Semantic Metadata in an Engine Producing e-Learning Content

A.Seyedeh Sara Mousavi Jabbari and B.Seyed Ehsan Mousavi Jabbari

Abstract— **Metadata is a structural data defining details such as: content, format, object and structure. Within the past years, the IEEE-LOM standard has greatly captured the metadata world in e-learning tools. However, by the emergence of semantic web, e-learning tools particularly the ones issuing e-learning content have upgraded the metadata operation by adding the semantic structure to the present standards and or completely changing the structure into a semantic one. In fact, this change makes it possible to implement a new set of tools issuing e-learning content for the added value to the applied assets.**

So, the current essay aims at studying the evolution of utilized metadata in e-learning issuing tools from the standard applied programs to its semantic presentation and putting it into practice in our engine producing e-learning content.

Indexed Terms—**Semantic metadata, reusability learning objects, learning object metadata, ontology**

I. INTRODUCTION

Learning objects can be reused only if they are easily located, evaluated, adapted, and adopted by educational practitioners. In order to facilitate this process of resource description, discovery and evaluation, learning objects should be appropriately described, classified and indexed using standard metadata. Metadata are data that describe a physical or electronic resource and can be used to manage collections of documents, images, and other information in a repository such as an archive or library. Metadata are helpful because they provide standard buckets for keeping data about almost any e-learning resources. The most broadly accepted meta-data standard in e-learning is IEEE Learning Object

Metadata (LOM) [1]. With a structuralist approach, it provides 60 elements as a means of developing more Comprehensive descriptions of learning objects and of providing support for user services. The Metadata standard is also included in key IMS and ADL specifications such as Learning Resource Meta-data (LRM) [2] and Shareable Content Object Reference Model (SCORM)[3].

LO is main unit of educational content and its meaning can be integrated into e-learning environment.

For that it is necessary that LO has these features:
- Reusability
- Accessibility
- Interoperability
- Portable and
- Durability

Advantages of objects learning in e-Learning environment:
- Flexibility- same resource can be used in different context
- Administration of content – it is possible to administrate resources in a simple way, because they

- Adaptively – it enables selection and usage of resources by the needs of context
- Economy – thanks to possibilities of re-usage
- Open source- that enables compatibility with different platforms. [4].

II. EMPLOYING IEEE-LOM STANDARD IN AN ENGINE PRODUCING LEARNING CONTENT

LOM (Learning Object Metadata) is a metadata format in SCORM. LOM can be determined for each level in SCORM including: Assets, SCO, Lesson, Course and Organization or as a separated file. It can be also embedded by using a space named XML.

Totally, LOM contains the following metadata sets:
- General metadata containing topic, language, etc.
- Life cycle metadata consisting of version, situation, etc.
- Metadata information including information about metadata itself
- Technical metadata holding requirements, formats, etc
- Educational metadata embracing the conditions of utilizing the resources
- Rights metadata making up of the conditions of utilizing the resources
- Relation metadata involving information related with other educational modules
- Annotation metadata covering the definitions about educationally using the modules
- Classification metadata encompassing keywords, explanations, taxonomy, etc.

Although these metadata are able to precisely describe the resources, LOM can handle metadata standards by utilizing XML [5].
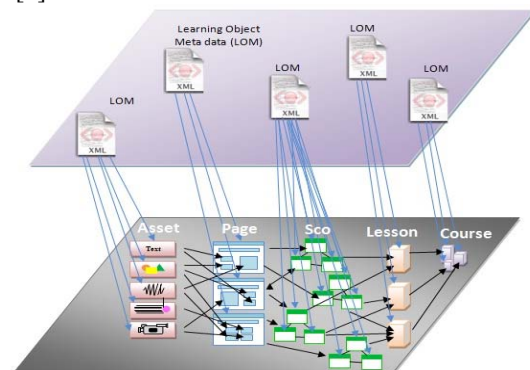

Fig. 1. The structure of e-learning content with IEEE-LOM standard

In the engine producing e-learning content that we have created, XMLs have been directed to define the main components of learning content like: Assets, Pages, SCOs, Lessons and courses by adapting appropriate Manifest. Fig. 1

displays the structural relationship between the Learning Object Metadata and the content components.

Thus, according to the structure defined in the above engine producing content, the following characteristics can be provided in the learning content:

- Making the search possible for user based on the defined learning components
- Facilitating the availability of similar data by user
- Prohibiting some users to access some information
- Training how to interpret data like: typing, coding, encrypting
- Teaching how to reuse the intended information
- Providing the users with the data background (such as the original resource or its main owner)
- Facilitating data management and control Appointing the relations between different data and other resources including dividing course into lessons, lessons to SCOs, SCOs to pages and pages to assets[6],[7].

