Taking advantage of LOM Semantics for Supporting Lesson Authoring

Olivier Motelet and Nelson A. Baloian

Universidad de Chile, Santiago, Chile
{omotelet,nbaloian}@dcc.uchile.cl

Abstract. Learning Object Metadata (LOM) is an interoperable standard focused on enabling the reuse of learning material for authoring lessons. Nevertheless, few work was done on taking advantage of LOM-semantics to facilitate retrieval of learning material. This article suggests an original approach which uses the structure of a lesson in order to automatically generate LOM-semantic-based queries whereas the user continues to formulate easy-to-write queries without semantic specifications. This proposal consists of a four-component framework attempting to consider the main issues of semantic-based retrieval of documents.

1 Introduction

One of the main motivations behind Learning Objects and Learning Object Repositories is to facilitate their reuse by as many people as possible. In order to make this possible, the characteristics of the learning objects should be exposed, in order to let other people locate and retrieve them. A very critic issue in this process is how to describe an object and how to search for it in order to find those who really would match the needs of a potentially user. The metadata describing a learning object is a fundamental characteristic enabling this process. In order to make the finding of a suitable learning object more accurate, the description of a learning object should not only consider the physical characteristics of the document, like the one proposed by the DublinCore Metadata Intitiative[1] but it should also be pedagogically relevant. The Learning Object Metadata (LOM) standard includes such data. Consequently, Learning Object Repositories (LOR) typically use this metadata for the storage and retrieval of learning objects. However, following this standard means that authors or people classifying the learning object should give some value to almost 60 metadata attributes in order to fully describe the object according to the IEEE LTSC LOM specification[2]. Also users trying to retrieve the learning material may have to deal with this problem. Such a fastidious task is not compatible with making learning material sharing a customary activity for regular teachers. Several researchers have already described this problem and propose the

[1] http://www.dublincore.org/
[2] http://ltsc.ieee.org/wg12/

automatic generation of metadata as a way to solve it[1–3]. Basically, metadata generation systems are intended to improve the performance of metadata exploitation systems[4]. Similarly, metadata exploitation system should influence metadata generation system specifications. However, the topic of exploiting the metadata of learning objects is still in its beginnings. The typical way of making use of the metadata for retrieving relevant learning material is making a query ala Google on all the attributes independently of their nature. More advanced exploitation systems called recommender systems make use of the experience and opinion of other people having already used this material (see [5, 6] for a sample). Baloian et al.[7] use LOM and user/system modeling as a base of a collaborative recommender system for learning material. Duval and Hodgins[1] suggest a collaborative filtering system based on rating and pattern recognition. Both type of systems benefits from the semantics of LOM, i.e. the semantic structure of the data, to rate the didactic material and facilitate its retrieval. This article presents an approach that benefits from LOM semantics for retrieving learning objects to fit in a certain learning context. This procedure is aimed to support an instructor during the authoring of an entire course syllabus based on learning material retrieved from different repositories without having to provide explicitly all the metadata values for querying the repository. Moreover, this approach may also help to automatically generate metadata for a learning object which exists inside a coherent course syllabus. This method is complementary to a recommender system since it is based on the existence of a graph that structures and relates the learning material to support the process. In order to introduce our work, learning object retrieval based on LOM semantics is discussed. Then, learning object graphs are brought in and their dependency with LOM semantic is studied. Next, integration and processing of LOM-semantic-based retrieval is presented. Finally, a framework of a system implementing our approach is drawn.

