Flexible Learning Object Metadata

Christopher BROOKS, Gordon MCCALLA, Mike WINTER
Advanced Research in Intelligent Educational Systems Laboratory
Department of Computer Science, University of Saskatchewan
Saskatoon, SK, Canada
cab938@mail.usask.ca, mccalla@cs.usask.ca, mfw127@mail.usask.ca

Abstract. By far the most popular specification for learning objects is the IEEE Learning Object Metadata (LOM) standard. In it are outlined 76 different elements that correspond to pedagogical, technical, and administrative aspects of learning objects. This standard, however, has proven to be ineffective for creating computer adapted dynamic courseware.

This paper outlines some initial research we are doing in acquiring, describing, and using learning object metadata. Instead of the IEEE LOM, we argue for a more flexible approach to both defining and associating metadata with learning objects. By creating domain, educational, and learner characteristic ontologies, content can be dynamically linked to those competencies that are observed in a running e-learning system. This provides for a set of evolutionary metadata, where software agents can inspect multiple metadata instances for a given learning object and reason over them for a particular goal. As more metadata instances are added to the system, agents are expected to be able to provide more accurate reasoning, eventually leading to the dynamic delivery of personalized course content.

Introduction

Perhaps the most widely used and accepted learning object specification is the IEEE Learning Object Metadata (LOM) standard [14]. This standard identifies 76 different aspects by which a learning object can be annotated, and is supported in some way by all major learning object repositories and e-learning platforms. One would think then, that learning objects should be rich with metadata markup, and that the interrogation of such metadata by a content management system could be used to dynamically assemble a course. This, however, is not the case – content management systems are increasingly static, with even relatively simple rules-based sequencing specifications seeing little to no adoption [6].

The ability to dynamically assemble a course from learning objects is an important goal within the educational technology community. Nonetheless, current e-learning standards and specifications are both too restrictive in the variety of metadata they capture, and too lax in how they express the structure of such metadata. Many learning object repositories support only a few of the fields available, and most do not support an external query format (e.g. [17]) which could be used by computer agents to retrieve objects from the repository. The result is that nearly all learning object based courses are created directly by instructional designers, who align content they have explicitly hand crafted for a given educational purpose. This purpose generally includes both an educational outcome (e.g. "understand relational operators in ECMAScript at a level such that the student can apply them to new situations") as well as an educational instruction style (e.g. a particular language, background, or learning style that a student is assumed to have). This makes dynamic delivery difficult, as an instructional designer must create many different versions of a course for the different kinds of purposes he or she hopes to achieve.

Instead, we argue that a more flexible method of associating metadata with learning objects will help in realising the on-demand assembly of courses for different educational purposes. A larger set of well-defined ontologies sufficient for particular purposes should be used instead of a single highly constrained taxonomy of values like the LOM. Further, the ontologies should be marked up in an unambiguous syntax such that they are able to be understood by software agents. This syntax must take into account the kinds of data types that agents are able to manipulate, and must appropriately codify metadata instances to conform to these. Finally, repository and content management software must be able to associate multiple metadata instances with a given learning object, and allow for agents to pick and choose those instances that fit their needs.

This paper is organized as follows; Section 1 outlines specific issues both we and others have had when trying to use the IEEE LOM with software agents. Section 2 outlines in general terms our approach, dubbed the *ecological approach*. Section 3 indicates how we are using semantic web techniques help to enable this approach in real e-learning systems. Finally, section 4 concludes the work by identifying some potential related areas for exploration.

1. Issues with the Learning Object Metadata Standard

The IEEE Learning Object Metadata (LOM) standard [14] provides a format for representing technical, administrative, and pedagogical metadata about a learning resource. While adopted by e-learning vendors and used in many other e-learning specifications (e.g. [18] [13] [10]), its use in actual deployed learning systems is sparse at best. The specification suffers from three main issues. Firstly, to create a conforming LOM document requires only a few of the many available fields be filled in. Friesen provides a compelling example of this in a study of 250 learning object metadata records (chosen from five different projects evenly) where only 36% of the elements were used more than half of the time, with many elements never used at all [11]. Further, the elements used often referred to custom or local vocabularies, a practice that effectively eliminates semantic interoperability. While it has been suggested that automatic metadata generation tools are a potential solution for this, work to date has been less convincing. For instance, in [8] a set of tools were developed to try and automatically generate LOM data by directly data mining the learning object content and the context in which it is being delivered. This worked only for a small set of fields (less than 25%) and proved to be an error prone process where many of the automatically generated fields disagree with the values set by content experts.

