LOIT: An indexing tool based on LOM ontology

Ouafia Ghebghoub\textsuperscript{a}, Marie-Hélène Abel\textsuperscript{b}, Claude Moulin\textsuperscript{b}, Mohamed Benmmohamed\textsuperscript{c}

\textsuperscript{a} Computer Science department, University of Jijel Jijel 18000, Algerie
E-mail: ouafia.ghebghoub@gmail.com
\textsuperscript{b} CNRS Heudiasyc, Compiègne University of Technology 60205 Compiègne CEDEX France
E-mail: {marie-helene.abel,claude.moulin}@utc.fr
\textsuperscript{c} LIRE Laboratory, Mentouri University BP 325 Ain El Bey, Constantine 25000, Algeria
E-mail: ben_moh123@yahoo.fr

Abstract. The Web has allowed the publication of many pedagogical resources. However, due to their important number, it is difficult to manage and exploit them efficiently. One way to solve this issue is to describe the resources with informations taken out from metadata schemas. This solution is not sufficient because schemas are not really shared formalisms and may contain semantic ambiguity. For preventing such ambiguities, we chosed to build an ontology whose objective is to better define the elements contained in description schemas. In this paper, we present the ontology building process used for indexing pedagogical resources. We precisely define our concept of semantic indexing. We present also the tool we have developed for indexing and retrieving pedagogical resources in a local knowledge base.

Keywords: Ontology, LOM, Metadata, Learning object, OWL

1. Introduction

A new form of learning, called e-learning, has emerged from the use of information and communication technology. The e-learning is based on the diffusion of resources through network (Internet or Intranet) or any other electronic medium. It is defined as “just-in-time education integrated with high velocity value chains. It is the delivery of individualized, comprehensive, dynamic learning content in real time, aiding the development of communities of knowledge, linking learners and practitioners with experts” [3]. The learning content used is called learning object (LO). However, the increasing number of available LO makes it even more difficult for learners and teachers to find the learning contents that best meet their needs.

To enable the localisation, access, sharing and reuse of these resources, it is necessary to describe them. A solution consists to use a set of shared metadata. This metadata can be information regarding the authors of learning materials, their fields of interest, their ideas, their learning objectives, and so on.

Currently, the most successful and most used standard for describing the LO is the LOM\textsuperscript{1} (Learning Object Metadata). However, the semantic ambiguity of some of its elements and their subjectivities, especially in the educational category, makes difficult the use of its descriptors.

To overcome this difficulty and to allow the understanding of these elements, they should be assigned with semantics including the definition of concepts, relationships between concepts, attributes and constraints. The use of the powerful representation of on-

\textsuperscript{1}http://ltsc.ieee.org/wg12/20020612FinalLOMDraft.html
ologies applied to LOM modelisation can make easier the resources description by using the semantic links between concepts and hence between related learning objects. Our approach is also motivated by the observation that a LO can be connected with other LOs through relationships described in an ontology.

Building an ontology from the LOM will force an unambiguous definition of concepts and relationships that makes explicit items. Each resource will be described by a concept rather than words that may be ambiguous.

Notice that the construction of an ontology with an operational objective faces a problem of choice of construction methodology and adequate representation formalism. It is essential that it be powerful enough to represent all the elements we want to incorporate into this ontology.

In this paper, we will discuss some difficulties in relation with the use of the LOM standard for indexing learning resources. Then, we justify why the approach we have adopted is relevant to overcome the metadata limits and present the construction method of our ontology. Thereafter, we illustrate how to use our ontology by means of two examples. The first example concerns the tool we developed for indexing and retrieving the LOs. The second concerns MEMORAe2.0 [2] project.

2. Learning object

The basic idea behind the creation of learning objects is the possibility to build components or small units that can be reused several times in different learning contexts.

Adopting this concept of small reusable units, [13] explain that often when teachers or learning content creators access for the first time to a learning material, they break it down into its components, and then assemble these components in order to build a material that supports their educational goals.

In order to find this learning material to use or reuse, it must be described efficiently. A resource that is not indexed, is an unusable resource because it cannot be found.

To develop and promote standards for learning technology, the IEEE\(^2\) consortium created in 1996 the Learning Technology Standards Standing Committee (LTSC). The committee then chose the term learning object to describe these small reusable units.