2 LOM-Semantic-based Retrieval of Learning Object

Google and other indexing engines typically provide interfaces for simple queries with a semantic based on logical operators. In such systems, these basic queries (*BQueries*) target complex indexes generated by document content analysis procedures. Learning Object Repositories generally offer interfaces for processing such simple queries. In these settings, BQueries concern all the elements of the objects' metadata set independently of their nature, making a string matching without using any semantic similarity of the terms. Although such a retrieval process is simple for the end-user, it does not benefit from one of the main advantages of metadata over indexes: their semantic classification. LOM exploitation systems should use this characteristic to overpass the limits of string-based indexing engines. In the currently existing LORs, users have to complete forms with all the fields of the learning objects metadata set in order to make a query considering which takes in account the metadata semantics . Indeed, query languages enabling semantic precision (for example XPath, XQuery or RDFQL) are too complex to be integrated at user-level. Form-based queries for retrieving learning objects is a time consuming and tedious task. Studies[8] show that authors of learning material do not properly generate complete and correct metadata. In the same way, we do not expect that users are willing to properly generate metadata for searching this kind of material. Processing semantic based queries involves many well-known problems characterized by the Artificial Intelligence[9]. In particular, a system for processing LOM-Semantic-based Queries (*LSQueries*) should be able to find relations between the vocabulary used in the query expression and the vocabulary used in the learning object repositories. If no relevant matching can be found, LSQueries should be approximated in order to effectively retrieve the desired material. Approximated outcomes could be reached for example by a process in which query restrictions are relaxed according to predefined or customizable strategies.

Some work dealing with the automatic production of queries enabling semantic-based retrieval of learning objects has been motivated by the difficulty of doing it 'by hand'. Typically, this kind of systems falls in the category of recommender systems based on complex recognition pattern methods or user profile analysis[7, 1]. Other systems assist users to generate LSQueries. Pinkwart et al.[10] present a system generating LSQueries for retrieving learning material with similar characteristics to the one being used at that moment by teacher and/or students. This method is particularly aimed to support learning in a collaborative context. Learning Management Systems (LMS) could also help users to generate LSQueries by providing information like the educational context, the expected learning time and the used language[1]. Our approach take advantage of the structure in which a learning object might be embedded in order to enable semantic-based retrieval of learning objects. For example, a syllabus of a certain learning unit might be represented with a graph, in which nodes contain the learning material and edges the relations between them.

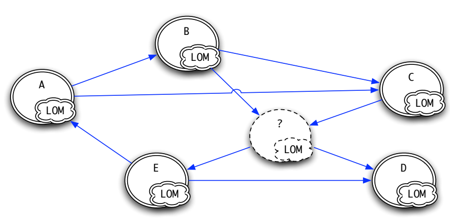

Fig. 1. Graph of learning objects during authoring process. One element is not yet referring a concrete material, but it is already characterized inside the graph.

As illustrated in figure 1, a lesson syllabus graph could refer to both, already existing material and material that has still to be provided or retrieved. Our proposal is to take advantage of the semantics of the graph, i.e. the nature of the edges relating the nodes, in order to identify some characteristics of the missing material. This process should enable the retrieval of educational resources suiting to the context defined by the lesson syllabus. In order to introduce this approach, the next part delves into the influence of the semantics of a learning object graph on the semantics of LOM.

3 Influence of Graph Semantics on LOM Semantics

From the beginning of the development of intelligent multimedia for learning application, authors have proposed the structuring of learning material in graphs. In [11], McCalla presents a number of self-adapting tutoring systems for supporting individual learners and he considers the graph as a key structure for the learning unit syllabus in order to achieve flexibility. Fischer[12] uses two different graphs to define a syllabus. First, a graph of concepts is built by means of a set of semantic relations. Second, a graph of material is defined based on a set of rhetorical relations. Using learning material metadata, the system generates semi-automatically the sequencing of the learning material. Baloian et al.[13, 14] use a graph structure for representing the syllabus of a learning unit. Such graphs are called Didactic Networks. Didactic Networks enables to generate several versions of a lecture according to different teaching styles. This functionality is based on the analysis of a predefined set of rhetoric relations between the didactic activities.

Independently of the type of graph used for structuring the learning material, an obvious fact is that the relations between two educational resources depend on their type and content. By definition, LOM reflects these characteristics. Consequently, in a learning object graph, the relations between two learning objects depend on the values of their metadata. Reversely, we can say that the relationship between the values of the metadata set of two linked objects of the graph is somehow aligned with the relationship represented by the link.

