Semantic Metadata for Creating and Describing Educational Resources

Brik Mourad
Computer Science Department Ferhat Abbas University
Setif Algeria

Touahria Mohamed
Computer Science Department Ferhat Abbas University
Setif Algeria

ABSTRACT
Nowadays, web contains a multitude of information sources and knowledge which often are represented as HTML files. These files can be used as educational resources (ERs) or may contain inside (piece of text, image …) important ERs. In this paper we propose an approach that allows us to extract and annotate these resources in order to store them into Learning Object Repository (LOR). The users can search ERs through this LOR from different points of view, especially educational type and domain topic. This LOR follows an adaptive research according to the user preferences.

Keywords
Metadata, Metadata Application Profile, Semantic Web, Ontology, Learning Object Repository.

1. INTRODUCTION
E-learning refers to the use of information and communication technology to diffuse learning resources, a learning resource is "any digital resource that can be reused to support learning" [1]. In this context, we distinguish three approaches that allow the creation of new educational resources:

    a) Writing;

    b) Combine multiple resources to form another;

    c) Extract it from another electronic resource.

The build of educational resources is the basis of e-Education which enhances the productivity in this domain However, The growing number of educational resources available leads to problems of access, management and sharing as well.

The current e-learning systems are based on metadata to annotate their learning objects. The use of metadata can significantly increase the effective description, search and retrieval of educational resource, but with current metadata, it seems clearly that the learning Object contents are inaccessible for a software agent. Concretely, the ERs have no knowledge neither about the taught domain nor educational field. Then, to receive the best search results the used keywords should be given from ontology. Gruber defines the ontology as 'the specification of a conceptualization' [2] furthermore, the use of the ontology brings a machine comprehensible data.

Our work consists in extracting a portion of text from an electronic document (html format), this result fragment is stand for an educational type (judged by the expert of the taught domain), and to annotate it afterward with a set of metadata in order to store them into a LOR. For this purpose we are developing a tool that inputs an HTML document (supposed that it contains one or many educational resources) and outputs an annotated ER of lower granularity with a semantic metadata schema. We focused in this work for the most important metadata that describing the educational resources and its contents to facilitate the usage of these last ones, in total, our proposition for a metadata schema is a LOM[1] (Learning Object Metadata) application profile.

Our approach to annotate the ERs components based on three ontologies *(domain, pedagogy, user preference),* furthermore the user preference ontology is used for personalized the research and the results. The approach is tested for the domain of "*computer programming.*"

2. ARCHITECTURE PROPOSED
Figure 1 illustrates our tool architecture. As the figure suggests, our tool has four following modules:

1- *ERs Extractor*: this module is responsible to load the original electronic document and display it as a set of paragraphs. The author extracts the ER by selecting the appropriate portion of text.

2- *ERs Annotator*: represent the most important module that responsible to annotate the ERs by different ontologies (Domain, educational type, user preference) and return as result an RDF[2] (Resource Description Framework) file that contains a set of metadata. The extracted learning resources are annotated semi automatically with metadata about the semantic content of the resources by concepts of domain ontology (i.e. this module extracts all existing concepts from ER in order to propose the most frequently repeated ones leaving the choice to the user to accept it or not), and automatically the relationships between the resource and the original document, note that the original document should be stored in our database with a simple metadata schema that describe his language, his URL[3] and his storing date in order to preserve the references of the ERs, and manually the metadata about the educational type *(definition, explanation…)* which we are using the educational type ontology.

3- *Search Engine*: represent the access point into LOR which author can be research of ERs by both domain concept and educational type thought specific queries. For example, user would like to find an *illustration* about *compiler*, stored recently by "*Mourad*" and written in French. This query can be divided into three parts, the first concerning the Topic of domain

---

[1] ltsc.ieee.org/wg12/

[2] www.w3.org/RDF/

[3] Universal Resource Locator

"*Compiler*", the second is about the educational type "illustration" and the last represent some preferences of the researcher about ERs as creation date, contributor name (who was stored the ER) and the language of ER. The LOR can make some inference about educational type as "an example is a type of illustration" this genre of information can be retrieval from the educational type ontology. However, if no resources are available, the system should be able to propose other similar resources with the same preferences because it has a supplement layer of knowledge (ontologies).