2.1. Definition

The definition of a learning object gave rise to several debates. In the document describing the LOM 1.0\(^3\) standard, a learning object is defined as any digital or non-digital entity that can be used, reused or referred to during learning, education, or training activities. This definition is seen by [15] as wide as it may include an object, a person or an idea. [12] refines it by adding that a learning object is an “independent and self-standing unit of learning content that is predisposed to reuse in multiple instructional contexts”.

In our work, we have adhered to the [15] definition that considers that a learning object is a learning material that can be selected, combined with another according to the needs of teachers and learners. It is also a learning content that should exist as such, and can be retrieved and indexed easily.

To enable retrieval and indexing of learning objects, the use of descriptions is necessary. One solution is to define a set of metadata for describing LOs.

2.2. Metadata and standards

The term metadata refers to data about data or information about information. It is defined as structured information describing a resource. It includes characteristics or properties such as file, format, size, location, etc. These information can be embedded into a resource or they can be stored separately.

In order to facilitate sharing and exchange of learning objects and their metadata as well, it is necessary to adopt a common metadata set to describe them. Several metadata schemas have been developed in a variety of user environments and disciplines. They are designed principally to standardize the description of resources. Some of them have been recognized by the organization for standardization such as ISO (The International Standardization Organization), they have become standards.

In March 1995, the first basis for a documents descriptors system was proposed. The Dublin Core\(^4\) (DC) standard is a set of 15 simple elements such as title, creator, subject, language, etc. where each element is optional and repeatable. This schema was not

\(^2\)Institute for Electrical and Electronic Engineers, http://ltsc.ieee.org/wg12

\(^3\)http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

\(^4\)http://dublincore.org/documents/usageguide/elements.shtml
designed specifically for the e-learning. It was thought by some that it is not comprehensive and suitable for education domain. Therefore it has been adjusted to be applied to learning resources.

In August 1999, a DCMI (Initiative Metadata DC) Education Working Group\(^5\) has been established to reflect the manner in which the DC standard can be adapted to educational resources. The DC Education helped enriching the initial elements of DC.

Other metadata schemas have been designed to take into account the specificities of certain professions or fields of applications. In the field of education and e-learning, working groups have attempted to define and specify suitable description elements. One of the most complete and successful standards for describing learning resources is the Learning Object Metadata.

3. LOM standard

The standard developed by the IEEE LTSC, was approved by the IEEE-Standards Association on June 12, 2002. 1484.12.1 - 2002 Learning Object Metadata specifies the syntax of the elements composing it such as element names, definitions, vocabularies, occurrences number and data types.

The LOM data model is a hierarchy of elements (45 elements in LOM 1.0) grouped into nine categories such as general, life cycle, meta-data, educational, technical, right, relation, annotation and classification.

The LOM descriptors can be used to design the e-learning system for the indexing of learning objects. To do this, they must be implemented in a structured language.

3.1. LOM implementation

The LOM metadata instances are implemented using XML or RDF (Resource Description Framework)\(^6\) binding [10]. Data in XML binding is considered as a tree having no semantics. For example, it is possible through XML binding to distinguish between title or description of a learning object, to deduce that these elements belong to the “General category” but we cannot guess the meaning of the “Aggregation level” or the “structure” of a LO which are two elements in the “General category” LOM (see example in [5]).

XML binding is probably easily implemented but remains inadequate for the representation of LOM elements since it does not express the meaning of its elements. In RDF binding, data have semantics and can be viewed as objects having properties that relate them to other objects. In addition to information provided by XML binding in the previous example, RDF binding provides also a definition of the title, description, aggregation level and structure among other things.

Therefore a RDF representation of the LOM is richer than the corresponding XML representation since RDF adds interpretations to LOM elements that are not explicit. With RDF, it is possible to describe a resource by a sequence of assertions representing its characteristics. That is why RDF is called metalanguage. Developed by the W3C, it was intended to structure the information available on the web to make its indexing easier.

