Towards a better understanding of learning objects’ content

Ramzi FARHAT  
Research Unit of Technologies of Information and Communication (UTIC)  
University of Tunis  
Tunis, Tunisia  
Ramzi.Farhat@esstt.rnu.tn

Bruno DEFUDE  
UMR CNRS-INT SAMOVAR  
Institut TELECOM, TELECOM SudParis  
Evry, France  
Bruno.Defude@it-sudparis.fr

Mohamed JEMNI  
Research Unit of Technologies of Information and Communication (UTIC)  
University of Tunis  
Tunis, Tunisia  
Mohamed.jemni@fst.rnu.tn

Abstract—Usually semantic metadata are introduced to support effective search of relevant learning objects from LOR (Learning Object Repositories). We propose to use them to assist authors of SCORM (Sharable Content Object Reference Model) like learning objects. In fact, when we design new learning objects by reuse of existing ones, there are 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 focus especially on how we use semantic metadata to produce indicators about the content of learning objects, more precisely about their complexity, heterogeneity and imbalance from a semantic perspective. To attend this specific objective we introduce the notion of “semantic space” and “semantic component”. Then we use them to compute metrics which are aggregated together to produce meaningful indicators.

I. INTRODUCTION

The reuse of learning objects can be applied at two different levels. The first consists on the reuse of the same learning object in different learning contexts and by different users. The second level consists on the reuse of learning objects to build new ones. For example, an author can reuse definitions, exercises and examples defined each one as a learning object to build a new one.

We are concerned by this second level of learning objects’ reuse which is already supported by many proposals like SCORM [1], IMS [2] and SIMBAD [3]. However, those proposals don’t introduce any specific authoring approach. Moreover, they don’t discuss issues related to author assistance in order to improve the quality of new learning objects composed by reuse.

In [4] we propose an authoring process in five steps. In the first step, an author searches for existing learning objects issuing a query on a repository (either local or distributed).

This query produces as output a set of learning objects’ references. They are reused during step two by author to compose a new learning object. Step three consists in a deep analysis of the new learning object to help author to have a good understanding of its production. After that, the author can choose between automatic generations of part of LOM [5] metadata (inferred from reused learning objects’ metadata) or defining them by hand. Finally in step five, high level pattern based conformity checking is applied. At the end, the new learning object is added to the repository.

In this paper we focus on the semantic indicators used in the third step of our authoring approach. As mentioned above, the goal of the third step is to help authors to have a good understanding of their production by using significant indicators. Some of them are called semantic indicators because they use semantic metrics.

In the two next sections we will introduce respectively the notion of semantic space and the notion of semantic component and their associated semantic metrics. In the fourth section we will explain how we use semantic metrics to build semantic indicators about content complexity, continuity and imbalance. In the fifth section we will discuss how those indicators can be used and interpreted by authors. In the last section we conclude the paper and introduce future works.

1 This work is partly funded by the APOGE project (French – Tunisian Utique CMCU)
II. SEMANTIC SPACE

We suppose that learning object’s content is defined by a set of concepts of an ontology domain (see figure 2). This can be done using keywords attribute of the IEEE LOM standard where keywords are controlled by an ontology domain. The semantic space of a given learning object is constructed upon the domain ontology by keeping concepts describing the content and specific/generic relationships among them. The semantic space of the learning object $lo$ is defined as following:

$$ SS_{lo} = \{SC_{lo}, SR_{lo}\} $$

The $SC_{lo}$ set represents domain model concepts describing content of the learning object $lo$. The $SR_{lo}$ set represents specific/generic relationships among concepts in the $SC_{lo}$.

Our objective here is to have the distribution of the concepts within a hierarchical view (only specific/generic relationships are considered) of the domain model. This particular view of learning object content’s concepts distribution is sufficient in our opinion to analyse its complexity, continuity and imbalance from a semantic perspective.