III. SEMANTIC METADATA MANAGEMENT SYSTEM BASED ON ONTOLOGIES

Semantic metadata can be defined as "… [Linking] related terms to one another" [8]. It can be also defined as "…the process of attaching semantic descriptions to Web resources by linking them to a number of classes and properties defined in Ontology".

Different areas of computer science have different interpretations of what "semantics" mean. For instance, in the domain of databases, metadata is thought of a conceptual schema that describes the structure of a database. The domain of information retrieval, might consider metadata as the set of keywords that describe the main theme of a document, or as a record that confirms to a specific schema (e.g. Dublin Core).

Sheth et al. have described these different depictions of metadata, organizing them into three types of semantics [9]: implicit, formal and powerful. Implicit semantics appear in unstructured text that has loosely defined and less formal structure (e.g. Information Retrieval). Formal semantics appear when the data representation takes a more rigid form (e.g. Knowledge Representation). Finally, powerful semantics imply the combination of simple syntactic structures to represent the meaning of complex ones[10].

Semantic metadata based on ontologies are applications that rely completely on domain ontologiesto define their metadata. They use RDF (Resource Description Framework) as a vehicle to express the semantics of a learning resource. For these applications, using RDF has advantages over the standard metadata approach .

First, any RDF data model is based on an "open world assumption" where the metadata is selected from heterogeneous ontologies. On the other hand, the majority of systems that implement LOM take a closed-world approach confining their metadata to that implemented by a particular LMS. Second, RDF allows for the creation of complex statements (i.e. metadata can be further annotated with more metadata). LOM, on the other hand, does not allow for the expression of complex metadata; it only supports extensions through taxonomic classification[11].

Therefore, in the engine producing content created by us,

we can exercise a format embracing an XML structure and apply IEEE-LOM to express assets, pages, SCOs, lessons and courses relying on their attributes and the process of practicing them in learning path and their inner relationships with each other. As a result, we have considered an ontology that it can supply us with such an exact, formulated and clear discrimination of the learning content. Initially, it is necessary to contemplate the content structure in two levels to reach a suitable ontology throughout the considered RDF:

- Upper Level: set of concepts related to the course topics, selected among the concepts defined in the domain ontology. (Relationship made by Semantic Metadata)
- Lower Level: learning resources (assets, pages, SCOs, Lessons, Course, etc) associated to the concepts. (With the learning resources' inner relationship)

Fig. 2 represents the relation on the defined level to create a convenient ontology in content tending to a learning path.
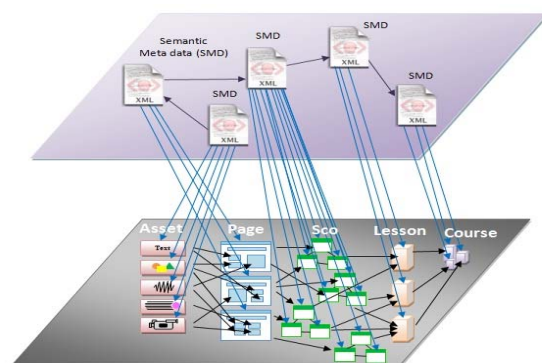

Fig. 2. The structure of e-learning content with semantic metadata based on ontology

The content with the properties of semantic metadata based on ontology can create learning paths and these learning paths add the characteristics below to the structure of e-learning content:

- A sequence of concepts selected during the ontology browsing. This sequence gives the order of access to the learning objects.
- A corresponding sequence of learning-objects, associated to the ontology concepts, that compose the customized course (by making relationships among semantic metadata)
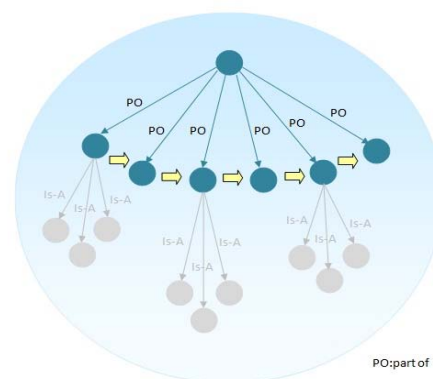

Fig. 3. Horizontal dimension in learning paths.