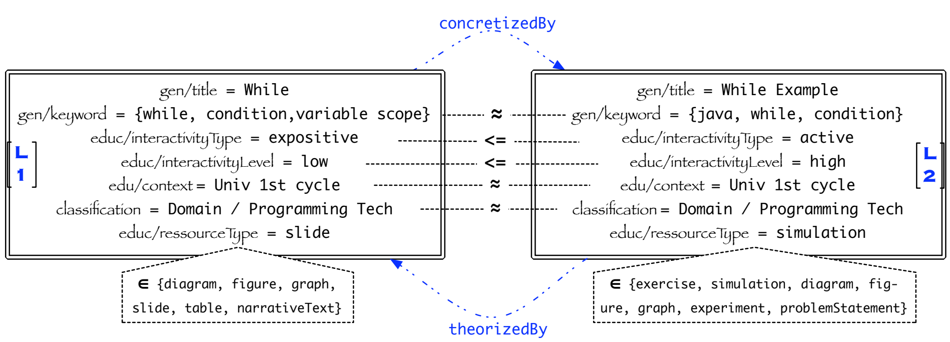

Fig. 2. Two learning objects $L1$ and $L2$ linked with rhetoric relations. These relations imply mutual influences between the LOM document values.

Consider the two learning objects $L1$ and $L2$ of figure 2. $L1$ theorizes $L2$ and $L2$ concretizes $L1$. Some *similarities* between the values of their metadata can be observed. For example, values for keywords, educational context, and classification are quite alike in both educational resources. This fact is not just a coincidence: we could derive the values of some metadata of one learning object from others by considering the relations between them. For instance, the nature of the relations between $L1$ and $L2$ imposes some *restrictions* on the material nature. Since $L2$ concretizes $L1$, $L2$ will deal with an exercise, a simulation, a diagram, an experiment, or a problem statement, whereas $L1$ will deal with a diagram, a figure, a graph, some slides, a table, or a narrative text. For the same reason, type and level of $L1$'s interactivity with the students will be certainly lower or equal than the ones of $L2$. Perhaps these assumptions may not have been valid for all potential users, so each community should define their own rules according to their needs. The important fact is that such rules provide relevant information for retrieving the learning material which is missing in a lesson graph. In particular, some rules may generate *restrictions* on the values of certain metadata. These restrictions could be used to formulate the queries to be sent to learning object repositories. In addition to that, other rules identify *similarities* between certain metadata. This similarities may serve to rank the query results. In the next part, this proposal is developed in a framework for semantic-based retrieval of learning objects during lesson authoring.

4 Semantic-based Retrieval during Lesson Authoring

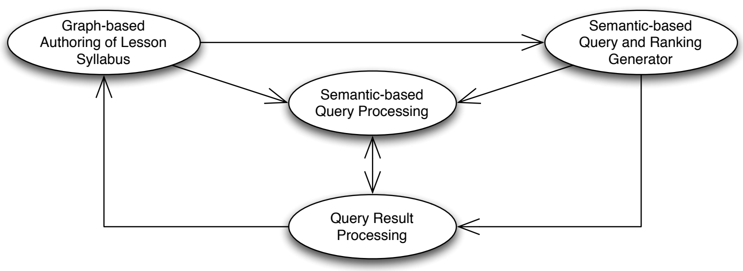

Fig. 3. Framework for semantic-based retrieval of learning objects during authoring of lesson syllabus.

Systems enabling retrieval of documents are basically composed of two central components, which deal with query processing and result processing. These components feed mutually each other: query processing may provide some results, and result processing may define new queries. In addition to them, a framework for semantic-based retrieval of learning objects during authoring of lesson syllabus should consider a third component reflecting the authoring process. As argued in the previous sections, this framework may also integrate a fourth component instantiating the generation of semantic-based queries and ranking information. This framework pictured by figure 3 is detailled in this section.

