



# The automatic generation of semantic annotations for didactic materials and the use of models in the material development to improve its reusability

Carmen L. Padrón, Paloma Díaz, Ignacio Aedo

## ► To cite this version:

Carmen L. Padrón, Paloma Díaz, Ignacio Aedo. The automatic generation of semantic annotations for didactic materials and the use of models in the material development to improve its reusability. 8th IEEE International Conference on Advanced Learning Technologies ICALT08. Learning technologies in the Information society, 2008, Santander, Cantabria, Spain. pp.855-860. hal-00703040

**HAL Id: hal-00703040**

**<https://hal.science/hal-00703040>**

Submitted on 31 May 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The automatic generation of semantic annotations for didactic materials and the use of models in the material development to improve its reusability

Carmen L. Padrón, Paloma Díaz and Ignacio Aedo  
DEI Laboratory. Computer Science Department  
Universidad Carlos III de Madrid, Spain  
[clpadron](mailto:clpadron@inf.uc3m.es), [pdp](mailto:pdp@inf.uc3m.es), [aedo](mailto:aedo@ia.inf.uc3m.es)

## Abstract

*The didactic material reusability, understood as its capability to be used in different learning situations or in diverse knowledge domains, is one of its most desirable features. It depends, among other factors, on the use of semantic annotations to describe relevant material characteristics, which can improve its localization and retrieval. In spite of the recognized value of semantic annotations, they are usually considered as optional. In this paper we describe the mechanism proposed in the MD2 approach to automatically generate semantic annotations within the didactic material development. Such mechanism tries to relieve users (developers) from the complexity of material's annotation and intends to obtain materials with some reasonable degree of reusability.*

## 1. Introduction

The use of learning technologies promises new opportunities for instructional reusability and personalized learning since they provide some means to describe main characteristics of materials, and infrastructure supports that help to localize, retrieve and reuse those materials in different educational situations or adapt them to the different educational preferences of learners. Instructional reusability and personalized learning are two basic features needed to face the challenges of the current information society. Bringing to reality such promises demand that those features will be somehow ensured within the development process of didactic materials.

The didactic materials reutilization capability depends, among other factors, on applying a material's design that is based on separation of the contents from its presentation and, the contents from its instructional design [11]. These separations help to severally reuse or adapt material or its components to different learning situations or in diverse knowledge domains. Moreover, the material should comply with the current learning technology specifications and standards, specially with those related to interoperability that allow materials to be used in heterogeneous contexts and those concerned to meta-data annotations which allows the

automatic or semi-automatic material search and retrieval. Thus, during the didactic material development those factors must be taken into account in order to achieve a reasonable level of reusability. But even if materials maintain a clear separation among contents, instructional design and presentation and are standard compliant, they can be hard to reuse. The main problem is that practitioners who intend to reuse them also need mechanisms to identify the material that fulfil their expectations, for which some kind of meta-data can be useful [9].

On the one hand, the didactic material has some pedagogical properties that can be relevant for reuse purposes such as its pedagogical strategy type, learning objectives supported, and time needed to achieve those objectives, level of difficulty, kind of material or its granularity. On the other hand, there is pedagogical information about the knowledge domain which is also important for reusing materials like the name of the topic or theme enclosed in its contents, the kind of learning competences that topic can cover, its knowledge prerequisites, the relation of that topic with other topics on the same knowledge domain, etc. And last but not least, some technical support features of material can be also essential for its reuse such as the medium used to publish or deliver it, its format, presentation time or the type of learner's interaction it promotes. A material which has been described using all those 3 perspectives, can have more chances to be discovered and adapted or reused when a learner or practitioner requires another material which has in common some of its described requirements or features. Unfortunately, most popular approaches to develop didactic material development like RELOAD [12], LAMS [8] and CopperAuthor [1] consider the semantic annotation as optional and it must be carried out by experts of each of the aforementioned perspectives. In this paper we introduce a mechanism to deal with these issues by automatically generating semantic annotations from the models generated during the development process. In this way, a set of semantic annotations compliant to standards are obtained, increasing the potential of material reusability and relieving developers of the complexity of creating them manually. In the section 2 we briefly present the MD2 approach, which supports some of e-Learning standards and specifications like IMS LD [5] and IMS LRMI (IEEE LOM)

[7] and proposes a model, a method and architecture for the didactic material development. Section 3 is devoted to present some related work and to explain how semantic annotations are generated from the models in the MD2 approach. Finally in section 4 some conclusions and future works are outlined.