An RDF statement is a triple consisting of a subject, a predicate and an object, where the subject is referred to by an URI. RDF allows the definition of appropriate schema vocabularies (RDFS). RDFS adds to RDF the definition of classes, types of properties, the statement inheritance, etc. Existing RDF vocabularies were also reused in RDF binding for related specifications to LOM such as Dublin Core and vCard\(^7\) specification which is a standard format for personal data exchange. Remember that the RDF binding does not represent relationships between LOs but connects directly URIs denoting objects (see figure 1). Certainly, RDF binding adds semantics to LOM elements but remains inadequate since the lack of RDF and RDFS expressiveness does not allow to represent all the desirable characteristics.

For example, the elements “title” and “entry” which are mandatory in LOMFR\(^8\), which is a standard designed to extend the terms of the LOM. Using RDF we cannot specify if a property is mandatory or limited to a single use for any resource. The only possible constraints on the properties with RDF are those of domain and range. We can neither specify some deduction rules like : if a learning object is of an atomic structure then its aggregation level has a value of 1 and at least two in the other structure types. RDF and RDFS formalism does not allow to express the inverse of a relationship which can be interesting to define in

\(^5\)http://dublincore.org/groups/education

\(^6\)http://www.w3.org/RDF/

\(^7\)IEEE Learning Object Metadata RDF binding. http://kmr.nada.kth.se/static/ims/md-lomrdf.html then link VCARD RDF

\(^8\)http://www.lom-fr.fr/
some cases. Consequently this lack of expressiveness of RDF and RDFS led us to a more complete representation of LOM in a more powerful formalism.

3.2. Use of LOM

Based on the model and its bindings, several tools have been developed to create metadata, and various learning object repositories (LOR) have been created to store and index LO in common areas.

Some editors may be downloaded and used locally such as LomPad9, others are used online. With the exception of SHAME10 which is based on the LOM RDF binding, most of these editors generate files compatible with the XML binding. However, if these tools allow to describe the learning resources and to store these descriptions, they do not use them for the recovery of learning objects.

LOR have their own editors and have more complex systems that allow to describe the LO. These repositories store these LO along with their descriptions. In addition, they allow to retrieve these described LOs. Among these LORs, we find the ARIADNE11 project of the Foundation for the European knowledge pool and the American collection of links towards sharable learning objects, MERLOT12 (Multimedia Educational Resource for Learning and Online Teaching).

However implemented in XML or RDF, the use of LOM metadata raise a number of problems that we try to summarize in the following section.

3.3. Difficulties with using LOM

Although the standard is widely used, it is strongly criticized to be complex and to have a large number of elements. Also, many problems are detected during its use.

In addition to the large number of elements, the nature of the information described complicates further its use. Many details such as difficulty or age do not appear in the content of the LO, as opposed to the title or language which may be explicit, which needs human intervention anyway.

This information can also be objective or subjective [4]. Objective information are those whose values consist of statements of facts objectively determinable such as author or date of creation. Whereas subjective information is intended to represent a person’s point of view on certain aspects such as the difficulty of a learning resource.

To facilitate the indexing by using subjective metadata, LOM provides a list of predefined values. However, these values are not always well defined and selection criteria is arbitrary. It is the case, for example, for the interactivity level element in the educational category which takes its values as : very low, low, high and very high.

Nevertheless in ARIADNE, elements belonging to the educational category (context, interactivity level, difficulty) are rarely used, and even when they are used, they take average values similar to defaults values [9]. The most used elements are the basic metadata such as title and author name.

These problems also are due to an incorrect interpretation and misunderstanding of the elements as well as their values. Indeed, the semantic ambiguity of some LOM elements, and especially their subjectivity in the educational category, makes the description of resources difficult [8].

It seems clear that the difficulties encountered when using LOM metadata are not only caused by the complex structure of the schema or the large number of its elements but also by the result of users’ misinterpretation for certain elements.

3.4. Some solutions to difficulties with using LOM

In order to solve part of the problems associated with the use of LOM, several researchers have proposed to combine ontologies to certain elements of the LOM.

- Indeed, [7] determine the elements of the LOM for which they suggest the use of ontologies. These elements are the structure and learning strategies.
– [14] propose to associate the element “Taxon-path” in “Classification category” to the concepts of an ontology. In this way we can associate, for example, for the “prerequisite” value of the “purpose” element the term Boolean logic of ontology and for the “discipline” value of the same element the term programming. This allows to say that the discipline of this LO is programming and prerequisites are Boolean logic.