4.1 Graph-based Authoring of Lesson Syllabus

The lesson graph component is responsible for supporting the authoring of the lesson syllabus. In the present implementation, this component is based on a previous work: LessonMapper[14, 15], a Java application for authoring lesson graphs. This platform is extended to assume the process illustrated by figure 4. In this example, the author of the syllabus of a lesson about programming languages is looking for some learning material in order to introduce the concept of loops. First, he creates a new node $L4$ characterized with the key-sentence `loops`. Then, he specifies that $L4$ introduces $L1$, $L2$, and $L3$ by creating some links of type `introducesTo`. On one hand, the key-sentence is used to formulate a BQuery (see section 2), which is provided to the query processing module. On the other hand, the graph semantics are processed by the generator component in order to set down a LSQuery reflecting the lesson context.

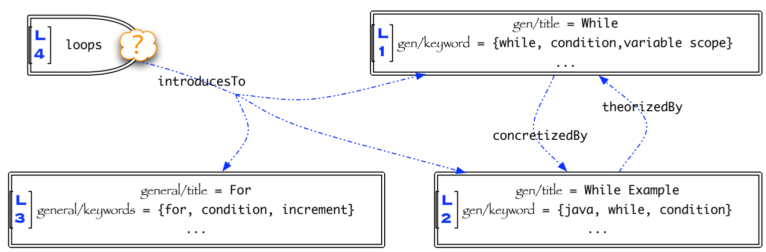

Fig. 4. L1, L2, and L3 are learning objects with LOM description. The instructor is looking for a learning object L4 in order to introduce L1, L2, and L3.

4.2 Semantic-based Query and Ranking Data Generator

The generator component is intended to provide semantic-based queries to the query processing component and also ranking information to the result processing module. Various implementations of this component may be developed. For instance, this module may process pattern analysis or user profile matching, or simply recover some relevant information from LMS. However, this work focuses on another alternative taking advantage of the influences of the graph semantics on LOM semantics. In this approach, which was introduced in section 3, query and ranking data generation results from processing the semantics of a lesson graph with a set of rules. This part first presents a language for specifying these generation rules. Then, rule diffusion in a recursive process is described.

Generation rule specification language. As argued before, generation rules should be well suited to the teaching/learning habits of the potential users. Therefore, rule set should not be hard-coded, but defined thanks to a Domain

Specific Language (DSL). We designed such a language embedded in XML elements for portability purposes. Our implementation includes some mathematical operators: *max*, *min*, *union*, *intersection*, *sum*, *product*, *subtraction*, and *division*. In the example of figure 4, since $L4$ introduces to $L1$, $L2$, and $L3$, the `keyword` metadata of $L4$ may have some similarities with the `keyword` metadata of $L1$, $L2$, $L3$. In order to postulate such a statement, the user should define the following rule, in which v stands for the value set of `keyword` metadata of all the educational resources related with `introducesTo`:

```
<similarity attribute="general/keyword" relation="introducesTo">
           UNION(v)
</similarity>
```

In order to specify restrictions on metadata values, our language also provides a set of boolean operators: $=$, $!=$, $<$, $<=$, $>$, $>=$, *contains*, and *containedIn*. For instance, since $L4$ introduces to $L1$, $L2$, and $L3$, the semantic density of $L4$ should be inferior or equal to the minimum semantic density of $L1$, $L2$, and $L3$. Such a restriction is generalized with the following rule:

```
<restriction attribute="general/semanticDensity" relation="introducesTo">
            '<=' MIN(v)
</restriction>
```

In order to compute this kind of rules, comparable elements should provide an order value. Practically, RDF vocabulary may include such a value as shown in the following example:

```
<lom_edu:InteractivityLevel rdf:ID="MediumInteractivity" order="15"/>
<lom_edu:InteractivityLevel rdf:ID="HighInteractivity" order="20"/>
```

Nevertheless, since most learning object repositories are not able to process restrictions based on vocabulary comparison, such properties is then expressed in terms of value set. For instance, a restriction of type '$\leq$ `mediumDifficulty`' is transformed in '$\in \{$`veryEasy`, `easy`, `mediumDifficulty`$\}$'.