4- Storage: this module stand for Storage of ERs in a format html and its metadata file generate by the annotator, also this module is responsible for displaying ERs according to the user visual preferences (police, background...).

The approach is implemented in Java programming language. Additionally, we use the Jena (Java Semantic Web Framework)[4] to generate the appropriate RDF files. These files are uploaded into repository which we used sesame[5] for built it (an open source RDF database with support for RDFS inferencing and querying), dedicated for searching over metadata, as well as for reasoning over the ontologies. For the construction of the ontologies, we use Protégé[6], an extensible, platform-independent environment for creating and editing ontologies [3].
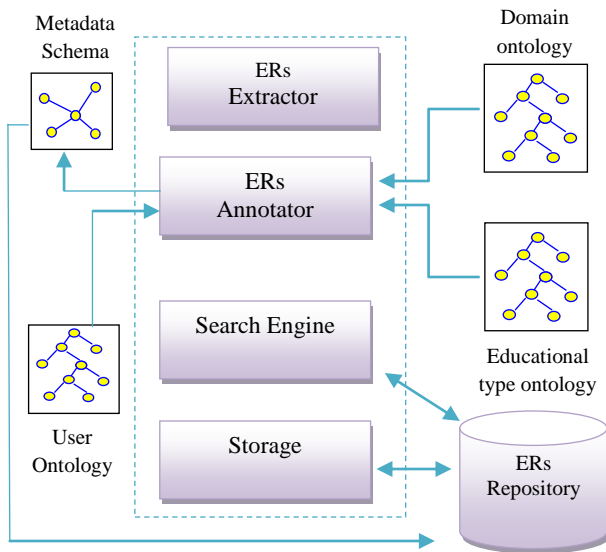
Figure 1: Architecture proposed for our tool

What does our tool provide?
Our tool is destined to authors' community that can be experts, teachers and others; its aim is to create new ERs by extracting it from an electronic document available on the Web and storing them into interoperable LOR. These ERs may be reused in

several contexts such as generating new educational materials (courses, chapters…), or just to find relevant ERs.

3. METADATA STANDARDS FOR E-LEARNING
Metadata for ERs are important to facilitate search, access, use and reuse of ERs. The work described in this paper is mainly focused on set of metadata (metadata schema). This set of last one must be contains all relevant elements that satisfied our requirement to describing and indexing the ERs.

Metadata is defined as data about data, to permit interoperability we need a standards. By the following we describe some of the common standards used in educational settings.

One simple but significant metadata schema is Dublin Core by the Dublin Core Metadata Initiative (DCMI) [4], Each Dublin Core element is defined using a set of 15 attributes like Title, Identifier, Language, and other. The major advantage of this schema is that contains a set of simple elements that can be used for all domains including education domain, furthermore a set of additional qualifiers "Qualified Dublin Core" are developed to further refine the base elements of this schema that allows more efficient indexing. We refer the reader to [4] for more details.

The Institute of Electrical and Electronics Engineers (IEEE) provides a standard for the description of learning resources widely used in educational context and applied in several LORs, called IEEE 1484.12.1-2002 Learning Object Metadata Standard (LOM). This schema based on nine categories (figure 2) and each of these categories contains several elements and thus, in total, LOM provides 76 data elements
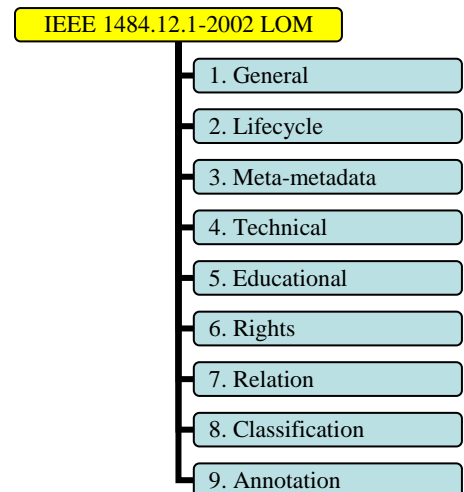

Figure2: LOM Categories

Work on the LOM schema has started in 1998, it is important to note that the use of all elements is facultative, and the user can work only with a sub set of this schema that is appropriate for his context.

4. METADATA APPLICATION PROFILE
The normal way to ensure the interoperability between other LORs is to define a profile based on a standard.