– [11] propose an approach based not only on the use of a single ontology but on several ones, in order to mitigate the LOM’s lack of expressiveness problem.

In most of these works, one or more LOM elements are substituted by concepts from ontologies. If this approach avoids some semantic ambiguity problems, it does not have a good understanding of all elements and it does not express all the relationships that may exist between these elements.

Our approach is to build a domain ontology based on the LOM’s description schema. Using a LOM’s ontology for indexing learning objects will allow a better understanding of the elements and their values and thus facilitate their descriptions. This ontology will enable a better definition of the LOM elements and of the learning object. It will also determine the relations which will enable to link the LOs to each other. That’s why we’re putting ahead the “structure” element in the “General category” which will allow to infer semantic links between LO. To describe the learning content, our approach propose to use the concepts of an application ontology for the “keyword” element in “LOM general category”.

4. Ontology building

To construct the ontology of LOM, we began by studying the structure of the LOM 1.0 and XML files generated by an editor as LOMPad. We built the ontology with multilangual approach. Concepts, attributes and individuals have labels and comments in French and English. For a French use of our ontology, we have included definitions of terms from LOMFR standard established by the French Association for Standardization AFNOR. However for the sake of flexibility, we relied on the original structure of the LOM developed by the IEEE consortium. In this way, a user can switch from English to French easily.

4.1. Building rules

Several methodologies have been proposed to develop ontologies [6]. For building our ontology we haven’t used a particular methodology. We started from the RDF binding and have defined different rules:

- R1: the learning object and LOM categories are represented by concepts that are LearningObject and LOMCategory;
- R2: each category of LOM is represented by a sub concept of LOMCategory;
- R3: the LearningObject concept is represented by a concept representing the LOM categories;

The LOM elements are in turn represented by a concept, an attribute or a relationship according to their type. Indeed, it lists three types of elements:

a. Simple elements (or atomic) with predefined values, such as the Structure element in the General category;

b. Simple elements without predefined values, such as the Description element in the General category;

c. Complex elements such as the Contribute element in the Life cycle category.

The representation rules Ri for the element e are:

- R4: if e is of type “a”, then it is represented by a concept associated with its category, and its values v are represented by individuals;
- R5: if e is of type “b”, then it is represented by an attribute associated with the concept representing its category or the concept representing the element to which it is attached;
- R6: if e is of type “c”, then e is represented by a concept associated with its category.

4.2. Building process

In the following we present examples how we applied the defined rules to build our ontology.

By applying R1, we first created the LearningObject concept and the LomCategory concept. With R2, the LomCategory concept is specialized in nine sub-concepts representing the nine LOM categories. R3 al-
allowed us to associate the LearningObject concept to the sub-concepts that represent the nine LOM categories using nine relations. This way, the LearningObject concept is attached to the LomGeneralCategory concept by the relationship “hasGeneralCategory”.

The LOM General category defines the notion of identifier for a learning object as “a globally unique label that identifies this learning object”. The identifier is characterized by a “catalog” and an “entry”. The “Identifier” is an element of type “c”, and so we apply $R_6$. It is represented by a concept of the ontology. The Catalog and Entry elements are of type “b”, and so they are represented by attributes of the Identifier concept. A relationship “hasIdentifierElement” between the two concepts, LomGeneralCategory and Identifier, is linked the identifier to the LOM general category.

The “Language”, “Description”, “Keyword”, “Coverage” and “Title” are also elements of type “b”, and they are therefore represented with attributes of the LomGeneralCategory concept. The “Structure” and “AggregationLevel” elements are of type “a”, and they are therefore represented as concepts. Two relationships, “hasStructureElement” and “hasAggregationLevelElement”, are created whose domain is the “LomGeneralCategory” concept, and ranges are the concepts “Structure” and “AggregationLevel” respectively. The values for the Structure element (Atomic, Collection, Networked, Hierarchical, Linear) are represented by individuals. We apply our rules in the same way for the remaining categories.

Since our approach is also motivated by the fact that learning objects can be linked with other objects, we will give details on some of these relationships as follows.

