Checking Semantic Consistency of SCORM like Learning Objects

Ramzi FARHAT1,2, Bruno DEFUDE1 and Mohamed JEMNI2

1UMR CNRS-INT SAMOVAR
Institut TELECOM, TELECOM SudParis
Evry, France

2 Research Unit of Technologies of Information and Communication (UTIC)
Ecole Supérieure des Sciences et Techniques de Tunis
Tunis, Tunisia

ramzi.farhat@esstt.rnu.tn, Bruno.defude@it-sudparis.fr, Mohamed.jemni@fst.rnu.tn

Abstract—Semantic metadata are not yet fully integrated in most learning metadata profiles. Moreover, when it is supported it is only used to improve the quality of results returned by search engines in LOR (learning object repositories) and LMS (Learning Management Systems). Within the framework of APOGE1 project, we want to demonstrate the usability of semantic metadata, when they are provided, during the authoring phase to assist authors to improve the consistency of new designed learning objects. The added value of our approach is maximized in the case of SCORM (Sharable Content Object Reference Model) like learning objects’ authoring. Indeed, when we have learning objects designed by reuse of existing ones there is risks to have weaknesses mainly due to a bad understanding of the true nature of the reused objects. To overcome those risks we have defined an automated authoring assistance approach. In this paper we put the emphasis on two semantic based methods used in our approach: semantic consistency checking and learning object’s semantic space analysis.

Learning object; semantic metadata consistency; authoring approach

I. INTRODUCTION

Educative metadata are in the core of works around learning objects (LO). They are mainly used in the discovery of learning objects within repositories. However some research works demonstrates that those metadata are also important for other kinds of use such as personalization of the learning experience by comparing some metadata to learners’ profiles [1][2].

SIMBAD [3] is one of these approaches which use a rich metadata profile including semantic metadata. This metadata profile is used to improve the relevance of results returned by the search engine when a user is looking for a LO and it is used for personalization purposes too. Semantic metadata cover three aspects: prerequisites (precondition on the access), content and acquisition function (concepts that are added to the learner profile if he/she succeeds with this LO). Prerequisites and acquisition function entries contain three parts <concept, role, level>. The concept information is taken from a domain ontology describing the education concepts and the relationship between its. The role is one of the educative roles that a LO can play concerning a learning concept such as exercise, example, definition, etc. The level is a discrete value from {low, intermediate, high}. For example, if a LO has as prerequisite the triple <‘Relational databases’, ‘definition’, ‘intermediate’> this means that learners must have – according to their profile - the required knowledge about relational databases with an intermediate level to be able to access to this LO content. The content of a LO is described by a set of couples having the following structure: <concept, role>.

With the introduction of numerous learning object repositories, a new way to author LOs appears, we call authoring by reuse [4]. In this authoring mode, an author constructs new LOs by assembling some existing ones combined together using dedicated operators. SCORM [5] and SIMBAD are examples of such models of composed LOs. SIMBAD proposes a richer composition model with three operators (sequence, alternative and parallel). If it is very easy for an author to design LOs in such a way, it is not so easy to define consistent LOs. The problem we are focusing in this article may be formulated as follows: how semantic metadata can be used to discover possible anomalies and weaknesses of composed LOs during the authoring process?

The complexity of this task is increased by the fact that the reused LOs can be recursively composed by other ones. We called author view (the higher level of abstraction) the abstract composition graph. It is the graph designed by the author to describe the navigation strategy between the reused LOs. And we called system view (the lowest level of abstraction) the concrete composition graph. It is the graph generated by the system by replacing each composed LO by its composition in a recursive way until having a graph with only atomic LOs (assets in the SCORM vocabulary). These

1 This work is partly funded by the APOGE project (French – Tunisian Utique CMCU)
notions are illustrated by the example of figure 1. In most cases, the user view can be very different from the system view, complicating the task of the author.

Figure 1. Illustration of the levels of abstraction of a composed LO: the case of the LO lo_903

We propose in this article two methods to assist authors constructing consistent LOs based on the use of semantic metadata. In the next section we describe the semantic consistency checking method. In the third section we present the semantic space analysis method. Finally, we conclude and give some open issues.