In addition to a lack of instance data, many learning object based products released have poor support for the full LOM. Expressed in the Learning Object Metadata Best Practices guide, "Many vendors expressed little or no interest in developing products that were required to support a set of meta-data with over 80 elements…[and the] burden to support 80+ meta-data elements on the first iteration of a product is too great for most vendors to choose to bear". [3] The end result is partial implementation, where many of the more complex fields (which happen to be the most useful for dynamic courseware generation agents, as they relate a learning object directly to domain vocabularies) are discarded, and only the simple fields (such as title, or description) are kept.

While the goal of the LOM was to make data available to both human interpreters (generally teachers and instruction designers) and computer agents, the standard provides several examples of poor data typing, leading to potentially ambiguous situations. Some of our previous work has identified that the version and lifecycle elements of the LOM are generally stored as arbitrary human readable text, and are thus unreliable for automatic processing [5]. Further, Friesen indicates that even those elements that are required by the standard to be strictly data typed are often not, as was the case of vCard contact information, where none of the documents in a 3,000 instance test set were found to conform [11].

In addition to internal issues with applying the metadata standard, there are external issues. In particular, most learning object repositories allow for only a single metadata instance to be associated with a learning object. Anecdotal evidence observed in assigning metadata to a set of computer science learning objects [9] suggested that inter-rater reliability is often quite low, especially when only a few fields are chosen by authors. This appears to be a general trend in educational metadata, whether the resource being described is a tutorial, discussion thread, or other digital artefact. By restricting learning objects to single instances of metadata, repositories are significantly limiting both the quality and quantity of information that can be expressed about a given resource.

2. The Ecological Approach

We are working on implementing an alternative theory of metadata, called the *ecological approach*, to overcome the defficiencies present in the standards based approach. In the ecological approach the e-learning system keeps a learner model for each learner, tracking characteristics of the learner and information about the learner's interactions with the learning objects they encounter. After a learner has interacted with a learning object, the learning object is associated with an instance of the learner model. The information in such a learner model instance can include

- information about the learner, including cognitive, affective, and social characteristics and their goal(s) in accessing the content;
- information about the learner's perspectives on the content itself, including the learner's feedback on the content, the learner's knowledge of the content (as determined, for example, by a test administered during the learner's interactions with the learning object);
- information about how the learner interacted with the content, including observed metrics such as dwell time, number of learner keystrokes, patterns of access, etc.;
- information about the technical context of use, including characteristics of the learner's software and hardware environment;
- information about the social context of use, including links to the learner model instances attached to learning objects previously encountered by the learner.

Over time, each learning object thus slowly accumulates learner model instances that collectively form a record of the experiences of all sorts of learners as they have interacted with the learning object. The collected learner model instances can then be inspected for patterns about how learners interacted with the learning object, for example that learners whose knowledge has been evaluated as weak did not have long dwell times, or that learners with certain cognitive characteristics did well. The sequence of learner model instances for a particular learner forms a "learning trail" through the learning object repository, and this trail can also reveal interesting patterns of success and failure for the learner.

There are an enormous number of patterns that can be found when inspecting actual learner behaviour. The key to finding meaningful patterns is the *purpose* (in the sense of [20]) for which the patterns are sought. Each such purpose places its own particular constraints on what patterns are meaningful, how to look for these patterns, and how to use what these patterns reveal in order to achieve the purpose. Thus, determining whether to recommend a specific learning object to a particular learner may require comparing this learner to other learners on important characteristics and then looking at how similar learners have evaluated (or been evaluated on) the content (and, moreover, the characteristics considered to be important are themselves determined by the learner's own goals). On the other hand, determining whether a learning object is now obsolete may require an examination of all learners' evaluations of the content, trying to extract temporal patterns in the evaluations that show how recent learners like or dislike the content. The key point is that it is the purpose that determines what information to use and how it is to be used. An ideal goal for a real time e-learning system is that this determination be made *actively* (in the sense of [15]) at the time the purpose is invoked, so that no *a priori* interpretation needs to be given to the information; however, time constraints on executing the data mining algorithms may mitigate against such real time computation in many circumstances.

In sum, then the ecological approach promotes the notion that information gradually accumulates about learning objects, the information is about the use of the learning object by real learners, and this information is interpreted only in the context of end use. The approach is ecological because over time the system is populated with more and more information, and algorithms emulating natural selection based on purposes can determine what information is useful and what is not.