## 2. The MD2 approach

Didactic materials are herein considered as a conjunction of a set of contents, in the form of learning objects, and a pedagogical strategy, defined by an instructional design, and they take the form of units of learning to guide the educational process. This definition takes into account current e-Learning standards and specifications and it is based in two important concepts in the e-Learning domain: the learning objects and the units of learning [10].

MD2 is an approach for developing didactic material whose rationale is “the development can be effectively supported if we count on means to describe the main features of the material (namely, contents and pedagogical strategy) taking into consideration the real abilities of material developers, who are not always and at the same time experts in pedagogy, in the knowledge domain and in technical issues”. The approach consists of a method, a model and architecture for their implementation.

The method intends to assist users during the different stages of didactic development: selection, composition and evaluation [10]. Underlying the method, there is a model that provides a number of elements to formally specify the features of the didactic material using four views: knowledge domain (KD), pedagogical information (P), technical support (S) and quality (Q-U). The main idea behind this approach is to use this model elements or descriptors to provide automatic mappings from complex technical descriptions of learning technology standards and specifications to simpler and closer descriptions to practitioners about material requirements. Thus, the relationships amongst the different MD2 model elements from K, P and S views are used by the method to guide the selection and composition stages of the development. Meanwhile another set of MD2 model elements (Q-U) are used in the evaluation stage to help to analyze the material’s usability, pedagogical value and quality. Those elements had been defined taking into consideration previous experiences related to evaluation of Educational Hypermedia [5].

The method also defines a mechanism to automatically annotate material features taking advantage of the relations among elements of different MD2 model views and the descriptions they provide about material characteristics. Hence all values of MD2 model elements with the material’s development rationales will be automatically stored as extended semantic annotations of the material, allowing its future localization and its retrieval for reutilization purposes.

Next section goes further into automatic generation of semantic annotations presented in this paper.

## 3. Automatic generation of semantic annotations

To the best of our knowledge there are no design or development methods which deal with the automatic generation of semantic annotations in the learning design arena, but there are authoring tools like HyCo [2] which semi-automatically generates annotations using some elements of IMS LRMI (IEEE LOM). It automatically generates the values for some IMS LRMI elements according to information from the structure of created hypermedia book and it provides support to manually create values for the rest of IEEE LOM elements relying on the user knowledge. Also approaches from the software engineering discipline like KA2 initiative [4] are worthy of consideration since it automatically produces semantic annotations using a process of extracting annotation-relevant information from texts and multimedia resources.

The MD2 method includes a mechanism that generates semantic annotations, which is automatically triggered when the evaluation stage ends. This mechanism creates a set of extended semantic annotations. We consider them extended because they comprise descriptions of material features that comply with IMS LRMI (IEEE LOM) [7] as well as annotations concerning development rationales. Those development rationales include information about the set of actions taken during the different stages of development, their results and the reasons leading to take such actions. Our approach is different from the used in HyCo since MD2 supports different types of materials (learning objects, assessment and units of learning), automatically generates two different types of annotations and the descriptions about material features are obtained from the relations among the different MD2 views meanwhile HyCo uses the structure of hypermedia book that can be represented as learning object to generate the annotations. Next we explain each of these types of annotation and their influence on the potential reusability of material.

### 3.1 Automatic generation of LRMI semantic annotations

The descriptions of material features compliant to IMS LRMI (IEEE LOM) specification covered by the MD2 approach include a subset of elements grouped on four of the nine categories defined by the specification: General, Technical, Classification and Educational [7]. These categories are closely related to the three perspectives presented in the Introduction section. For their selection we also took into account which ones of the 100 elements defined by the specification were the most common used features to search and retrieve materials [7]. In table 1, the *S\_standarElements* descriptor is presented and it includes mappings from elements of the different MD2 model views to the LOM descriptors of the aforementioned categories. The availability of such standardized descriptions about material features can help to search, localize and potentially reuse or adapt material or its components.

Table 1. Mappings from MD2 elements to selected LOM descriptors

LOM descriptors included in the MD2 element <i>S standarElements</i>		MD2 elements
Category	Element	
General	title	<i>KD_topic</i> , <i>P_materialType</i> , developer, affiliation
	aggregation-level	<i>P_materialType</i>
	language	<i>S_language</i>
Technical	format	<i>S_medium</i>
	requirement-operating-system	<i>S_medium</i>
	requirement-browser	<i>S_medium</i>
	duration	<i>P_effort</i>
Classification	classification-purpose	<i>KD_discipline</i> , <i>P_skills</i>
	classification-taxopath-source	<i>KD_discipline</i> , <i>KD</i> ontology name
	classification-taxopath-taxon	<i>P_skills</i>
Educational	learning resource type	<i>P_materialType</i>
	interactivity level	<i>S_interaction</i>
	difficulty	<i>P_difficulty</i>
	typical learning time	<i>KD_estimatedTime</i>

### 3.2 Automatic generation of annotations about development rationales

MD2 generates another kind of annotations, which are concerned to development rationales and are extracted from different stages of the development process.