4.3. Relationship deduced

In the LOM “relation” category and especially for the element “kind”, we note in particular the values “is part”, “requires”, “has version”. These values can actually link a learning object with other learning objects. Therefore we have used the proposed values for this element to create relationships that will link a learning object to another, that is to link an instance of the concept LearningObject to another instance of this same concept. Although it was described in a particular category, we nevertheless considered that the induced relations had for domain and range the concept LearningObject and not the one of the original category.

As an example, a course on databases under the form of a website. This learning object consists of several web pages, each showing the contents of a chapter. Assume that the first chapter entitled “Introduction to databases and Database Management System (DBMS)” is followed by a second devoted to “relational models”, and then another on the “SQL”. The first chapter contains a figure illustrating the DBMS’s architecture considered also as a learning object. In this case the relation is part of will allow to connect the 3 chapters throughout the entire course so that this website and its components could be indexed (see Figure 2).

The “structure” element of the “General” category also allowed us to create new relationships. If this element is set to “linear” value this means that the learning object is itself a set of fully ordered learning objects. We therefore deduce that this LO is made of other LOs, which can be linked by the relation “has next” or “has previous”. Using the same example as above and considering that the course structure on databases is linear, it is interesting to say that Chapter 1, entitled “Introduction to databases and the DBMS” is part of the course on databases, that it has next chapter 2 entitled “relational models”.

The value “hierarchical” of the “structure” element, which defines the learning object as a set of resources whose relationships can be represented by a tree structure, led us to create relationships like ascendant and descendant between learning objects that constituting the considered object.

To represent our ontology we chose OWL\textsuperscript{15} (Web Ontology Language) which is recommended by W3C and used Protégé\textsuperscript{16} and TopBraid\textsuperscript{17} editor.

5. Semantic indexing

In our view, semantic indexing allows to associate ontological elements, whether they are concepts (most frequently), relationships, attributes, or even individual (enumerated concept) to a learning object. A learning object can be associated with several ontological elements and on the other hand an ontological element can denote many learning objects.

However, the association must be semantically specified. This is why we consider that only a semantic description made up of statements instantiating relations

\footnotesize
\textsuperscript{15}Ontology Web Language: http://www.w3.org/2004/OWL/
\textsuperscript{16}http://protege.stanford.edu
\textsuperscript{17}http://www.topquadrant.com/index.html
Fig. 2. Instantiation example in the LOM’s ontology

and attributes of the ontology can represent a LO. In our model, there is a sequence of RDF triples. Indexing a LO comes, therefore, to create such a semantic description.

In this section, we begin by presenting the LOM ontology based tool that we developed for indexing and retrieving LOs. Next, we discuss the LOM ontology integration in the MEMORa2.0 project.

5.1. Presentation of the indexing tool

In order to overcome the problems associated to the use of the standard LOM in LORs, we developed a specific tool with two objectives: indexing a distributed set of resources referenced by their URL/URI, and helping to retrieve for these resources. In addition to these objectives, the tool provides a support for a good understanding of the different items.

5.2. Architecture

Our tool LOIT (learning objects indexing tool) is a component that can be added to any Java web application. It is built using the struts\textsuperscript{18} open-source framework that allows to develop web applications using MVC (Model-View-Controller) architecture. LOIT is composed of three main forms:

- A form that allows entering information to describe learning objects.
- A second form is used to enter search criteria according to the nine categories to perform an advanced search.
- A third form is used to enter search criteria to perform a simple search.

The first two forms rely on nine java classes, each representing a category of LOM (Figure 3). The two actions, index and search, are defined in the struts configuration file (struts-config.xml).

The user interaction is supported by JSP pages that present the ontology concepts and allow to insert values for describing or searching a resource (Figure 4). The knowledge base is managed by the Jena\textsuperscript{19} framework, whose last version extends the notion of property in order to access concepts using any keyword. The knowledge base exported in OWL also allows interoperability with systems using another technology. The possibility of finding and retrieving particular LO is a key issue for users in LOR.

\textsuperscript{18}http://struts.apache.org/

\textsuperscript{19}Jena, a Semantic Web Framework for Java : http://jena.sourceforge.net/
Internally, descriptions of resources are transformed in an RDF knowledge base, compliant with our ontology, that can be exported as an OWL file. Discovery of resources is performed thanks to SPARQL\(^20\) requests against this knowledge base. To make more explicit the standard elements, each label is a link that refers to the page of the adequate LOM ontology documentation.