There are many possible applications for the ecological approach in e-learning. The approach could underlie the design of

- a study aid, for example to retrieve for a learner relevant papers from a cache of such papers for a graduate student trying to learn about an area of research (e.g. [19]);
- a recommender system, to recommend some content to a learner that is relevant to his or her current task (e.g. [16]);
- an instructional planner, to plan out a sequence of content pages of relevance to a learner, sort of an individualized curriculum of study;
- a group formation tool, to suggest to the learner a group of other learners relevant to solving a particular task or learning about a particular subject (e.g. [22]);
- a help seeker, to find another learner who can help the learner solve a problem he or she has encountered (e.g. the I-Help system [12]);
- a reminder system, to keep a learner updated with new relevant information, say from the web, that is relevant to the learner's goals;
- an evaluation tool, to allow learners' interactions with educational content to be studied by instructional and cognitive scientists, in particular to look at the experiences of all learners or particular types of learners with some educational content;
- an end-use tagging system, to automatically derive educational content tags from pre-established ontologies based on the experiences of the actual users of the content, and that can be parameterized by end use variables such as type of learner, success/failure of the educational content for each type of learner, etc. A variant of this is possibility is the ability to refine, modify, or change pre-assigned metadata based on inferences from end use;
- an "intelligent" garbage collection system, to determine the on-going relevance of educational content and, if necessary, to suggest modifications or even that it be deleted as no longer being useful to learners (e.g. as discussed in [4]).

3. A Semantic Web Approach to Supporting the Ecological Approach

3.1 Introduction

As described more fully in [6], the Department of Computer Science at the University of Saskatchewan has developed a set of e-learning applications which includes both a discussion forum system (asynchronous and synchronous), as well as a learning object-based content management system. Each of these systems is connected to the Massive User Modelling System (MUMS) [7] – a piece of semantic web based middleware which allows these systems to create small packages of learner modelling information called events. These events are marked up using the Resource Description Framework (RDF), and correspond to one or more RDF schemas[1]. Events are then forwarded from e-learning applications to higher level applications, in particular software agents, where they can be analysed and acted upon.

Our initial work in applying the ecological approach to learning object metadata is based primarily around our content management system, the iHelp LCMS. This system delivers standard IMS Content Package [18] formatted learning objects to learners, and typically includes text, video, and interactive exercises. In addition to reading the content, learners must complete a short quiz both before and after the learning object is delivered. This quiz contains multiple choice questions which are related to the content that is being taught. Each question/answer pair in a quiz is mapped to a particular domain concept expressed in our domain ontology, as well as an entry in an educational objectives ontology. Our domain ontology is a large (1,000+ node) RDF graph that represents the relationships between concepts covering basic computer science for non-majors, focusing on web technologies (HTML, ECMAScript, etc.) and the history of computer science. Our educational objectives ontology is based on the work done by Anderson et al. [2], which itself is built off of work done by Bloom et al. [1], and indicates the depth of cognition a student has demonstrated in a given topic.

Consider the following example taken from a lesson on operators:

Question: What is the result of the operation ((2<9) && (3>2))?
i) true
ii) false

If the learner answers true, it shows they can understand procedural knowledge in both of the topics "RealtionalOperators" and "LogicalOperators". If the learner answers false, they have not demonstrated any knowledge or ability in particular. Using the case of the former as an example, the results can then be expressed in RDF (shown graphically in figure one). It is worth noting that the content management system itself knows only about the user and the question/answers he or she has submitted (the left hand side of the figure) – semantic web rules can be used to value-add the RDF with a derived understanding of the competencies a student has gained after the fact.

[1] Examples of these schemas are available online at http://ai.usask.ca/mums/best_practices