Thus, during the selection stage the semantic annotations generated include the ideas that support selection decisions like the number of content resources needed, and per each one of them: the type of selection process (retrieval, creation or versioning); the kind of resources included on contents; the type of pedagogical strategy selected; the durability of the topic and right management issues.

The rationales related to the composition stage comprise whether the stated requirements have influences to modify the material presentation structure, which is an instance of unit of learning for the selected pedagogical strategy and it is represented according to IMS LD XML binding. The rationales also include what kind of adjustments was made to such structure and the results of content integration in that presentation structure.

Meanwhile rationales concerned to the evaluation stage include all the values of Q-U view elements which represent evaluation criteria [10], the computed and observed values for each evaluation objective (pedagogical value and usability), whether redesign is needed and in those cases, the recommendation proposed. Also when redesign is needed, the annotations include information about corrective actions

and the results of modifications done in each stage of development according to the proposed recommendation.

The availability of such kind of annotations can be useful to guide and reduce development efforts in those situations when materials with similar development requirements are needed.

### 3.3 The mechanism for the automatic generation of annotations

A general architecture of the MD2 development framework has been presented in [10]. It includes a semantic annotations generation mechanism which interacts with the rest of mechanisms supporting the different development stages. The semantic annotations generation mechanism implements an annotation algorithm composed by the following steps

- I. Gather semantic information to describe
  - a. Material features
  - b. Development rationales
- II. Generate semantic annotation files
- III. Store didactic material with its annotations in the Repository.

In step Ia. (Gather Material Features), all information that has been gathered during the selection and composition stages concerning the knowledge domain (KD), pedagogical information (P) and technical support (S) views provides data to describe material features according to IMS LRMI (IEEE LOM). The relationships between the elements of these three MD2 views (KD, P and S), play an important role to achieve such a goal since they establish which information must be used in each of the selected LOM descriptors, as has been shown in Table 1.

The step Ib. (Gathering development rationales) aims to log all data from the executions of modules supporting the different development stages and the reasons why certain action has been taken. This step uses the data flows from the MD2 selection, composition and evaluation mechanisms to the annotation mechanism. The development rationales include the name of each development stage, the MD2 model elements used in the stage, the results obtained in the development stage and the decisions taken through the process to obtain them. They also comprise information about the material redesign when it is needed, the results of proposed recommendations and the modified values of related MD2 elements.

The step II (Generate semantic annotation files) uses the data gathered in the previous steps to store them in 2 different annotation files. For the general material descriptions is created a *lom.xml* file using the IMS LRMDI XML schema meanwhile for the development rationales annotation file *developmentrationales.xml* is used the XML schema defined by the MD2 method. Thanks to this step the mechanism is able to provide developers with the semantic annotations.

Once the distribution package for the material is created according to the IMS CP [5], step III is in charge of adding the *lom.xml* file to the package and stores the package material with its development rationales file into the

Repository. By this way, when a developer has to create a new material with similar requirements, the material previously created can be retrieved from the Repository during the selection stage using information from the semantic annotations. According to the degree of the coincidence found respect to the descriptors included in the *lom.xml* file, one of its components (content or strategy) or both could be reused in the new material. Furthermore, annotations concerning development rationales annotations can be used to guide the new development and also can be a good start point to automatically generate materials.