We consider that the learning object’s content is described by a set of entries. Each one is composed by a couple of data: a concept from the domain model and an educative role played by the content regards to the concept (e.g. Introduction, Exercise, Definition, etc.). Let’s $LO_{222}$ be a learning object having the following semantic metadata describing its content:

$$ \{<\text{Algorithm}, \text{Introduction}>, <\text{Algorithm}, \text{Definition}>, <\text{Type}, \text{Introduction}>, <\text{Simple type}, \text{Introduction}>, <\text{Variable}, \text{Definition}>, <\text{Variable}, \text{Example}>, $$
$$ <\text{Constant}, \text{Definition}>, <\text{Constant}, \text{Example}>, <\text{Corps}, \text{Introduction}>, <\text{Instruction}, \text{Definition}>, <\text{Writing}, \text{Definition}>, <\text{Writing}, \text{Example}>, $$
$$ <\text{Reading}, \text{Definition}>, <\text{Reading}, \text{Example}>, <\text{Affectation}, \text{Definition}>, <\text{Affectation}, \text{Example}>, <\text{Affectation}, \text{Exercise}>, <\text{Conditional}, \text{Definition}>, $$
$$ <\text{Conditional}, \text{Exercise}>, <\text{Iteration}, \text{Definition}>, <\text{Iteration}, \text{Example}>, <\text{Iteration}, \text{Exercise}>\} $$

And we will consider the following ontology as a model covering the algorithmic domain:

By using both semantic metadata and the domain model the following semantic space can be calculated (see figure 3):

![Figure 3. LO_222 semantic space](image)

From the semantic space we have defined four metrics. The first one is the number of concepts within the semantic space. In the case of LO_222 the number of concepts is 13. The second one is the number of connected graphs within it. A connected graph is a graph such that there is a path between any pair of nodes In the case of LO_222 the semantic space contains one connected graph. The third metric is the distance in terms of transitions between connected graphs if there are two or more. In the case of LO_222 there is one connected graph so the distance is considered as equal to zero. And the fourth metric is the weight of each concept. We mean by “concept’s weight” the number of learning objects (reused to compose the one under analysis) which consider the concept in their content.

III. SEMANTIC COMPONENT

We call “semantic component” a group of concepts in semantic proximity within the semantic space of a learning object. Each semantic component is a connected graph. Two different concepts are in the same semantic component when the distance between them is equal or less to a fixed threshold. This distance is computed in term of number of transitions in the shortest path between two concepts in the semantic space.

Thus, we can redefine a semantic space as a couple $\{CS, s\}$ where $CS$ is a set of semantic components and $s$ the threshold of semantic proximity.

Each semantic component $c_i$ is a set of concepts $C$ where:

$$ \forall \{c_x, c_y\} \subseteq C, d(c_x, c_y) \leq s $$

If we take as example the learning object LO_222 and a semantic proximity threshold $s$ equal to two, figure 4 shows the five different semantic components within the LO_222 semantic space.
From the semantic component we have defined three metrics. The first one is the number of semantic components within the semantic space. In the case of LO_222 there are 5 semantic components. The second metric is the number of intersections between them. In the case of LO_222 this is equal to 7. The third metric is the weight of each semantic component. This weight is computed in term of number of learning objects considering at least one of the concepts of the semantic component in their content.

### IV. SEMANTIC INDICATORS

Those different metrics are not directly significant for a learning object's author. In the same time they describe concrete properties of a learning object. Thus we have defined meaningful indicators (from authors perspectives) based on formulas using learning objects' metrics. Those indicators cover three aspects: complexity, continuity and imbalance of learning object content.

#### A. Content complexity

The content of a learning object is considered as complex when it is difficult to be assimilated by learners. One of the factors influencing the complexity is certainly the number of concepts and how they are spread in the domain ontology. In fact if the learning object's content presents a number of concepts exceeding a threshold they can be considered as complex. This threshold of complexity must be fixed by experts depending on learning context (educative domain, learners’ age, etc).

However, the number of content concepts is not the unique metric influencing the content complexity indicator. In fact, for example if we have few concepts within the content but if they are sparse in the domain ontology the content is still complex to be assimilated by the learner. Another example, if we have a high number of concepts and those concepts are similar the content is not necessary difficult to be assimilated.