II. SEMANTIC CONSISTENCY CHECKING

A. Overview of our design by reuse approach

We propose to adopt a design process in five steps [4]. In the first step, an author searches for existing LO issuing a query on a repository (either local or distributed). This query produces as output a set of references to LOs. During step two, from these objects the author can compose a new one. Step three consists in a deep analysis of the new LO to help author to have a good understanding of its production. After that, the author can choose between automatic generations of part of LOM metadata (from those of reused LOs) or define these metadata by hand. Finally in step five, high level consistency rules (pedagogical ones) are applied. At the end, the new LO is added to the repository.

The first question we address in this article is about the consistency of author-defined semantic metadata (prerequisites, content and acquisition) compared to the semantic metadata of the reused LOs. To help author to answer this question we propose a method covering two aspects: (i) the verification of prerequisite satisfaction of the reused LOs and (ii) the verification of the content consistency by comparing the composed LO’s content to the reused LOs’ content.

B. Prerequisites satisfaction

For each learning object there is a set of prerequisites that must be satisfied by learners’ profiles who want learn the LO’s content. When an author composes a new learning object by reuse of existing LOs the question is if all the prerequisites of the reused LOs are satisfied?

Prerequisite is satisfied in two cases: (i) when the prerequisites are part of the learner profile or (ii) when the prerequisites are satisfied by a previous LOs’ acquisition function in the abstract composition graph.

Let us consider as example the learning introduced by the figure 1 and having the identifier lo_903. It has the abstract composition graph at the top of the Figure. We suppose that these LOs have the following semantic metadata:

- Semantic metadata of lo_903:
  - Prerequisite: {“Relational theory”, “principal”, “Intermediate”}, {“Relational database”, “Introduction”, “Beginner”}
  - Content: {“Relational databases”, “example”, “exercise”}
  - Acquisition function: {“Relational databases”, {“principal”, “example”, “exercise”}, “intermediate”}

- Semantic metadata of lo_902:
  - Prerequisite: {“Relational theory”, “principal”, “Intermediate”}
  - Content: {“Relational databases”, “Principal”, “Example”}
  - Acquisition function: {“Relational databases”, {“principal”, “example”}, “Intermediate”}

- Semantic metadata of lo_006:
  - Prerequisite: {“Relational databases”, “Principal”, “Intermediate”}
  - Content: {“Relational databases”, “Exercise”}
  - Acquisition function: {“Relational databases”, “Exercise”, “Beginner”}

The lo_902 LO has one prerequisite that is satisfied because it is part of lo_903’s prerequisites. So, this prerequisite must be satisfied by the learner profile accessing lo_903.

The unique prerequisite of the lo_006 LO is also satisfied but in this case by the acquisition function of lo_902 which precedes it in the composition graph.

This example demonstrates that even if the abstract composition graph is simple and the number of semantic metadata is small, it is not so easy to the author to check the prerequisites satisfaction by himself. That is why we propose a checking algorithm to do that automatically.

To introduce the related algorithm we must first define an order relation (≺) between <concept, role, level> triples. We consider that c ≡ c’ when these two concepts c and c’ are in relationship via a generic/specific relation or a rhetoric relation (contrast, similar or equivalent) within the domain ontology.
The following algorithm describes the verification of prerequisites satisfaction. In this paper we will use the notation below:

\( G_i \) is the abstract composition graph of the LO \( i \). \( P_i \) is the set of prerequisites of \( i \). \( A_i \) is the set of acquisition functions of \( i \). \( C_i \) is the set of contents of \( i \). \( \text{prev}_i \) is the set of LOs which must be taken by the learner before \( i \) according to the abstract composition graph defined by the author. We will use \( \text{clo} \) as identifier for the composed learning object under verification.

