Relation-based heuristic diffusion framework for LOM generation[1]

Olivier Motelet [2]
DCC - Universidad de Chile

Abstract. Learning Object Metadata (LOM) intends to facilitate the retrieval and reuse of learning material. However, the fastidious task of authoring them limits their use. Motivated by this issue, we introduce an original method for LOM generation based on relations between LOM documents. These relations significantly influence the attribute values. We formulate this influence with heuristics of acquisition, suggestion and restriction. A diffusion framework for these heuristics is suggested. In the context of relation-based graphs of LOM documents, this framework models the recursive processing of the heuristics. The generated values could then be used to assist users in generating LOM documents.

Keywords. learning material reuse, learning object metadata, metadata generation

1. Introduction

Since a few years, reuse of learning material is becoming a leitmotiv for researches on computer-aided education. A first obvious motivation is the economic interest of reusing learning material instead of authoring it indefinitely. Other motivations can be found in the pedagogical area. For instance, learner centered education focuses on the individual needs of learners. In such a pedagogical context, both the teaching and the learning material should adapt to a large variety of situations. In order to cope with this task, various researches aims at providing intelligent adaptation of didactic material to the learner profile (see [Mur99,Mer01,Bru96] for a sample). Despite very interesting results, these approaches imply an important bootstrapping cost due to the building of content-specific rules. For a setting in which the teacher remains responsible for learning material adaptation, literature suggests material should be easily adaptable during the class. It supposes the material to be sufficiently varied in order to precisely suit learners' profile. If such a generic material remains difficult to build from scratch, it could reasonably be the result of an emergent collaborative effort of teachers. To achieve this goal, the storage and the retrieval of learning material is needed. Most efforts in this area deal with the concept of *learning objects* referring to shared digital educational material [Wil00]. Learning objects (LOs) are described and referenced by *Learning Object Metadata* (LOM). LOM documents and their associated LOs are stored and retrieved in *Learning Object Repositories* (LORs).

[1]This work was partially financed by the IT Cooperation Center Chile-Korea and E-Lane Project
[2]Correspondence to: Olivier Motelet, DCC - Universidad de Chile, Avenida Blanco Encalada 2120, Tercer Piso, Santiago, Chile , C.P. 837-0459, Tel (+56 2) 678.4365, E-mail: omotelet@dcc.uchile.cl

Sharing learning objects is presently related with the fastidious task of instantiating the almost 60 metadata attributes of the IEEE LTSC LOM specification[1]. However, if we aspire to see regular teachers sharing and reusing learning material, the use of learning material distribution system should remain as light as possible. For this reason, the research community seriously focuses on the metadata generation issue [DH04,SDM+04,Dow04]. Following this direction, our work introduces an original framework to facilitate LOM instantiation.

First, this article presents researches about metadata generation. Then, it states that the relations between LOM documents influence the attribute values of these documents. Some heuristics are defined to formulate this influence. Next, a framework for diffusing recursively the effects of these heuristics is presented. Finally, the model is discussed.

2. Metadata Generation

Today, most tools for authoring LOM documents are form-based. Even if a form facilitates the use of the hierarchical syntax of XML, it does not support the instantiation of the LOM attributes. Most researches agree on the fact that metadata should be automatically generated [SDM+04].

Almost half of the LOM specification refers to the work of the Dublin Core Metadata initiative (DCMI[2]). Recent tools dedicated to the automatic generation of DCMI attributes give considerable results [Irv04]. They typically extract the information from the document content. However, most attributes of LOM are not processed by such generation tools and in particular those concerning educational topics. Indeed, the educational information generally remains implicit in the learning material. Recently, natural language processing was used to efficiently generate the educational attributes of LOM [YFL04]. However, this work is dedicated to the processing of particular learning objects describing lesson plans.

Duval et al. [DH04] suggest to extract educational information from the course authoring tools. During a same authoring session it is expected that some characteristics are shared by all the learning objects (e.g. author, educational context, typical age range, or language). Therefore, the authoring tool could hold this information so that it should properly be diffused to the metadata of learning material. Pinkwart et al. [PJO+04] study another source of information for generating metadata; they relate various versions of a same learning material. Nevertheless, this approach focuses on a specific context of collaborative learning.

In [Gre04], Greenberg affirms that the best metadata generation option remains to integrate both human and automatic processes. According to this trend, metadata suggestion emerges as another important topic in the field of metadata generation. Crystal's studies [Cry03] confirm the benefits of context exposition during metadata authoring. In practice, Hatala et al. [HR03] extract suggestions for metadata values from inheritance, accumulation, content similarity, and semantic similarity between learning objects.