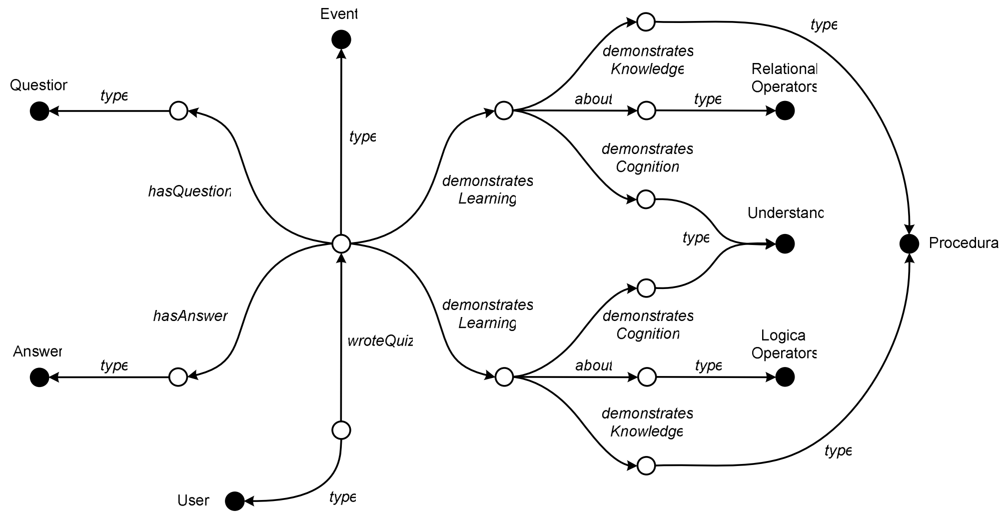

Figure 1: Graphical representation of RDF Model. Empty circles are instances while filled in circles are classes. Namespace prefixes and instance values have been omitted to aid in readability.

This form of collecting student competencies is relatively agile – there is no need to indicate "correct" and "incorrect" answers, instead, the instructional designer can indicate explicitly which answers demonstrate which competencies. This is especially useful in multiple choice tests where there is often one best answer, but several other options still demonstrate some smaller set of that knowledge. Further, the open world model of RDF allows for an arbitrary number and type of statements to be made about a students' interaction with a quiz. This is useful in making multiple assertions about student knowledge (as in above). We anticipate that associating known misconceptions (a form of bug libraries) with given answers will allow us to further value-add the learner model.

Once assessment data has been collected, it can be attached to the learning object. Instead of just associating the raw data with the object, the data can be summarized by subtracting all of those competencies demonstrated in the pre-test from those demonstrated in the post-test. This then shows the net gain in knowledge the student achieved by interacting with the learning object.

3.2 Issues with this Approach

A significant challenge within the educational technology community is in making e-learning artefacts interoperable. This challenge is the primary reason standards such as the IEEE LOM exist. The problem is that once an implementation deviates from a standard (as in our approach), interoperability begins to get severely hampered. To combat this, we anticipate that agents native to our metadata repository will be needed to reason over and summarize metadata to convert it to a more standardized form for export. Unfortunately, this is a lossy process, as in many instances summarized data cannot contain multiple records, and the new semantics we would like to introduce (through deeper user modelling) are meaningless to external repositories.

During the 2004-2005 regular session we began implementing this approach and collected assessment data from approximately 50 students using our online course, as well as feedback from the instructor. During this time it became apparent that students were frustrated with the pre-tests as they often lacked sufficient knowledge to understand the question being asked. We are addressing this by changing the pre-test from actual assessment to a declaration of self knowledge, where students would indicate what they felt their level of knowledge in each topic is using a likert scale. Values on this scale would then be mapped to approximate entries in the educational outcome taxonomy being used. In addition, there is a delicate balance between asking too many questions, risking that some of them may be off topic, and not asking enough, thus missing potential useful metadata entries. This is a trade off we are trying to mitigate by working closely with the instructor and instructional designers for the course.

Finally, it should be noted that while this technique allows for the collection of metadata about a given learning object, it does not in and of itself predict if that learning object is going to be useful for a given future learner. It is likely that we will investigate the use of probabilistic models, such as those presented in [23], when actually constructing an automated instructional planner.

4. Conclusions

Metadata specifications, in particular the nearly ubiquitous IEEE LOM standard, have yet to be proven effective at capturing enough metadata at an appropriate level to be used by automated instructional planners. In an attempt to bring automatic instructional planning to our online courseware, we have proposed a more lightweight metadata collection method, dubbed the *ecological approach*. In this approach arbitrary metadata statements are made about learners and attached to them in the form of a learner model. As a learner interacts within the learning environment, their model is associated with the learning artefacts that exist. In this way metadata is descriptive (based on actual observed interactions), as opposed to prescriptive (assigned by a content expert).

To being to concretize this approach we have started to collect competency lists for each learner both before and after they interact with a learning object, by way of pre and post quizzes. The difference between these lists results in the topics that the given learning object has taught to that particular learner. By associating these lists with the learning object, we are able to form a corpus of evidence that can be used by an instructional planner when sequencing objects together for other learners.

It should be noted that this is just one way in which we are applying the ecological approach. A larger research agenda built around our learner modelling middleware, the Massive User Modelling System (MUMS) [7] is also being pursued. In this we are capturing student interaction with both synchronous and asynchronous discussion forums, such as postings read, time dwelt on a posting, chats participated in, and general availability online. While an immediate end-goal for capturing this is to augment our peer help system (as described more fully in [21]), we also intend to associate this semantic data with learner models, which will then be attached to various artefacts in our systems. We anticipate that this, as well as the extra user interaction information that we are capturing from our content management system (e.g. which learning objects were read, how long they were read, what order they were read in, etc.) will prove useful when trying to adapt learning resources (from peer helpers to traditional learning objects) for personalized instruction.