Learning paths can figure out the relationships between

learning components and concepts through two distinct dimensions of horizontal and vertical. In horizontal dimension, learning sequence is determined due to main topic (starting from a giving concept) and browsing the ontology (through the decomposition relations). As shown in fig. 3, main topic with other components can make a PO (Part Of) relationship.

Vertical dimension in learning paths follow a set of specialization links holding both up and down views in which:

- Up: Synthesis and wrap-up of the topics
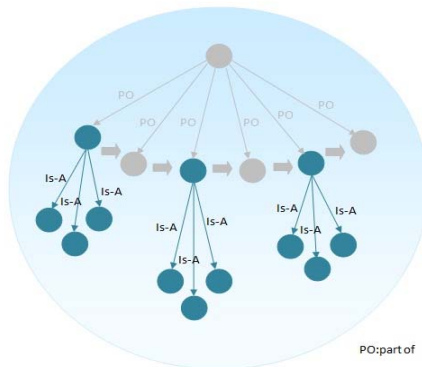- Down: Further elaboration of a given topic

Fig. 4. Vertical dimension in learning paths

Fig. 4 illustrates the vertical dimension in learning paths to elaborate the topics with an IS-A relation. Finally, we introduced the ontology model reclining on extending the application of IEEE-LOM to semantic metadata based on ontology (case study).

IV. ONTOLOGIES AND SEMANTIC METADATA FOR CONTENT

Both the terms metadata and ontology have variable interpretations depending upon circumstances in which they are used.

Metadata means information about information and it can be used for different purposes such as to describe media characteristics, content processing, and actual content semantics (Boll et al. 1998).
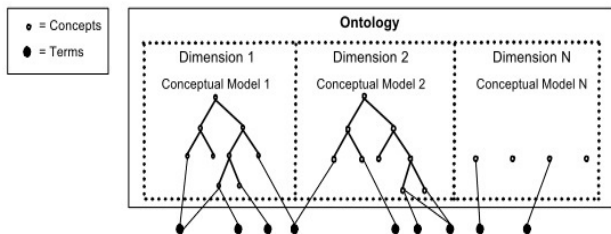

Fig. 5. Different Aspects of a Domain Ontology

With the definition of the word ontology we stay away from the more complicated definitions used for example in knowledge acquisition and representation (Shapiro 1992). In our paper ontology means a set of formally specified metadata structures consisting of commonly agreed concepts that bear a limited sense of meaning within them. Ontology describes the semantics and can cover multiple different angles, dimensions, of the content. With these dimensions the ontology should be able to cover the semantic needs that are needed to produce and deliver the content to the customer. The following fig. 5 visualizes the relations between different aspects of our ontology.

Ontology comprises a set of concepts and concept relationships representative to the problem domain. Concepts and their relations define *conceptual models* for classifying (or "tagging") information objects under different *dimensions*. Dimensions are typically independent from each other and they have their own conceptual models. We have preferred hierarchies in the conceptual models due to their computational and representation advantages, but this is not a mandatory requirement. An example of dimensions is Location, Subject, or Author of a document. Subject and Location -dimensions have well-defined hierarchical conceptual models, whereas Author typically contains only the name of the author without deeper conceptualization.

Boll et al. (Boll et al. 1998) divides metadata into content-independent and dependent classes. Content-dependent metadata can be further divided into direct content-based and content-descriptive metadata having both domain-independent and domain specific subclasses. This classification captures well metadata dependency on its environment, but it does not describe how the metadata can be used. We acknowledge the dependency aspects, but use a more process-oriented approach by classifying metadata into three categories: semantic, structural, and control metadata. Even though all these three aspects are important, this paper concentrates mostly on discussing different aspects of semantic metadata.

Semantic metadata complies with its ontology and describes the meaning of the content. Semantic metadata must be produced in a common systematic format and the production must follow agreed guidelines. If the content provider does not provide semantic metadata in a systematic way or if a common agreement for ontologies is missing, the results are not compatible. The keywords of a news story are an example of semantic metadata.

Structural metadata describes the format and technologies used with the content, but does not consider what the content actually means. An example of structural metadata is the time encoding in a video clip.

Control metadata contains production and delivery related metadata. This metadata assists in determining the status of the content and rights to access and use the content.

When producing semantic metadata and developing content models, the resulting ontology must be able to describe both the incoming information feeds and the needs of the customers. If either one is ignored, the ontology is seriously impaired. In addition, the ontology has to be able to cope with the dynamic nature of the information feeds and customer interests. This means, that there must exist methods for reflecting changes back to the ontology.