Other works focus on limiting the set of possible values (scope and vocabulary) (1) to ensure the quality and consistency of the metadata [Dow04], and (2) to facilitate machine processing of the results [QH04]. This consideration is directly related to the concerns

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

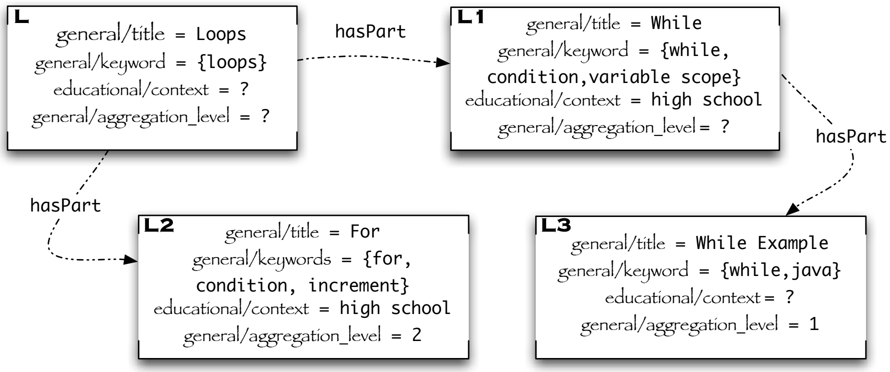

Figure 1. Sample of LOM documents in completion process.

of the semantic web. In this sense, the definition of an ontology for LOM [QH04] is a forward step.

3. Relation Influence on LOM documents

This section describes how the context drawn by the relations between learning objects can be used to generate, suggest and restrict the instantiation of LOM documents.

LOM not only describes learning object content but also its learning style and its intended learning context. This pedagogical information is particularly useful for teachers for retrieving appropriate material. It typically refers to the learners' profile, the learning environment and the global lesson surrounding the associated learning object. The relations with other learning objects are also part of this pedagogical context. For that reason, these relations are explicit in the LOM documents.

We state that a specific relation between two LOM documents could imply that the value of the attributes of one will influence the value of the attributes of the second. Let consider the four LOM documents pictured in figure 1. In this example, the LOM documents L and $L1$ are respectively linked to $L1$ and $L2$, and $L3$ with an hasPart relation. According to the definitions of the general/keyword attribute and the hasPart relation, it seems logical that L could also hold the keywords of $L1$ and $L2$. Similarly, $L1$ could inherit the keywords of $L3$. This logic is based on a subjective interpretation of the LOM specification and a generalization of this statement remains out of the scope of this work. Nevertheless, it is obvious that such an heuristic could facilitate LOM instantiation. In the same example, we also observe that the value of the educational context of document L have significant possibilities to be the same that the educational context of $L1$ and $L2$. As argued in the previous section, such a suggestion is of interest for the user instantiating LOM documents. Finally, the definitions of the general/aggregation_level attribute and the hasPart relation impose that the aggregation level of L cannot be inferior to the aggregation level of $L1$ or $L2$. Similarly, the aggregation level of $L1$ cannot be inferior to the aggregation level of $L3$. LOM document instantiation should also consider such restrictions. In this example, three types of heuristics have been illustrated. They are heuristics of acquisition, suggestion and restriction. Note that the specific values of our examples did not influence the definition of these heuristics. In fact, each heuristic depends only on a specific attribute and a specific relation type.

4. Heuristic Diffusion Framework

This section defines a framework describing how the acquisition, suggestion and restriction heuristics are processed. We call it diffusion framework for its recursive nature.

In the previous section, we introduced an acquisition heuristic for the LOM attribute general/keyword and the relation hasPart. This heuristic applied on the LOM document L of figure 1 returns the keywords of all the documents linked to L with an hasPart relation, concretely {while,condition,variable scope,java,for,increment}. Consider a function ϕ_{Acq} returning the general/keyword values of the LOM documents associated to the document L with the hasPart relation. Then,

$$\phi_{Acq}(L, \text{general/keyword, hasPart}) = \{ (L1, \{\text{while,condition,variable scope,java}\}), (L2, \{\text{for,condition,increment}\})\}$$

With such a function, the acquisition heuristic for general/keyword and hasPart can be formulated for any LOM document l:

$$AcqHeur(\text{general/keyword,hasPart})(l) = \bigcup v_i / (l_i, v_i) \in S, l_i \in \mathcal{L}, v_i \in \mathcal{V} \quad \text{where } S = \phi_{Acq}(l, \text{general/keyword, hasPart})$$

This definition is tight to the couple (general/keyword,hasPart). For instance, an acquisition heuristic for (educational/semantic_density,hasPart) would not implement a union but a mean balanced by the values of an additional attribute[3].

