support the construction of queries that can be forwarded to the five log data repositories (or more that could be added) via a mediator and to interpret their responses. The components of the ontologies that cannot be mapped to the common ontology are system-specific concepts that have no representation in the common ontology. These unmatched elements will be analysed in the future.

[1] http://protege.stanford.edu/
[2] http://dito.ais.fraunhofer.de/digalo/webstart/index.html

Melis, E., McLaren, B.M., & Solomon, S. (2008). Towards Accessing Disparate Educational Data in a Single, Unified Manner. In P. Dillenbourg and M. Specht (Eds.), *Proceedings of the Third European Conference on Technology Enhanced Learning (EC-TEL 2008)*, Lecture Notes in Computer Science 5192 (pp. 280-283). Berlin: Springer.

In Fig. 2 the top-level structure of the common ontology is depicted. For instance, the Action class is connected with the Session class through the relation action_in_session (represented here by an arrow between the two classes). These classes and relation have mappings to four of our five separate ontologies (only Digalo does not have an equivalent). Likewise, the other concepts and relationships in the common ontology have been mapped to our five ontologies, where possible. Our next step is to experiment with how our mediator allows us to access the data of the separate repositories through common queries.

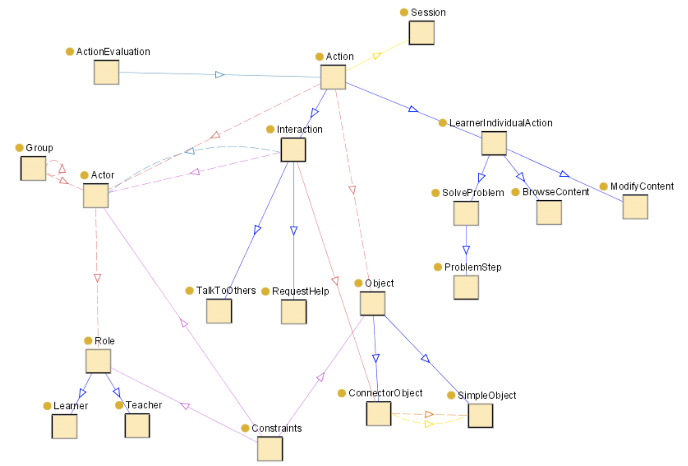

Fig. 2: The high-level structure of the common ontology

Acknowledgments. We would like to thank A. Harrer, R. Baker, J.-M. Adam, J.P.David, A.Martínez Monés, and T. de Jong for their support. This work was funded by the Kaleidoscope NoE.

References

[1] Kaleidoscope. Kaleidoscope Project Report D31.2, Project "Interaction Analysis (IA)," Library of Interaction analysis methods, (2005)
[2] Anderson, J.R.: Spanning Seven Orders of Magnitude: A Challenge for Cognitive Modelling. Cognitive Science, 26(1), 85--112. (2002)
[3] Newell, A.: Unified Theories of Cognition. Cambridge, MA. Harvard Univ. Press (1990)
[4] Kärger, P., Ullrich, C., Melis, E.: Integrating Learning Object Repositories Using a Mediator Architecture. Innovative Approaches for Learning and Knowledge Sharing, 1st European Conference on Technology Enhanced Learning: pp. 185--197. (2006)
[5] VanLehn, K., Koedinger, K.R., Skogsholm, A., Nwaigwe, A. Hausmann, R.G.M., Weinstein, A., Billings, B.: What's in a Step? Toward General, Abstract Representations of Tutoring System Log Data. User Modelling 2007: 455--459. (2007)
[6] Martínez, A., de la Fuente, P., Dimitriadis, Y.: Towards an XML-based Representation of Collaborative Interaction. International Conference on Computer Support for Collaborative Learning, pp. 379-384. (2003)
[7] Adam, J.-M.: Demonstrator of an Infrastructure for Collection and Exchange of Experimental Traces, Kaleidoscope Initiative report D7.8.1 (2006)