Generation rule diffusion. The difficulty to properly generate metadata for educational resources imposes to consider potential incompleteness of LOM values in the lesson graph. In order to assume this situation, we suggest to benefit from the graph structure and propagate restrictions and similarities through the whole graph. In our implementation, this propagation process is based on the recursive framework introduced in a previous work[16]. In this framework, restrictions and similarities are not only based on the metadata of other educational resources, but also on the set of restrictions and similarities generated for these resources. Basically, this model introduces propagation and composition principles for restrictions and similarities. This feature enables the recursive processing of the rules and it limits the side-effect of metadata incompleteness.

4.3 Semantic-based Query Processing

This component is responsible for defining a query Q summarizing the products of the authoring component and the different implementation of the generator component. Afterward, Q is distributed to a set of learning object repositories.

Query Formulation. On one hand, the authoring component generates a BQuery, i.e. a query not considering any semantic restriction. On the other hand, each implementation of the generator component provide semantic restrictions formulated as LSQueries. The summarizing query Q is first a conjunction: $Q = \texttt{BQuery} \bigwedge \texttt{LSQuery1} \bigwedge ... \bigwedge \texttt{LSQuery}n$. Later, Q may be relaxed (for example in a disjunction) by the query result processing component. In our prototype, Q is written in Xquery. However, the choice of a query language should first depend on the compatibility with the targeted repositories.

Query Distribution. Query distribution deals basically with the communication with the learning object repositories. In our present implementation, distribution is limited to a local repository. Nevertheless, distribution should definitely be considered in order to reach sufficient sources for making a retrieval system interesting. An interface like Simple Query Interface (SQI)[17] may support this process. If we expect teaching/learning communities to use specific local vocabulary for sharing the educational resources[3], the terminology used in the query may be different from the one used in the learning object repository. However, such usage involves complex vocabulary distribution and interoperability issues[9].

4.4 Query Result Processing

The query result processing component deals with the answers of the consulted repositories. First, it is responsible for the presentation of the results. Then, according to the subjective analysis of the user, the first query may be relaxed and/or some didactic material may be reused.

Result Visualization. The generator component is susceptible to generate data enabling to rank the results returned by the learning object repositories. In our implementation, ranking information stands in the similarity set produced by processing the generation rules. Presently, educational resource matching some similarities has a better rank than material matching less similarities. Such a service may also be implemented with collaborative filtering techniques. Moreover, information visualization techniques may efficiently support the user in browsing the query results[18].

Query Relaxation. In case that a lesson author is not satisfied with query results provided by the system, he/she could reconsider some part of the generated query by relaxing the restrictions imposed on some attributes. For example, the restrictions done on the `general`, `lifeCycle`, `technical`, and `classification` categories of LOM may be relaxed to enlarge the search to educational resources matching with a certain pedagogical context but not limited to a specific discipline or format. The learning objects resulting from this relaxation process may offer interesting hints for defining methods supporting the particular educational objective of the authored lesson. Further work should be done on LOM semantics in order to offer a set of pedagogically-sounded relaxation strategies.

LO Reuse. Once the syllabus designer has selected one or more learning objects to be reused, a recontextualisation phase is required. This process deals with the adaptation of the retrieved material to a specific use context. Format, language, style and copyrights issues have to be managed, but these topics remain far out of the scope of this article.

5 Conclusion

This article presents an original approach for enabling LOM-semantic based retrieval of learning objects during the lesson authoring process. Our proposal differs from existing semantic-based retrieval systems because it is originally based on the analysis of the semantics of the lesson graph. For that reason, it could be used in complement to other approaches based on user profile, pattern analysis, or material similarity. For the same reason, our system focuses on a specific context: lesson-syllabus authoring based on graph.