For all LOM documents, the original value of an attribute and the value due to heuristic processing are conceptually different. The first will be invoked with a function $AcqVal$. The second will be retrieved with a function $AcqDif$. The latter is a union of the original value of the attribute and the acquisition heuristic results for this attribute. Concretely, applying these functions on the document L and its general/keyword attribute gives:

$$AcqVal(L, \text{general/keyword}) = \{\text{loops}\}$$
$$AcqDif(L, \text{general/keyword}) = \{\text{loops,while,condition,variable scope,java,for,increment}\}$$

We formally define an acquisition heuristic diffusion framework as the set of the functions introduced in this section.

Definition 4.1 (Acquisition Heuristic Diffusion Framework) *Let $\mathbf{A}$ be the set of attributes of the LOM specification, $\mathbf{T}$ the set of relation type between LOM documents, $\mathcal{L}$ the set of LOM documents, $\mathcal{V}$ the set of generic values, and $\mathcal{R} = \mathcal{L} \times \mathbf{T} \times \mathcal{L}$ the set of existing relations between LOM documents, then*

- $AcqVal : \mathcal{L} \times \mathbf{A} \rightarrow \mathcal{V}$
- $AcqDif : \mathcal{L} \times \mathbf{A} \rightarrow \mathcal{V}$. *With* $l \in \mathcal{L}$, *and* $a \in \mathbf{A}$,
  $AcqDif(l, a) = AcqVal(l, a) \cup \bigcup_{t \in \mathbf{T}} AcqHeur(a, t)(l)$
- $AcqHeur : \mathbf{A} \times \mathbf{T} \rightarrow \mathcal{L} \rightarrow \mathcal{V}$
- $\phi_{Acq} : \mathcal{L} \rightarrow \mathbf{A} \times \mathbf{T} \rightarrow 2^{\mathcal{L} \times \mathcal{V}}$. *With* $l \in \mathcal{L}$, $a \in \mathbf{A}$, *and* $t \in \mathbf{T}$,
  $\phi_{Acq}(l, a, t) = \{(l', AcqDif(l', a)) / l' \in \mathcal{L} \bigwedge (l, t, l') \in \mathcal{R}\}$

[3] Such complex heuristics can be found at http://www.dcc.uchile.cl/∼omotelet/heuristics.pdf

The recursive aspect of the framework is introduced by ϕ_{Acq} calling recursively the diffusion function $AcqDif$ on the related LOM documents. If this version of ϕ_{Acq} suits our simple example, in a more realistic context it also needs to manage with recursion depth and cycle prevention, but this is part of future work.

A suggestion heuristic diffusion framework is similar to the acquisition framework except the definition of ϕ_{Sug}. In this case, both the suggestion diffusion and the acquisition diffusion should be recursively called. With $a \in \mathbf{A}$, $t \in \mathbf{T}$, and $l \in \mathcal{L}$,

$\phi_{Sug}(l,a,t) = \{(l', AcqDif(l',a) \cup SugDif(l',a))/l' \in \mathcal{L} \bigwedge (l,t,l') \in \mathcal{R}\}$

A restriction heuristic diffusion framework differs from the previous frameworks since a restriction consists of couples operator-value. The previous section introduced a restriction heuristic tight to the couple (general/aggregation_level,hasPart) (noted (Agg,hasPart)). In the LOM document set of figure 1, we can observe that $ResHeur(\text{Agg,hasPart})(L1) = \{(\text{'}\geq\text{'}, 1)\}$ and $ResHeur(\text{Agg,hasPart})(L) = \{(\text{'}\geq\text{'}, max(1,2))\}$. Such results can be obtained thanks to a function ϕ_{Res} processing recursively not only the restriction diffusion but also the acquisition restriction. Let $\mathbf{O}$ be the set of operators for characterizing constraints. Then,

$\phi_{Res} : \mathcal{L} \times \mathbf{A} \times \mathbf{T} \rightarrow 2^{\mathcal{L} \times 2^{\mathbf{O} \times \mathcal{V}}}$. With $l \in \mathcal{L}$, $a \in \mathbf{A}$, and $t \in \mathbf{T}$,

$\phi_{Res}(l,a,t) = \{\big(l', \{(\text{'='}, AcqDif(l',a))\} \cup ResDif(l',a)\big)/l' \in \mathcal{L} \wedge (l,t,l') \in \mathcal{R}\}$

With this function, the restriction heuristic of (Agg,hasPart) can be defined for any LOM document l:

$$\begin{aligned} ResHeur(\text{Agg,hasPart})(l) = & \big\{(\text{'}\geq\text{'}, max(v_{ij}))/o_{ij} = \text{'}\geq\text{'} \vee o_{ij} = \text{'='}\big\} \\ & \big/(l_i,(o_{ij},v_{ij})) \in S, l_i \in \mathcal{L}, o_{ij} \in \mathbf{O}, v_{ij} \in \mathcal{V} \\ & \text{where } S = \phi_{Res}(l, \text{Agg, hasPart}) \end{aligned}$$