---

[4] http://jena.sourceforge.net/

[5] www.openrdf.org ; it is important to note that there are many third-party stores that can be used through the sesame API as OWLIM for OWL reasoning

[6] http://protege.stanford.edu/

An application profiles (AP) enable "mixing and matching" metadata elements, in order to meet specific requirements for a particular context. As an example, some communities may want to make certain elements mandatory or restrict the value space of a particular element [5].

In our work and for annotating the ERs, we have adopted a set of elements which are most relevant for storage and allow an adaptive search, thus we have used some elements from the following LOM categories (figure3): general, lifecycle, relation, education and classification, in total, we are defined a best-practice subset of 9 attributes which is summarized in the table1.
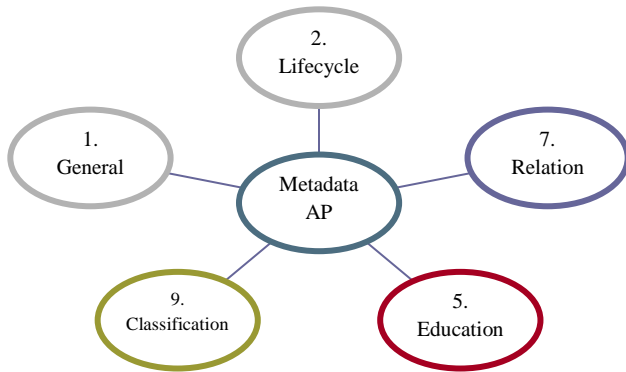
Figure 3: Metadata Application Profile Categories

Table 1: A.P Elements and value spaces

Category	Elements LOM	Value Space	R
General	1.1 Identifier	Any	A
	1.2 Title	Any	O
	1.3 Language	En, Fr,…	A
	1.4 Description	Any	O
Lifecycle	2.3.2 Entity	Instance of user preference ontology	M
	2.3.3 Date	W3c date format	O
Relation	7 Relation	Identifier of the original document	A
Classification	9.4 Keyword	Value taken from domain ontology	M
Education	5.2 Learning Resource Type	Value taken from educational type ontology	M

(R: Requirement; O: Optional; M: Mandatory; A: Automatic)
Note that   all categories are not repeatable.

4.1 Overview of proposed categories
We describe by the following overview the proposed categories. A set of attributes   are focused for resources retrieval and other are used for adaptation as well as for description.

General category: generally this section represents an identification of ERs. The search engine displays this information as a result of a specified query.

Lifecycle category: in order to store the contributor (instance of user ontology) and contribution date

Relation category: its aim is to keep reference of the original document

Education category: refers to the domain concept (ontology)

Classification category: refers to the educational type of the ERs (ontology)

4.2 Elements used for adaptation
As indicated in the previous section, the following table2 presents the set of metadata that can be used for adaptation.

Table2: Elements used for adaptation

Elements	Description
1.3 languages	Represent what the researcher prefers as language
2.3.2 Entity	represent the preferred author by the researcher
2.3.3 Date	Used when the researcher specify the desired age of ERs in the LOR.

4.3 Application profile binding
In order to benefit from metadata we must represent it by a common format that permits a machine-readable.

There are several possibilities to represent LOM metadata descriptions on the Web. One is to use an XML binding (current proposal), another alternative is the use of an RDF binding, RDF statements are triples with a subject indicating the resource being described, a property and a value for this property for the described resource (the property value is eventually another resource).

In the vision of the Semantic Web, not only the metadata space value is represented in ontological database, but also the metadata schema itself can be expressed in semantic web language and thus, to implement our metadata A.P We selected the RDF binding, table 3 shows the elements of the AP and the used attributes, this mapping is described in [6] as the syntactical representation of the LOM standard.