### 3.4 An example of automatic generation of semantic annotations

Let us introduce an example to help to understand how semantic annotations are generated within the MD2 approach. In this example, all evaluation stage details have been omitted, aiming to make the example simpler. If the reader is interested in a complete example, it is presented in [10]. There is a development situation where a teacher from a Spanish university needs to create a material to guide a learning process that is related to Fundamental Data Structures, especially about the topic Linked List Data Structure. He will use an authoring tool which implements the MD2 approach and there is also available a Repository. Step 1 of MD2 method asks him to select the topic to be covered by the material and his answer is the aforementioned topic. According to the method this information is used to consult a knowledge domain Ontology about Computer Science learning and to obtain data to complete all values of KD view like discipline (*KD\_discipline- Programming Fundamentals*), general learning objectives addressed by the selected topic (*KD\_generalObjectives*) and the estimated time to reach them (*KD\_requiredTime*) or the topics related to the selected one (*KD\_sTopic*). Then the method Steps from 2 to 8 ask teacher about some pedagogical features of the material. Step 2 ask to select a type of material from a list of possible types supported (i.e. unit of learning -UoL, learning object, and assessment) and he choose **UoL** that constitutes the value for the *P\_typeMaterial* element. At this point the teacher also is asked to select a type of pedagogical strategy from the list of the most appropriated strategies for learning processes related to Computer Science. The values of that list are defined according to the definition of *P\_strategy* element, he selects **Problem based strategy** and it is stored as the value for *P\_strategy*. Step 3 demands him to select which are the required concepts and learning objectives to attain from the list of cognitive skills to be addressed by the selected topic, which has been assigned to *KD\_generalObjectives* in Step 1. Some of the data from that list are (1) **Knowledge- Know- Linked list data structures representation, operations and use**; (2) Comprehension-Describe-Linked list data structure memory allocation. The teacher selects the first one and those values will be stored into the *P\_skills* element. Step 4 ask teacher to define the estimated effort to execute material and to effectively achieve the learning objectives from a list of legal time values restricted by the 6hrs defined in *KD\_requiredTime* element.

The teacher selects **2 hrs** and that value is assigned to *P\_effort* element. Step 5 asks teacher to define the kind of items needed for the material. To do so, a list of possible types of items or components is presented. Its values are proposed following the *P\_typeMaterial* definition and they can be whole strategy, some activities or some activity structures for that strategy. The teacher selects **some activities structures**. Step 6 asks teacher about the number items he considers needed for the material and her answer is **3**. Step 7 requires from teacher to supply information about level of difficulty for the material. Thus he must select an option from the list of possible fuzzy values [High, Medium, and Low]. His answer is **Medium** and this value is assigned to the *P\_difficulty* element. Next, Step 8 asks to the teacher to define if some score threshold is needed to consider that learning goals are achieved. Then he must select an option from a list of two valid values (Yes, No). He chooses Yes and he is prompted to provide a range of values for the score threshold and he gives a range between 2 and 5 points. This information is used to obtain suitable value for S view elements related to the e-Learning specification used for material delivery or presentation.

Method Steps from 9 to 11 ask teacher to provide some information about the technical support needed for the material. Thus, step 9 asks teacher to define if the material needs time presentation restrictions, he is asked to select an option from a list of two valid values (Yes, No). He selects Yes and he is prompted to provide a range of values to restrict the material (UoL) execution time. His answer is also controlled by a maximum value corresponding to the to the Step 4 answer. In this case he chose **2 hrs** and this value is used by the composition mechanism to control the UoL execution time. Next in Step 10 teacher must select the desired supporting medium for the material from a list of possible media (Printed, hypertext, hypermedia-web, pervasive computing element) defined by the MD2 meta-model for the element *S\_medium* and he also is asked to select the language of presentation. For the first question, he selects hypermedia-web and this value is assigned to the mentioned S view element. Such value is used to determine which will be the e-Learning standard or specification XML binding used to define the presentation structure, in this example is IMS LD. Also according to the *S\_medium* definition, selected value will provide information about the format of presentation (**XHTML**) and the requirements for the operating system (**any operating system**) and browser needed for the material visualization (**any browser**). The Step 11 asks teacher to define what kind of learner interactivity with the material is needed to effectively achieve the stated learning outcomes. He is prompted to select of one type from the list of fuzzy values [High, Medium and Low]. He selects **Medium** and that answer is assigned to *S\_interaction* element. Its also used by the MD2 inference mechanisms based on S view definition to the adequacy of his selection with respect to his answer to Step 10 and to confirm the proposal about the presentation and delivery standard. For the second question about the language, he must select a value from the list of possible presentation

languages defined for the element *S\_language*. He selects **Spanish (es)**.

All the information gathered in these 11 Steps is used by selection mechanism to select the appropriate contents for the material contents. In this example, was not found a *developmentrationales.xml* file for materials of the same type the teacher needed, thus was not previous developing experience obtained to guide this development. Thus during selection, 6 different contents (4 explanations and 2 simulations in form of learning objects) were retrieved from the Repository since their features matched to the gathered requirements. All of them were opensource and free to use thus, no rights management issues were found. And the teacher decided to make versions of 2 explanations related to Linked list operations to adequate them to desired difficult level (*P\_difficult* value). Also the data obtained from the 11 steps were used by the composition mechanisms to select the presentation structure i.e. a UoL template for Problem based strategy with 2 hrs of presentation time and with 3 activity structures, the first structure was devoted to Linked list representations, the second one to Linked list operations and the last one for Link list uses. Such structure does not need any modification. As result of composition the different contents were glued to the presentation structure to make up a coherent material. Then, when evaluation stage ends, the semantic annotation generation mechanism is triggered. Thanks to the communication among previous mechanisms and the annotation mechanism the information to execute the annotation algorithm is supplied. Hence the step Ia of annotation algorithm can obtain and assign the values of the selected LOM descriptors according to the Table 1. These values have been presented in this example with a bold type. Also the step Ib can gather the data from the aforementioned development rationales. Next the step II generates the *lom.xml* and *developmentrationales.xml* files and the step III adds the first one to the material distribution package and stores the package with the development rationales file in the Repository. Thus when a new material about Linked list with other similar requirements will be needed, it will be possible to reuse the recently created material about Linked lists.