5. Discussion

The theoretical advantage of LOM documents for reusing learning material is limited by the difficulty to generate them. Motivated by this issue, this work introduced an original method for metadata generation based on relations between LOM documents. These relations offer relevant information for LOM attribute instantiation. In practice, this information is generated by three types of heuristics: acquisition, suggestion, and restriction. A diffusion framework models the recursive processing of these heuristics on a relation-based graph of LOM documents. An existing work on rule-based metadata generation [HR03] focused on analyzing LOM packaging frameworks like SCORM and IMS-LD, and the similarities between LOM documents. These analysis result in sets of suggested values. Since our approach studies a different mechanism, i.e. a relation graph, it is complementary to this work. Moreover, our system not only suggests values for LOM instances but also restrictions for these values.

Further work is needed in order to consolidate and implement the system. In particular, recursion depth and cycle prevention should be controlled. Since our approach takes benefits of relation-based graphs of LOM documents, a consistent set of relation types should be defined in order to increase the generation potential of the system. Moreover, it is necessary to define sound heuristics which may be based on a LOM ontology. Finally, well-defined taxonomies for LOM attribute values would increase the potential of our system.

Another perspective for our heuristic diffusion framework stands in the retrieval of learning objects. In a course authoring system based on graphs of LOM documents (e.g. [BPM03]), a LOM node without reference to a concrete learning object could be considered as a query on a learning object repository. Since this node is part of a relation graph, it could receive acquisition and restriction values from our system. This feature could significantly precise the query with the pedagogical context of the lesson being authored. Moreover, the suggestion values available in our framework could be used to effectively refine the query result ranking.

References

[BPM03] Nelson A. Baloian, José A. Pino, and Olivier Motelet. Collaborative authoring, use, and reuse of learning material in a computer-integrated classroom. In *CRIWG*, pages 199–207, 2003.

[Bru96] Peter Brusilovsky. Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 6(2-3):87–129, 1996.

[Cry03] Abe Crystal. Interface design for metadata creation. In *CHI '03: Extended abstracts on Human factors in computing systems*, pages 1038–1039. ACM Press, 2003.

[DH04] E. Duval and Hodgins. Metadata matters. In *International Conference on metadata and Dublin Core specifications*, 2004.

[Dow04] Stephen Downes. Ressource profiles. *Journal of Interactive Media in Education*, Special Issue on the Educational Semantic Web.(ISSN: 1365-893X), 2004.

[Gre04] Jane Greenberg. Metadata extraction and harvesting: A comparison of two automatic metadata generation applications. *Journal of Internet Cataloging: The International Quarterly of Digital Organization, Classification, and Access*, 6(4):58–82, 2004.

[HR03] Marek Hatala and Griff Richards. Value-added metatagging: Ontology and rule based methods for smarter metadata. In *RuleML*, pages 65–80, 2003.

[Irv04] K. Irvin. Comparing information retrieval effectiveness of different metadata generation methods. Master's thesis, UNC SILS, 2004.

[Mer01] D. M. Merrill. *Knowledge objects and mental models*, chapter The Instructional Use of Learning Objects. Association for Instructional Technology, 2001.

[Mur99] Tomas Murray. Authoring intelligent tutoring systems: an analysis of state of the art. *International Journal of Artificial Intelligence in Education*, 10:98–129, 1999.

[PJO+04] Niels Pinkwart, Marc Jansen, Maria Oelinger, Lena Korchounova, and Ulrich Hoppe. Partial generation of contextualized metadata in a collaborative modeling environment. In *2nd International Workshop on Applications of Semantic Web Technologies for E-Learning*, 2004.

[QH04] Jian Qin and Naybell Hernández. Ontological representation of learning objects: building interoperable vocabulary and structures. In *WWW (Alternate Track Papers & Posters)*, pages 348–349, 2004.

[SDM+04] Bernd Simon, Peter Dolog, Zoltán Miklós, Daniel Olmedilla, and Michael Sintek. Conceptualising smart spaces for learning. *Journal of Interactive Media in Education*, Special Issue on the Educational Semantic Web.(ISSN: 1365-893X), 2004.

[Wil00] David Wiley. *learning object design and sequencing theory*. PhD thesis, Brigham Young University, June 2000.

[YFL04] Ozgur Yilmazel, Christina M. Finneran, and Elizabeth D. Liddy. Metaextract: an nlp system to automatically assign metadata. In *JCDL '04: Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, pages 241–242. ACM Press, 2004.