**Algorithm 1**

For all \( \text{clo} \in G_{\text{clo}} \) do

For all \( \text{lo} \in P_{\text{lo}} \) do

If \( \{ \exists x / \text{lo} \prec x, x \in (P_{\text{clo}} \cup A_i) / \text{prev}_i \} \) then

Return (false)

End

End

Return (true)

---

**C. Content’s semantic description consistency**

Within SIMBAD approach there are two levels of semantic descriptions. First, the semantic metadata filled by the author. Secondly, the semantic metadata of the reused learning objects. Those two semantic descriptions must be consistent and any possible inconsistency is probably a design anomaly. Thus it is necessary to check the content semantic consistency.

First, we must verify if the prerequisites are consistent compared to the semantic metadata of the reused learning objects. Here the verification is not the same as that done above. In fact we must verify that no prerequisite (of the composed learning object) is equal or greater that an acquisition function of one of the reused learning object.

For example, supposing that the \( \text{lo}_903 \) has as prerequisite \(<\text{"Relational databases"}, \text{"Exercise"}, \text{"Beginner"}>\). The question is: why demanding this prerequisite when we will propose a learning object offering the same kind and level of knowledge?

To verify this facet of content’s semantic description consistency we have defined the following algorithm.

**Algorithm 2**

For all \( (p \in P_{\text{clo}}) \) do

For all \( (i \in G_{\text{clo}}) \) do

For all \( (a \in A_i) \) do

If \( (a \prec p) \) then

Return (false)

End

End

End

Return (true)

---

III. LEARNING OBJECT’S SEMANTIC SPACE ANALYSIS

In the section II we have presented an approach to exploit semantic metadata of the composed LO and LOs in the abstract composition graph to check the semantic consistency. The goal is to help author to discover possible anomalies and to fix them if any. In general the author must modify the incorrect semantic metadata entries and/or modify the structure of the abstract composition graph.

In this section we suppose that the structure and semantic of the composed LO are already consistent. However, how we can help author to improve his/her composed LO from an educative point of view?

In previous works [4][7] we have introduced semantic, structural and hybrid metrics to generate various measurements. They must help author to have a best view of the different facets of the learning object, especially hidden ones: measurements in relation with the concrete composition graph (CCG). The importance of CCG relies in the fact that it is composed of the concrete media files that will be provided to the learner. For example, the author can discover that the CCG provides too much alternatives to the learner. In some other cases he can discover that some paths in the CCG do not include exercises.

In this context we have investigated about the use of semantic metadata describing the content of LOs within the CCG in our authoring approach.

Thus we propose to introduce the notion of LO’s semantic space. A semantic space, in our approach, is a subset of the domain ontology containing only concepts
describing the content of LOs in the CCG and specific/generic relationship existing between them.

The LO’s semantic space must contain information defining what are the concepts provided to learners and how they are distributed within the domain ontology. The importance of this analysis relies on the fact that metadata of the composed LO and LOs in the abstract composition graph, generally summarize, contract and condense the information. Our goal is to give detailed and precise information that cannot be easily extracted and analyzed directly by the author.

A. Computation of LO’s semantic space

The semantic space is defined as the projection of the content’s concepts of all atomic learning objects (i.e. learning objects in the CCG) into the domain ontology and specific/generic relationships between them. The figure 2 shows three semantic spaces of three different LOs.

In a formal way the semantic space \( S_{\text{clo}} \) is defined by two sets: \( S_{\text{clo}} = \{SC_{\text{clo}}, SR_{\text{clo}}\} \). The first set \( SC_{\text{clo}} \) contains the concepts included in the semantic space. The second set \( SR_{\text{clo}} \) contains the specific/generic relationships between concepts in \( SC_{\text{clo}} \). Concerning the notation, let \( O'_{\text{clo}} \) be the set of LOs in the concrete composition graph of the composed \( LO_{\text{clo}} \). Let \( R \) be the set of specific/generic relations between concepts in the domain ontology. We use the following algorithm to compute the semantic space.