## 4. Conclusions and Future works

In this paper we have presented the mechanism proposed by the MD2 approach to automatically generate extended semantic annotations while you are developing didactic materials. It is novel in the sense that automatic generation of semantic annotation includes standardized metadata derived from knowledge domain, pedagogical and technical descriptions for material development and also provides annotations about development rationales. Thanks to this mechanism the MD2 approach can help to ensure some degree of reusability of the obtained material and lightens developers from the complexity of the semantic annotation process.

The implementation of presented mechanism has been concluded and we are currently working on its assessment in real development situations. We already obtained some preliminary information that can confirm that presented

mechanism lightens developers from the cognitive load of materials annotation. Further results of these evaluations will give us more information about previous conclusion and related to the usefulness of the automatic generated semantic annotations to achieve the desired and demanded instructional reusability.

**Acknowledgements:** This paper is part of the MODUWEB project (TIN2006-09678), funded by the Ministry of Science and Technology, Spain

## 5. References

- [1] CopperAuthor project website. Retrieved December 3, 2007 from <http://www.copperauthor.org/>
- [2] Berlanga, A. and Garcia, F. (2004) Towards Semantic Metadata for Learning Elements in Proceedings of the Fifth International Conference on Information Technology Based Higher Education and Training, 2004. ITHET 2004. pp 572- 577
- [3] Díaz, P. (2003). Usability of hypermedia educational e-books. D-Lib Magazine. 9(3). Available at <http://www.dlib.org/dlib/march03/diaz/03diaz.html>
- [4] Erdmann, M., Maedche, A., Schnurr, H.-P., Staab, S. (2000). From Manual to Semi-automatic Semantic Annotation: About Ontology-based Text Annotation Tools available at [http://www.aifb.uni-karlsruhe.de/WBS/Publ/2000/COLING\\_meretal\\_2000.pdf](http://www.aifb.uni-karlsruhe.de/WBS/Publ/2000/COLING_meretal_2000.pdf)
- [5] IMS-CPIM. IMS Content Packaging Information Model. Version 1.1.2 final specification, 2001. [http://www.imsproject.org/content/packaging/cpv1p1p2/imscep\\_infov1p1p2.html](http://www.imsproject.org/content/packaging/cpv1p1p2/imscep_infov1p1p2.html)
- [6] IMS-LD. IMS Learning Design. Version 1.0 - final specification. Retrieved April, 2005 from <http://www.imsglobal.org/learningdesign/index.html>
- [7] IMS Learning Resource Meta-data Specification. Retrieved November, 2007 from IMS global website [http://www.imsglobal.org/metadata/mdv1p3pd/imsmd\\_bestv1p3pd.html](http://www.imsglobal.org/metadata/mdv1p3pd/imsmd_bestv1p3pd.html)
- [8] Learning Activity Management System (LAMS) foundation website. Retrieved December 28, 2007 from <http://lamsfoundation.org/>
- [9] Littlejohn, A. Issues in reusing online resources in Reusing Online Resources: A sustainable approach to e-learning. Editor Littlejohn A. Kogan Page London/ Kogan Page US, 2003.
- [10] Padrón, C. L., Díaz, P. and Aedo, I. (2006) MD2 Method: The Didactic Materials Creation from a Model Based Perspective in Innovative Approaches for Learning and Knowledge Sharing. W. Nejdl and K. Tochtermann (Eds.). LNCS 4227, pp. 366 –382, 2006.
- [11] Polsani, P. "Use and abuse of reusable learning objects." Journal of Digital Information, Volume 3, Issue 4. (2003). Retrieved January, 2007 from <http://jodi.ecs.soton.ac.uk/Articles/v03/i04/Polsani/>
- [12] RELOAD Reusable E-Learning Object Authoring and Delivery project Retrieved December 5, 2006 from <http://www.reload.ac.uk/>