Mining for Content Re-Use and Exchange -- Solutions and Problems*

Agathe Merceron[†]. Carlos Oliveira[‡],
Michel Scholl[†], Carsten Ullrich[§]
[†]Ecole Supérieure d'Ingénieurs Léonard de Vinci, [‡]Faculty of Engineering of the University of Porto, [§]Saarland University
agathe.merceron@devinci.fr, colive@fe.up.pt,
michel.scholl@devinci.fr, cullrich@ags.uni-sb.de

1 Problem Statement

Valuable information can be gained by mining metadata of educational resources. However, if the mined data is annotated using IEEE Learning Objects and Metadata standard (LOM), then significant pedagogical information is missing.

In this paper, we will describe how an ontology of pedagogical objects that captures the “pedagogical semantics” of a learning resource helps data mining to retrieve more precise information. Two case studies illustrate the advantages of an ontological approach. Finally, we will point out shortcomings of this approach and propose a solution based on fusion of ontologies.

2 An Ontology of Pedagogical Objects

In the emerging educational specification for learning content SCORM, course structure is built up out of self-contained units of learning. The description of these resources is covered in LOM, which provides a detailed description of learning resources. But LOM does not directly address the pedagogical purpose of a resource. The available Educational elements, in particular the "Learning Resource Type", do not include recommended values to, for instance, describe that a web page provides a definition or a counter-example of an item.

Ullrich [2004] describes an ontology of pedagogical objects that provides a vocabulary capturing the “pedagogical semantics” of a virtual or text-book learning resource. In general, each paragraph in a text book and each learning object in an online course serves a particular pedagogical role. These roles are reflected in the classes of the ontology.

The ontology distinguishes between two main classes, pedagogical concepts and satellite elements. Concepts describe the central pieces of knowledge, the main pieces of information being taught in a course. Subclasses of concepts are fact, definition, and different kinds of laws and processes. Satellites provide additional information about the concepts. Direct subclasses of satellites are interactive element, example, evidence, and explanation.

* This work was supported by IST TEL NoE Kaleidoscope.

3 Case Studies

We will now describe two case studies that show how data mining benefits from information contained in a pedagogical ontology.

3.1 Mining of Learning Paths

Teachers need to be aware of whether and how learning resources are used by students while learning. This can be achieved by mining students's logs and homework for associations. Association mining algorithms produce results of the form A,B->C, 80%, 95% which describe the relation that if students use learning resources A and B, then they use learning resource C with a support of 80% and a confidence of 95%. Here, a support of 80% signifies that 80% of the students use learning resources A, B and C. Confidence is a measure of how much C is implied by A and B. The higher the value, the greater the dependency.

If learning resources are annotated using a pedagogical ontology, and additional user related information is taken into account, much finer association rules giving more information to teachers can be obtained.

An example of an advanced rule is the following: students who study concept A, then example B, then example C complete successfully test D with a support of 80% and a confidence of 95%. This gives hints to reuse resources and compose a course for new students. Resources that make students successful should be recommended (as 'preferred learning tracks') to other students or should be preferred when building a new course.

Additionally, dominant learning styles in learner groups can be assessed. Let’s assume an association shows that students study a concept then solve 10 exercises in a row and complete successfully a test with a support of 25% and a confidence of 98%. This puts in evidence practical students who learn by doing. To cater for them, teachers have to reuse and share with their fellow teachers many exercises.

This approach generalizes the use of associations presented in Merceron and Yacef [2003] where students' homework is mined to find mistakes made while solving exercises in propositional logic.

3.2 Mining for Quality Control

Most Learning Management Systems (LMS) give access to several statistics related to the levels of activity of course participants and to the course content usage. Typical reports include the number of published documents, the number of viewers, the last access, etc.. Systems give also access to general statistics such as browser and operating system.

This is useful for teachers to understand what is happening during the course, but provides little insight into the factors that influenced the pedagogical results. For instance, most LMS do not provide the specific kinds of aggregated information about courses and users which are of interest for learning managers.

Oliveira and Domingues [2004] describe how information discovery of behaviours and trends can be improved by using multidimensional data analysis and mining technologies, while providing the basis for adaptive learning and automatic course creation.

Several types of metadata were considered in this context: *student interaction metadata* provides a measure of the progress and rate of learning and of the collaboration extent; *content and courseware metadata* enables quality control of the provided content and path optimization based in the pedagogical purpose of content; *evaluation and assessment metadata* helps to determine satisfaction and effectiveness of learning and registers grades and other scores.

Multidimensional views on these administrative metadata related with the course areas, the content types, the communication tools, the user profiles or the time axis, allow a deeper knowledge on the system usage and, consequently, better informed decisions on investment, user training, content reuse or pedagogic methodology developments.

Examples of usage analysis include questions like "what are the preferred document types?", or scenarios like "why do we have high rates in message answering delay?" or "what if we do not allow internal linking to teachers' homepages?".

While LOM and SCORM provide a framework for the representation and processing of the metadata, they fall short in including the needed semantic density for more specific pedagogical tracking. We thus argue that the use of a pedagogical ontology provides a higher level of decision support analysis and mining, based in qualitative issues like: the pedagogic methodologies used, the collaborative degree of activities or the understanding expressed in the assessments.

For example, we could make an analysis of the completeness of existent learning objects, in terms of concept description, example coverage and interactivity level, and explore detected trends. The correlation between learner interaction with the content and exercise scoring can yield important clues on the relevance of pedagogic methodologies used and raise management alerts where untypical patterns are detected.

4 Limitations

It is often the case when sharing pedagogical resources between separate sites that sites are not only physically distributed but more important have different usages. Even in the case they run the same scenario, significant heterogeneity in the pedagogical resources and description is expected. In our context it might happen that some other e-learning community would like to use a similar but different pedagogical ontology. One solution is the fusion of the two ontologies or an upgrading of the initial one to take into account the requirements and enrichments proposed by the other. If the requirements for updates are infrequent and hit a few terms only the fusion or updating of an ontology is realistic. In contrast if the two communities have divergent –although similar-goals and evolutions, a looser connection between ontologies is to be preferred.

A simple known model consists in connecting two similar terms from different ontologies by an isa link. Then, if term t in peer 1 is connected to term t' in peer 2, all resources under t become resources under t'. Given such a model, an interesting issue is to discover resources in neighbour peers and to find the best strategies for this distributed query processing, as described in Tzitzikas and Meghini [2003].

References

[Ullrich, 2004] Carsten Ullrich. Description of an Instructional Ontology and its Application in Web Services for Education. Submitted to the *Third International Semantic Web Conference*, 2004.

[Merceron,Yacef, 2003] Agathe Merceron and Kalina Yacef. A Web-Based Tutoring Tool with Mining Facilities to Improve Learning and Teaching. In *Proc. of the 11th International Conference on Artificial Intelligence in Education*, Sydney, Australia, 2003, U. Hoppe and F. Verdejo and J. Kay Eds., IOS Press, pages 201-208.

[Oliveira, Domingues, 2004] Carlos Oliveira and Manuel Domingues. Multidimensional Analysis of Administrative Data in eLearning Systems. In *Proc. of the 10th European University Information Systems Conference*, Bled, Slovenia, 2004, V. Mahnic and B. Vilfan Eds., University of Ljubljana Press, pages 172-178.

[Tzitzikas, Meghini, 2003] Yannis Tzitzikas and Carlo Meghini. Query Evaluation in Peer-to-Peer Networks of Taxonomy-based Sources. In *Proc. of the 10th International Conference on Cooperative Information Systems*, Catania, Italy, 2003.