**Algorithm 5: Computation of semantic space**

*For all \( (i \in \ O'_{\text{clo}}) \) do*

*For all \( (c \in C_i) \) do addConceptToSC(\( S_{\text{clo}}, c \))

*End*

*End*

*For all \( (r \in R) \) do*

*If fromConcept(\( r, c_i \)) and toConcept(\( r, c_j \)) and \( \{c_i, c_j\} \subset S_{\text{clo}} \) then addRelationshipToSR(\( S_{\text{clo}}, r \))

*End*

*End*

To help authors of composed LOs to have a better understanding of their LOs, we propose an analysis of the semantic space. Our analysis provides three kinds of information about the semantic space: shape, width and depth.

B. Shape of a LO’s semantic space

In our approach we provide to authors their semantic space’s shape. This kind of information must help the author to verify if the distribution of the concepts is in consistency with what it is expected.

The figure below shows three characteristic shapes of three different composed learning objects. In the case (a) we have a characteristic shape of a learning object with a content dealing with concepts from the general to the specific (or the opposite). In the case (b) we have a characteristic shape of a learning object with a content covering grouped transversal concepts. In the case (c) we have an irregular shape with concepts scattered all other the domain ontology which reflects in most cases a badly designed learning object.

C. Width and depth of a LO’s semantic space

To have more precise information than those given by the shape, we propose to analyze the LO semantic space to determine its depth and width.

We consider the width of a semantic space as the number of “leaf” nodes present in the space. A “leaf” node is a node with no child is within the semantic space. Let consider the example of the three semantic spaces in figure 5. In the first case (a) the width is 1 because only one node (marked with x in the figure) do not has a child within the semantic space. In the second case (b) the width is 5. In the last case (c) the width is 5 too.
Algorithm 6: Computation of semantic space width

\[
i \leftarrow 0 \\
\text{Width} \leftarrow 0 \\
\textbf{For all} \ (c \in SC_{clo}) \ \textbf{do} \\
\quad \textbf{IF} \ \exists r \in SR_{clo}/toConcept(r,c) \ \textbf{do} \ i \leftarrow i+1 \\
\quad \text{End} \\
\text{Width} \leftarrow \text{Max} (\text{Width}, i) \\
\text{End} \\
\textbf{Return} \ (\text{Width})
\]

The depth of a semantic space is the size of the longest path between root nodes (nodes with no parent within the semantic space) and leaf nodes (nodes with no child within the semantic space).

The depth of a semantic space is the size of the longest path between root nodes (nodes with no parent within the semantic space) and leaf nodes (nodes with no child within the semantic space).

Let consider the example of the three semantic spaces in the Figure 2. In the first case (a) the depth is 3. In the second case (b) the depth is 1. In the last case (c) the depth is 1 too.

The width and depth can be interpreted in different ways. For example, they can indicate if there is adequacy or not between the content and the granularity of a LO. Generally, large depth and width are characteristic of a learning object with large granularity (e.g. Lesson), else it is the case of a small granularity (e.g. Asset).

Moreover width and depth can be used to control the homogeneity of the LO compared to a reference set of LOs. This reference set can be composed of all the LOs in a repository. Here the goal is to conserve the homogeneity of the repository. In other cases the reference set can contain the published LOs of the same author. Here the goal can be to verify if the new composed LO is similar or no to author published LOs. Another possible use case can be the comparison of the new LO with the most used (or most rated) LOs in a repository.

IV. CONCLUSION

Semantic metadata can be used in several ways and for several purposes. In this paper we have used semantic metadata as a support for author assistance in the case of learning objects designed by reuse of existing ones. We have underlined the semantic metadata usability to (i) check the semantic consistency of a learning object and to (ii) generate LO’s semantic space based indicators. This effort is justified by the complexity leaded by learning objects’ reuse. With automated semantic checking and analysis, an author will have more feedbacks, indicators and information to fix or to improve his/her designed learning object before its publication in repositories.

We think that LO’s semantic space analysis can produce high level guidelines to design learning objects. We are working on this direction, conducting experiments on a set of SCORM resources. Experimentation results must help on the classification and automatic discovery of typical anomalies. Moreover, a correlation between semantic indicators and weaknesses can be established too based on authors’ behavior.

REFERENCES