V. JUSTIFICATION FOR SEMANTIC METADATA

Metadata production and utilization require resources and effort. There are, however, a number of reasons why it is advisable to produce and use semantic metadata.

• *Size*. Semantic metadata is a condensed representation of the content. It captures the essential semantics of the source. With textual content the savings depend on the detail level of the produced semantic metadata. With other source formats such as video or audio the savings are, however, much more substantial. Instead of using megabytes of original content, many tasks related to the content production and distribution can be performed with a much smaller semantic metadata representation.

• *Support for multiple formats*. Metadata can support many different formats with the same representation format. If we consider the difficulties borne with the management of multiple different formats such as text, images, video, or audio, we can easily understand the advantages of having semantics represented in a single uniform way.

• *Expressing hidden/author's views*. This can be considered to be either an advantage or a disadvantage. On one hand semantic metadata forces the author to express the message explicitly in the semantic metadata. This helps to define the key facts making the content clearer. On the other hand semantic metadata reveals the hidden message author wanted to express without stating it explicitly.

• *Common view on machine-usable level*. Because key information is stored explicitly in the metadata, there is less room for different interpretations what the content is about.

• *Saves computation*. Content analysis especially with the video and audio formats requires often extensive computing. If we have analyzed the content during its creation, we do not have to perform the analysis later on the fly.

• *Higher information quality*. High quality metadata increases the possibilities to produce and deliver accurate information to the customers.

On the other side of the coin are at least the following questions

• *Ontology creation*. One should not underestimate the difficulties related to building a good ontology. The creation of an initial ontology requires a special set of skills as well as expertise on both the domain and the customers' needs. Moreover, ontology may require multiple iterations before it is usable.

• *Expressiveness of the ontology*. It is naïve to claim that textual metadata is able to express all possible aspects of the content. Emotions and subjectivity are two examples of the broad range of difficult challenges in this field.

• *Effort required in the metadata creation*. Metadata creation can be very expensive in terms of human resource consumption. It might also be very difficult to pinpoint the exact value the metadata is producing.

• *Dynamic nature of metadata structures*. Metadata structures change over time, which produces a number of questions related to managing already existing metadata.

• *Degradation of the information*. Not only the structures change, but the overall correctness and value of information changes over time. We must be able to produce and manage multiple versions of both the metadata and the content.

• *Different qualities of different media content types*. If we try to use the same the metadata format for different content types, we have to understand and produce metadata to cover all the unique aspects of these types. For example, image metadata does not have to take into account temporal relations, whereas for metadata describing a video clip this information is essential.

Most of these questions will be discussed later in the paper when we introduce our ontology development framework. Some of the metadata advantages appear during the authoring, some later during the delivery or consumption of the content. It seems, that the more complex the content production and manipulation is, the more advantageous it is to use metadata in the process. For example, content personalization benefits from semantic metadata that can be used in selecting actual content to be delivered to the customer's site. This approach decreases network traffic and is more flexible than transferring and processing huge amounts of original content.

VI. NATURE OF ONTOLOGIES FOR CONTENT

Librarians have worked for centuries to find usable ways of describing and categorizing information. Automatic information filtering systems, such as SmartPush (SmartPush 1999), try to route information by matching explicitly defined content metadata with customer profiles that describe user needs. The specification of these needs is based on a well-defined domain model.

An ontology can be more than a taxonomy or classification, and can include multiple types of relationships between concepts. One possibility for a rich semantic model that has been applied to the news domain is the approach used in Perspecta (Perspecta 1999) [fig. 6].
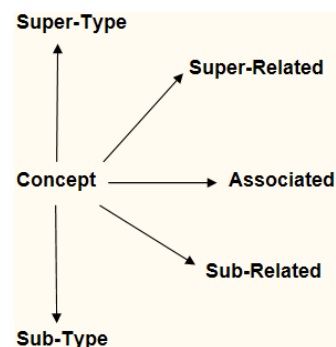

Fig. 6. A Semantic Model of Concept Relationships

• Super/sub-type concepts are ontological parents/children of the target concept; they are of the same type as the target concept, but more general or more specific. For example, software company is a sub-type of company.

• Super/sub-related concepts are more general/specific than the target concept, but not of the same type. Software company products is a more specific concept than software company, but not a sub-type of it. Thus, software company products is sub-related to software company.