References

[1] B.S. Bloom ed., *Taxonomy of Educational Objectives, The classification of educational goals, Handbook I, Cognitive Domain*, David McKay Company Inc., 1956.

[2] L. Anderson and D. Krathwohleds., *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*, Addison Wesley Longman Inc., 2001.
[3] *IMS Learning Resource Meta-data Best Practices and Implementation Guide, Version 1.1*, IMS Global Learning Consortium Inc., 2003
[4] Bannan-Ritland, B. , Dabbagh, N., and Murphy, K., "Learning Object Systems as Constructivist Learning Environments: Related Assumptions, Theories and Applications.," In The Instructional Use of Learning Objects (on line version). AIT/AECT, 2000,
[5] C. Brooks, *Versioning of Learning Objects*, master's thesis, Saskatoon, SK, Canada, University of Saskatchewan, 2005.
[6] C. Brooks, L. Kettel, and C. Hansen, "Building a Learning Object Content Management System" *World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education* (E-Learn 2005), Association for the Advancement of Computing in Education (AACE), 2005. In submission.
[7] C. Brooks et al., "The Massive User Modelling System" *7th International Conference on Intelligent Tutoring Systems* (ITS04), Springer-Verlag, 2004, pp. 635-645.
[8] K. Cardinaels, M. Meire, and E. Duval, "Automating Metadata Generation: the Simple Indexing Interface" *The 14th International World Wide Web Conference 2005* (WWW 2005), International World Wide Web Conference Committee (IW3C2), 2005.
[9] J. Cooke et al., "Computer Science Tutorial Page"; http://www.cs.usask.ca/resources/tutorials/csconcepts/index.html.
[10] *Shareable Content Object Reference Model (SCORM) Content Aggregation Model (CAM) Version 1.3.1*, Advanced Distributed Learning, 2004
[11] N. Friesen, Final Report on the "International LOM Survey", tech. report Document 36C087, Canadian Avisory Commiittee for ISO/IEC JTC1/SC36, 2004.
[12] J. Greer et al., "The Intelligent Helpdesk: Supporting Peer-Help in a University Course" *Fourth International Conference on Intelligent Tutoring Systems* (ITS 1998), Dpringer-Verlag, 1998, pp. 494-503 .
[13] *IMS Learning Design Information Model*, IMS Global Learning Consortium Inc., 2003
[14] *IEEE P1484.12.1-2002, Draft Standard for Learning Object Metadata*, IEEE, Inc., 2002
[15] G. McCalla et al., "Active Learner Modelling" *Intelligent Tutoring Systems 2000* (ITS2000), 2000.
[16] M. Recker and D. Wiley, "A Non-Authoritative Educational Metadata Ontology for Filtering and Recommending Learning Objects." *Interactive Learning Environments Journal: Special Issue on Metadata*, 2001, pp. 1-17.
[17] *Simple Query Interface (SQI) for Learning Repositories, Version 0.8*, 2004
[18] *IMS Content Packaging Information Model, Version 1.1.4*, IMS Global Learning Consortium Inc., 2004
[19] T. Tang and G. McCalla, "Smart Recommendation for an Evolving E-Learning System" *Workshop on Technologies for Electronic Documents for Supporting Learning, International Conference on Artificial Intelligence in Education* (AIED 2003), 2003.
[20] J. Vassileva, G. McCalla, and J. Greer, "Multi-Agent Multi-User Modeling in I-Help" *User Modeling and User-Adapted Interaction: Special Issue on User Modelling and Intelligent Agents*, vol. 13, no. 1, 2002, pp. 1-31.
[21] M. Winter, B. Daniel, and C. Brooks, "Towards Automatic Discovery of Peer Helpers from an Large Message Board System" *Workshop on Usage Analysis in Learning Systems, the 12th International Conference on Artificial Intelligence in Education* (AIED 2005), 2005. In submission.
[22] M. Winter and G. McCalla, "An Analysis of Group Performance in Terms of the Functional Knowledge and Teamwork Skills of Group Members" *Workshop on User and Group Models for Web-based Collaborative Environments, 9th International Conference on User Modeling* (UM 2003), pp. 35-45.
[23] J.-D. Zapata-Rivera and J. Greer, "Inspectable Bayesian student modelling servers in multi-agent tutoring systems" *International Journal of Human-Computer Studies*, vol. 61, no. 4, 2004, pp. 535-563.