Table 3: Metadata AP binding RDF

Elements LOM	Used Attributes
1.1 Identifier	dc:identifier
1.2 Title	dc:Title
1.3 Language	dc:language using dcterms: RFC1766
1.4 Description	dc:description
2.3.2 Entity	dc:creator and lom:Entity using vCard:FN
2.3.3 Date	dcterms:created using dcterms: W3CDTF

7	Relation	dcterms:ispartof
9.4	Keyword	dc:subject
5.2	Learning Resource Type	Rdf:type

The corresponding abbreviations for namespaces used are:
  vCard: http://www. w3.org/2001/vcard-rdf/3.0/
  dc: http://purl.org/dc/elements/1.1/
  dcterms: http://purl.org/dc/terms/

Example

```
<rdf:Description rdf:about="&MyVoc:R150">
 <rdf:type rdf:resource="&Role:Introduction"/>
    <dc:Creator rdf:resource="&User:Teacher_1"/>
    <dc:Title>Java development kit<dc:Title>
    <dc:language>
          <dcterms:RFC1766>
              <rdf:value>en</rdf:value>
          </dcterms:RFC1766>
    </dc:language>
    <dcterms:created>
          <dcterms:W3CDTF>
              <rdf:value>2010-11-26 </rdf:value>
          </dcterms:W3CDTF>
    </dcterms:created>
 <dc:Subject rdf:resource="&Dom:Java"/>
</rdf:Description>
```

5. ONTOLOGICAL FOUNDATION

Ontologies, which have recently got a lot of attention in the context of the Semantic Web, provide a shared and common understanding of a domain that can be communicated between people and application systems like agents. They are developed to facilitate knowledge sharing and reuse.

Our architecture proposed requires uploading at first time all ontologies (figure 3) into LOR in order to benefit from reasoning on this knowledge base, in the other hand the ontologies can be augmented by human agents using the ontology editor Protégé [3] and then can be re-uploaded at any time using the sesame API.
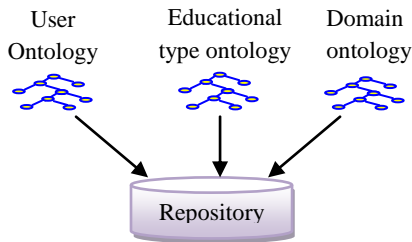
Figure 3: uploading the ontologies

Furthermore, it is necessary to have languages in order to create an adequate ontology that represents the semantics of the information; OWL[7] (Ontology Web Language) represents the most known language, which can represent ontologies and metadata. OWL provides relations between classes, cardinality, equality and richer typing of properties.

----

[7] www.w3.org/2004/OWL/

5.1 Domain Ontology

The goal of domain ontology is to provide a common conceptual vocabulary to members of a community of users who need to share their information in a particular domain. For testing our approach we reuse the domain ontology "Computer programming" that is used for the project LOCO-Analyst [7]

5.2 Educational Type Ontology

In order to describe educational resources that can be automatically re-used for educational purposes such as course generation, and based on the Ontology of Instructional Objects [8], we have developed an ontology that enumerates the educational type as shown below (figure 4):
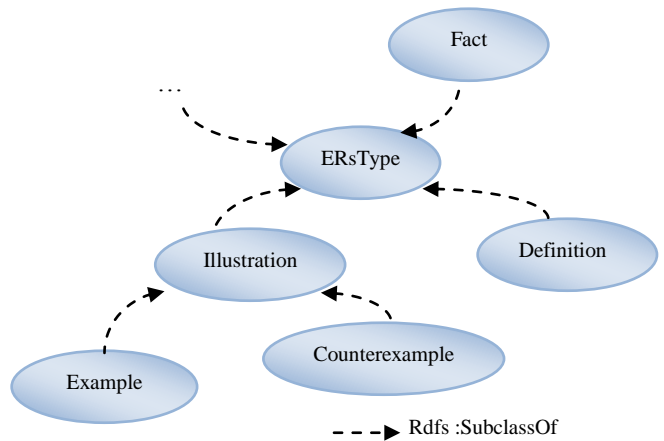
Figure 4: Graphical Representation of part of the Educational Type ontology

5.3 User preferences Ontology

The user model (also called learner model) manages information about users. In our context, its ontology is restricted to information about users' visual and content preferences of ERs.
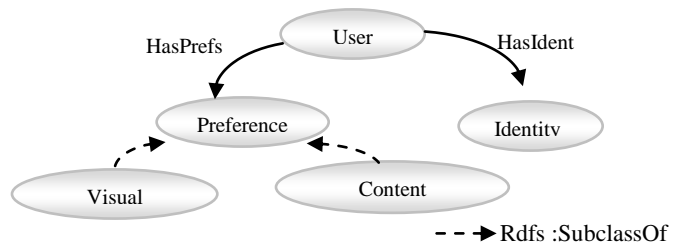
Figure 5: Graphical Representation of part of the User Model ontology

The proposed model is an individual model which stores informations about the users such as identity and preferences. As shown in figure 5, this part of ontology illustrates its main classes and properties, for example HasPrefs linked the user's class with his preferences class that can be Visual or content preferences.