• Associated concepts have a "see also" relationship to the target concept. These rich relationship types enable the expression of deeper domain-specific knowledge. Both super/sub-type and super/sub-related relationships can be used for describing hierarchical ontologies.

The problem with news is that the domain is open and unbounded. Comprehensive computer-based representation

of an ontology that covers whole the news domain is impossible to create. A typical news producer organization has access to information feeds, such as wire services from Reuters and AP. These news feeds have their own classifications of content that we refer to as information feed ontologies. These feeds are typically treated as raw information for more thorough news reporting, and even if the content from information feeds would be used as such, they typically require re-categorizing.

Therefore a news ontology needs to be based on journalistic judgement by the content provider, and we define this ontology as the provider ontology. The content provider tries to cover its content domain in a way that would be most useful to its customers. This provider ontology can be explicitly available for customers for making selections for their information needs, and typically serves as a basis for customization for different individuals and communities of news consumers. The customers may also have defined their own ontology, a customer ontology, which has to be linked to the provider ontology. These ontologies and their relations are described in the following fig. 7.
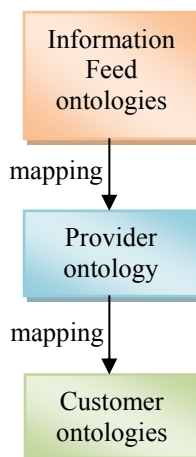

Fig. 7. Mappings between ontologies

Ontology mapping means defining comparable and relating concepts to facilitate the usage of content over heterogeneous data sources. If mapping is not performed, the semantic metadata is not compatible. An example of the usefulness of ontology mappings can be derived from agent based systems, where the adoption of a shared ontology through mappings allows agents to interoperate without misunderstanding, and to retain a high degree of autonomy and flexibility.

XI. CONCLUSION

Ontologies can organize learning objects in contents and maintain a semantic accessibility for them and this directs instructors to provide learning process and assessment based on processes by creating learning paths in making content and to receive the knowledge they need. Learners creatively carry on learning paths by having a suitable learning content. For this approach, semantic queries of ontology are required to be arranged for improving the reusability of the learning process. Moreover, the structural techniques of learning paths are verified in accordance with their limitations and prerequisites. This is possible when a unified structure be defined by utilizing metadata standards like IEEE-LOM for learning object. Consequently, essential capabilities are equipped for making the content intelligent through expanding it toward semantic metadata based on a convenient ontology model.

REFERENCES

[1] IEEE, "1484.12.1-2002 Standard for Learning Object Metadata," http://standards.ieee.org/catalog/olis/learning.html.

[2] IMS Global Learning Consortium, "Learning Resource Metadata," http://www.imsglobal.org/metadata/.

[3] ADL, "Sharable Content Object Reference Model (SCORM) 2004 2nd Edition," http://www.adlnet.org/downloads/70.cfm

[4] O. Ghebghoub, M.-H. Abel, and C. Moulin, "Learning Object Indexing Tool Based on a LOM Ontology," *Proceedings of the Eighth IEEE International Conference on Advanced Learning Technologies (ICALT '08)*, pp. 576-78, 2008.

[5] Duval, E., IEEE standard for Learning Object Metadata(LOM). Accessed on May 8, 2006. Available on linehttp://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

[6] Hossein , Keynejad .E-Learning Content authoring tools and introducing a Standard Content Constructor Engine.2011 ,2nd Symposium on Computing in Science & Engineering. iscse 2011 ,Kusadasi , izmir, Turkey.   June 1-4 , 2011.

[7] Hossein, Keynejad. Introducing an Intelligent E-learning Content Constructor Engine. *2011,3nd International conference on education and management technology, ICEMT 2011* , Shanghai , china , August 19-21,2011.

[8] Haase, K. Context for semantic metadata. *Proceedings of the 12th ACM International Conference on Multimedia*. 2004. New York, NY, United States: ACM.

[9] Scerri, S., C. Abela, and M. Montebello. SemantExplorer: A Semantic Web Browser. In IADIS International Conference WWW/Internet 2005. 2005. Lisbon, Portugal.

[10] Sheth, A., C. Ramakrishnan, and C. Thomas, Semantics for the Semantic Web: The Implicit, the Formal and the Powerful. *International Journal on Semantic Web & Information Systems*, 2005. 1(1): p. 1-18.

[11] Brooks, C. and G. McCalla, Towards flexible learning object metadata. *International Journal of Continuing Engineering Education and Life-Long Learning*, 2006. 16(No.1/2): p. pp.50 - 63.