Based on this theoretical analysis and on empirical study we have defined the content complexity indicator by the following formula:

$$
\left( c_1 \frac{\log(x)}{\log(\text{Max}(\alpha, x))} + c_2 \frac{\log(y)}{\log(\text{Max}(\beta, y))} \right) \times \frac{E}{\sum c_i}
$$

$x$: number of concepts in the content  
$y$: number of semantic components in the content  
$\alpha$: threshold of complexity in term of number of concepts  
$\beta$: threshold of complexity in term of number of semantic components  
$c_i$: weighting coefficient  
$E$: scale

The formula is based on two terms. The first one increases the value of complexity indicator when the concepts number goes up and the second term increases the value of the indicator when the number of the semantic components goes up. The empirical study shows that the values must go up in a logarithmic way.

In the case of the learning object LO_222, with weighting coefficients $c_i$ equal to one, thresholds $\alpha$ and $\beta$ equals respectively to 100 and 10 (in practice experts must fix those values according to the size of the domain ontology and the educative context) and a scale from zero to ten, the value of the indicator is calculated as following:

$$
\left( \frac{\log(13)}{\log(100))} + \frac{\log(5)}{\log(10))} \right) \times \frac{10}{2} = 6.28
$$

#### B. Semantic continuity

Moving from the study of a concept to the study of another one in the same learning object is certainly easier if there is a transition via intermediate concepts. We call this property “semantic continuity of the content”.

In this context we use two metrics to define the semantic continuity indicator. The first metric is the number of connected graphs in the semantic space. In the best case it must be a unique connected graph. The second metric is the number of intersections between the semantic components. A large intersection number is a positive sign.

Based on this theoretical analysis and on empirical study we have defined the semantic continuity indicator by the following formula:

$$
\left( c_3 \frac{1}{1 + \sum d_j} + c_4 \frac{1}{1 + \text{min}(0, y - (n - l))} \right) \times \frac{E}{\sum c_i}
$$

$x$: number of connected graphs in the semantic space  
$y$: number of intersections between semantic components  
$n$: number of semantic components  
$\sum d_j$: sum of distances between connected graphs  
$c_i$: weighting coefficient  
$E$: scale

The formula is based on the sum of two terms. The first term decreases with the sum of distance between connected graphs in the semantic space. The second term increases
when the number of intersections between semantic components goes up.

In the case of the learning object LO_222, with weighting coefficients \(c_i\) equal to one, and a scale from zero to ten, the value of the indicator is calculated as following:

\[
\left( \frac{1}{1 + \frac{1}{1 + \left| \min(0,7 - (5 - 1)) \right|}} \right) \times \frac{10}{2} = 10
\]

C. Semantic imbalance of content

The learning object’s semantic space summaries the concepts treated by the content and how they are spread in the domain model. However there is no information about how each concept or group of similar concepts is covered by the content. In fact the goal is to check if there is some balance or not.

We propose in this context a semantic imbalance indicator based on the weight of each concept or semantic component. The weight is calculated in term of reused learning objects within the content in the case of a composed learning object.

We have defined the semantic imbalance indicator by the following formula:

\[
\left( c_a \frac{\log(1 + (M - m))}{\log(1 + D)} + c_b \frac{\log(1 + (M' - m'))}{\log(1 + D)} \right) \times \frac{E}{\sum c_i}
\]

\(M\): maximal concept weight within the semantic space
\(m\): minimal concept weight within the semantic space
\(M'\): maximal semantic component weight within the semantic space
\(m'\): minimal semantic component weight within the semantic space
\(D\): theoretical maximal difference between weights
\(c_i\): weighting coefficient
\(E\): scale

This formula is based on the sum of two terms that increase when the ratio between real weight and theoretical maximal weight goes up. The first term concerns concepts and the second term concerns semantic components.

In the case of the learning object LO_222, with weighting coefficients \(c_i\) equal to one, a maximum and a minimum weights of concepts equal respectively to 3 and 1, a maximum and a minimum weights of semantic components equal respectively to 6 and 2, a theoretical maximum difference of weight equal to 11, and a scale from zero to ten, the value of the indicator is calculated as following:

\[
\left( \frac{\log(1 + (3 - 1))}{\log(1 + 11)} + \frac{\log(1 + (6 - 2))}{\log(1 + 11)} \right) \times \frac{10}{2} = 5.45
\]

V. INTERPRETATION OF THE INDICATORS

The different indicators give values in a scale varying from the minimum value (zero) to the maximum value (E). But in the absolute, these values are not necessary significant to the author. That is why we propose two possible ways to give a significant interpretation of indicators’ values.