Lesson graphs are not specific to this work but explored by several researches in the community. Their main advantages are flexibility during course presentation and semantic-based sequencing of the lesson. We attempt to aggregate another advantage to lesson graphs: the semantic-based retrieval of learning objects. Our approach enables automatic generation of LOM-semantic-based queries, whereas the user continues to formulate easy-to-write queries without semantic restrictions. Such system is based on generation rules exploiting the influences of graph semantics on LOM semantics. The same approach is also used to rank the query results according to the context of the authored lesson. The model can be adapted to specific didactic behaviors since rules are defined with a simple domain-specific language. Moreover, a recursive diffusion framework limits the impact of a potential incompleteness of the learning material metadata.

The whole system is part of a four-component framework susceptible to integrate other methods for generating queries and ranking data. This framework integrates the authoring process as a legitimate component of learning object retrieval systems. Query and result processing are considered with the perspective of semantic-based retrieval of educational resources. Complex issues like vocabulary distribution and learning object re-contextualization remain opened.

Nevertheless, interesting perspectives are also emerging like the possibility to define pedagogically-sounded retrieval strategies.

References

1. Duval, E., Hodgins: Metadata matters. In: International Conference on metadata and Dublin Core specifications. (2004)
2. Simon, B., Dolog, P., Miklós, Z., Olmedilla, D., Sintek, M.: Conceptualising smart spaces for learning. Journal of Interactive Media in Education **Special Issue on the Educational Semantic Web.** (2004)
3. Downes, S.: Ressource profiles. Journal of Interactive Media in Education **Special Issue on the Educational Semantic Web.** (2004)
4. Brasher, A., Andrew, P.M.: Human-generated learning object metadata. R. Meersman et al. (Eds.) OTM Workshops 2004 **LNCS** (2004) 723–730
5. Rafaeli, S., Dan-Gur, Y., Barak, M.: Social recommender systems: Recommendations in support of e-learning. Journal of Distance Education Technologies **3** (2005) 29–45
6. Kristofic, A.: Recommender system for adaptive hypermedia applications. In Bielikova, M., ed.: IIT. SRC. (2005) 229–234
7. Baloian, N.A., Galdames, P., Collazos, C.A., Guerrero, L.A.: A model for a collaborative recommender system for multimedia learning material. CRIWG 2004 **LNCS** (2004) 281–288
8. Friesen, N.: International lom survey: Report. Technical report, ISO/IEC JTC1/SC36 sub-committee (2004)
9. Stuckenschmidt, H.: Query processing on the semantic web. Kunstliche Intelligenz: Special Issue on the Semantic Web **3/03** (2003) 22–26
10. Pinkwart, N., Jansen, M., Oelinger, M., Korchounova, L., Hoppe, U.: Partial generation of contextualized metadata in a collaborative modeling environment. In: 2nd International Workshop on Applications of Semantic Web Technologies for E-Learning. (2004)
11. McCalla, G.: The search for adaptability, flexibility, and individualization: Approaches to curriculum in intelligent tutoring systems. In: Foundations and Frontiers of Adaptive Learning Environments. Springer (1992) 91–122
12. Fischer, S.: Course and exercise sequencing using metadata in adaptive hypermedia learning systems. J. Educ. Resour. Comput. **1** (2001) 5
13. Baloian, N.A., Hoppe, H.U., Pino, J.A.: A teaching/learning approach to cscl. In: HICSS. (2000)
14. Baloian, N.A., Pino, J.A., Motelet, O.: Collaborative authoring, use, and reuse of learning material in a computer-integrated classroom. In: CRIWG. (2003) 199–207
15. Motelet, O., Baloian, N.A.: Introducing learning management systems standards in classroom. In Kinshuk, Looi, C.K., Sutinen, E., Sampson, D.G., Aedo, I., Uden, L., Kähkönen, E., eds.: ICALT, IEEE Computer Society (2004)
16. Motelet, O.: Relation-based heuristic diffusion framework for lom generation. In: AIED - Young Researcher Track. (2005)
17. Simon, B., Massart, D., Duval, E.: Simple query interface specification. In: CEN/ISSS Workshop on Learning Technologies. (2004)
18. Klerkx, J., Duval, E., Meire, M.: Using information visualization for accessing learning object repositories. In: - 8th International Conference on Information Visualization. (2004)