5.3. Search for learning objects

The simple search interface is intended for novice users who do not know the structure of the LOM and want to search without browsing the nine categories. The query is written using the LARQ\(^21\) technology.

---

\(^{20}\)SPARQL: http://www.w3.org/TR/rdf-sparql-query

\(^{21}\)http://jena.sourceforge.net/ARQ/lucene-arq.html/
which is a combination of ARQ\textsuperscript{22} and Lucene\textsuperscript{23}. It is a library written in java allowing to create different kinds of indexes and provides a full-featured text search engine. For example, the interface enables the user to search for resources whose keyword field contains the word "sql" and not the word "sparql", combining free text search from several fields (for more examples, see [5]).

To write queries, the wildcard characters are used, such as an asterisk (*) or a question mark (?) which respectively represent zero to several characters and one character. The symbols (+) or (-) are also used to represent respectively, the Boolean operators "and" and "not" (Figure 5).

\begin{figure}
\centering
\includegraphics[width=\textwidth]{query.png}
\caption{A simple search query}
\end{figure}

5.4. Advanced search

The advanced search interface allows users to fill in the fields that will enable them to regain the learning objects they need. In the same way, a user can use one or more elements of the nine categories. When he runs his request, the URL/URI of learning objects that answers the search criteria are displayed in user interface presented.

For example, in order to search for the resources written in French, with an atomic structure, aggregation level equal to 1 and associated with the “SQL” word, the user inputs values for language, structure, aggregation level and language elements respectively of the General category. The tool then builds the request shown in Figure 6. A list of triples linking the variables to the user values is created and the URLs of the corresponding resources are returned in the ?entry variable. In the query interface, the response is displayed as links.

The constructed ontology allowed us to build a tool that facilitates the use of a standard such as LOM for indexing the LO, and allows the reduction of noise and silence in a search. LOIT can be considered as LOR because if it can not store the LOs locally, it stores their links. This tool can be also used in LOR to facilitate the LO indexing by our ontology concepts.

6. Conclusion

In order to solve the problems of management, sharing and access to resources we have presented our approach to semantically indexing the learning objects through the concepts of LOM ontology that we constructed. We have explained the construction methodology of this ontology, then we presented the tool developed for indexing and retrieval of learning objects.

Our ontology based on the LOM mainly describes the form of resources and not directly their content. For example, it is possible to say that a resource is a very difficult exercise, but it is not possible to say that it is a math’s exercise about the Pythagora’s theorem.

In the MEMORAe2.0 project, the resources are organised by means of two ontologies [1]. The first one (domain ontology) describes the concepts of the 'organization' domain. The second ontology (application ontology) specifies notions which are used by members of a particular organization. This ontology concerns the course B31.1 on applied mathematics at the University of Picardy (France).

In the domain ontology, the resource concept is defined as a learning content that should exist as such, and can be retrieved and indexed easily. This definition fits perfectly the learning object concept defined in the LOM ontology. Therefore we decide to replace the modelling of Resource concept by the one of learning object concept. This one will inherit the sub-concepts and relationships of the Resource concept, in addition to concepts and relationships defined in LOM ontology. The ontology domain extended, thereby allows

\begin{verbatim}
PREFIX lom: <http://www.w3.org/2002/07/owl/lo.owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT ?entry
WHERE {  
    ?y lom:hasIdentifierElement ?x .  
    ?y lom:keyword "SQL" .  
    ?y lom:hasStructureElement "atomic" .  
    ?y lom:aggregationLevel "1" .
}
\end{verbatim}

Fig. 6. SPARQL query

\textsuperscript{22}http://jena.sourceforge.net/ARQ/
\textsuperscript{23}http://lucene.apache.org/java/docs/index.html
a better description. Previously the information used were only the author name, the description and the date of availability of the resources in the organizational memory.

The link between the ontology domain extended and the application ontology allows describe the contents of LO and their form with LOM standard. However the concept “keyword” associated with the concept representing “general category” becomes then redundant and less effective.

So, we plan to change free text in the keyword element by concepts among application ontologies concepts. These application ontologies can be downloaded by the users.

References