The first interpretation is based on the use of discrete values: low (values from 0 to 3.33), medium (values form 3.34 to 6.66) and high (values form 6.67 to 10). In fact, it is more significant for the author to know for example that the semantic imbalance of the content is medium. This information will drive the author to examine the content and to improve its continuity. This can be done by adding appropriate content (reused learning objects) to improve the balance.

The second interpretation is based on the comparison of the values to the reference values relative to a particular set of learning objects. This particular set of learning object can be the set of the author’s learning objects or the set of the best rated learning objects or the set of similar learning objects.

This formula is based on the sum of two terms that increase when the ratio between real weight and theoretical maximal weight goes up. The first term concerns concepts and the second term concerns semantic components.

In the case of the learning object LO_222, with weighting coefficients \(c_i\) equal to one, a maximum and a minimum weights of concepts equal respectively to 3 and 1, a maximum and a minimum weights of semantic components equal respectively to 6 and 2, a theoretical maximum difference of weight equal to 11, and a scale from zero to ten, the value of the indicator is calculated as following:

\[
\left( \frac{\log(1 + (3 - 1))}{\log(1 + 11)} + \frac{\log(1 + (6 - 2))}{\log(1 + 11)} \right) \times \frac{10}{2} = 5.45
\]

V. INTERPRETATION OF THE INDICATORS

The different indicators give values in a scale varying from the minimum value (zero) to the maximum value (E). But in the absolute, these values are not necessary significant to the author. That is why we propose two possible ways to give a significant interpretation of indicators’ values.

The first interpretation is based on the use of discrete values: low (values from 0 to 3.33), medium (values form 3.34 to 6.66) and high (values form 6.67 to 10). In fact, it is more significant for the author to know for example that the semantic imbalance of the content is medium. This information will drive the author to examine the content and to improve its continuity. This can be done by adding appropriate content (reused learning objects) to improve the balance.

The second interpretation is based on the comparison of the values to the reference values relative to a particular set of learning objects. This particular set of learning object can be the set of the author’s learning objects or the set of the best rated learning objects or the set of similar learning objects.
The author can discover that for the imbalance indicators he is not in conformance with the mean value of the set of learning object considered as a reference. This information will encourage him/her to re-examine the content and to fix the semantic imbalance of the learning object.

We have proposed two possible ways to help the author to interpret the calculated indicators’ value. But additional researches must be done to discuss the most appropriate way to assist authors to a best understanding of indicators’ meaning.

VI. CONCLUSION

In this paper we have demonstrated that we can provide semantic indicators about content and calculate them automatically. Those indicators are built step by step. The elementary data available about learning object content are semantic metadata and domain ontology. They are used to build semantic space which is composed of semantic components. Those two notions are introduced to compute metrics that can be used to evaluate semantic properties of the learning object content. They are then combined and used to generate three significant indicators covering important aspects: complexity, continuity and imbalance of content from a semantic perspective. The two first indicators can be used too in the case of simple learning objects (not composed by reuse) if semantic metadata are provided.

Those semantic indicators are used with others indicators in the third step of our authoring approach. The goal is to assist author (i) to understand the nature and the properties of his learning object, (ii) to reveal possible weaknesses and (iii) to encourage the iterative nature of the authoring process.

The entire approach is implemented within our LOR called COLORS [6] and by using logic programming with OntoBroker [7] engine. First tests demonstrate that our approach is adequate for SCORM like learning objects and that the semantic indicators are significant and helpful to authors. However, semantic metadata are not yet supported in the current metadata learning profiles. This fact is the main restriction to enable our author assistance approach in general and especially the use of semantic indicators. To overcome this problem we are working on an automatic semantic metadata generator based on the interpretation of learning objects’ LOM metadata and textual resources.

REFERENCES