6. RELATED WORK

Many other projects aim at creating and reuse of learning objects through different approaches. The TRIAL SOLUTION project

[9] is aimed for publication of personalized electronic document based on existing scientific books, its approach consist in slicing electronic book into learning resources and annotate them with metadata on author of the original book and semantic content to enable intelligent retrieval. These learning resources are stored into a repository available for teachers or students. The metadata is implemented in XML and the annotation is based on thesaurus for the book's domain. The core idea of TRIAL SOLUTION is similar to ours; our work aims to improve this idea by using three ontologies to annotate the learning resource in many facets in order to allow precise retrieval, and implementing the metadata in standard way in order to offer accessing in our repository from other context.

Other similar approach was implemented in PiggyBank [10], this last one enables collecting individual information items by extracting them from web pages, and store them with their metadata in a semantic bank (a communal repository of RDF to which a community of Piggy Bank users can contribute to share the information they have collected). Our tool goes a step further by adding an annotations module and builds a standard-based repository.

7. CONCLUSION AND FUTURE WORK

In conclusion, the resource repositories are the bases of e-Learning system construction. Semantic web offers a theory frame for the e-Learning resources sharing, alternation and intelligence. Domain ontology plays a very import role in it, other ontology like educational type augment the effective retrieval and offers precise research of resources. The user model ontology especially the preferences are to improve the search of such systems that restrict the search results to relevant resources relatively to the specific user.

The use of standard metadata schemas to develop an application profile ensures the interoperability between other systems; in our work we need the interoperability in order to allow the exchangeability of educational resources with other repository. The separation of different knowledge categories allows the users to enrich the basic ontological with *classes, rules, properties and axioms* thus allowing permanent evolution of this basis.

As future work we suggest to add some inference rules to metadata in order to enhance entering its values, we suggest also to extend our system to accept other kind of resources as image, video…   and to be able to read many other formats like PDF documents.

8. REFERENCES

[1] Wiley, D.A.: Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In: Wiley, D.A. (ed.) the Instructional Use of Learning Objects: Online Version (2000)

[2] T. Gruber, A Translation Approach to Portable Ontology Specifications, Knowledge Acquisition, vol. 5, no. 2, 1993, pp. 199-220

[3] Noy N. F., Fergerson, R. W. and Musen, M. A. The Knowledge Model of Protégé-2000: Combining Interoperability and Flexibility. EKAW 2000, 2000, pp. 17-32.

[4] Erik Duval, Neil Smith, Marc Van Coillie Application Profiles for Learning, Proceedings of the Sixth International ICALT'06, 2006

[5] Dublin Core Metadata Initiative, Dublin Core Metadata Element Set. Version 1.1, http://dublincore.org/documents/I 999/07/02/dces/.

[6] Nilsson, M., Palmér, M., Brase, J. The LOM RDF binding - principles and implementation- Proceedings of the 3rd annual ARIADNE conference, Katholieke Universiteit Leuven, Belgium 2003

[7] Jovanović, J., Gašević, D.,Brooks, C., Devedžić, V., Hatala, M., Eap, T. and Richards, G. (2008) LOCOAnalyst:semantic web technologies in learning content usage analysis. Int. J.of Continuing Engineering Education and Lifelong Learning, Vol. 18, No. 1, pp.54–76

[8] Carsten Ullrich. Pedagogically Founded Courseware Generation for Web-Based Learning -- An HTN-Planning-Based Approach Implemented in PAIGOS, September, Lecture Notesin Artificial Intelligence, Number 5260, Springer, 2008

[9] Buffa M., Dehors S., Faron-zucker C , Sander P : Towards a Corporate Semantic Web Approach in Designing Learning Systems: Review of the TRIAL SOLUTION Project. In: Proc. Of International Workshop   on Applications of Semantic web Technologies For E-learning, AIED, pp. 73-76. Amesterdam, Holland (2005)

[10] Huynh, D., Mazzocchi, S., & Karger, D. (2005). PiggyBank: Experience the Semantic Web inside your Web browser. In Proceedings of the 4th International Semantic Web Conference, Galway, Ireland (pp. 413-430).