

Annotation and research of pedagogical documents in a platform of e-learning based on Semantic Web

S. BOUKIL, C. DAoui, B. BOUIKHALENE, M.FAKIR

Information Processing and Telecommunication Teams
FST, Sultan MoulaySlimaneUniversity,
BeniMellal, Morocco.

Abstract—E-learning is considered as one of the areas in which the Semantic Web can make a real improvement whatsoever in finding information, or reusing of educational resources or even personalized learning paths. This paper aims to develop an educational ontology that will be used to annotate learning materials and pedagogical documents.

Keywords—*Ontologie ; Web sémantique ; XML ; RDF ; RDFS ; OWL ; métadonnées ; enseignement à distance.*

I. INTRODUCTION

At its creation by Tim Berners Lee in the early 90s, the goal of the Web was to allow any user to access and share large amounts of information on the net, and very quickly the Web has achieved its objectives. Therefore, we have a large volume of information, but no control of content. The Semantic Web is the solution proposed by the W3C (World Wide Web Consortium), this new vision of the Web aims to make Web resources understandable not only by humans but also by machines. To achieve this goal, W3C begins to develop new and better languages: XML (Extensible Markup Language), RDF (Resource Description Framework), OWL (Web Ontology Language) ... etc. The Semantic Web has used engineering knowledge to provide a tool for knowledge representation and it seems that ontologies were most suitable for such an environment. An ontology is a conceptual system that enables the sharing of knowledge between humans and computers and between computers. The work presented in this paper lies at the intersection of the Semantic Web with the field of distance education. This work aims to achieve a double goal: the first one is to design an application ontology that describes the educational materials used for university education, and the second one concerns the development of an application for annotation and retrieval of documents in educational exploitation of metadata that describe them.

II. SEMANTIC WEB

A. Semantic Web Hierarchy Model

The Semantic Web is a vision of the Web in which information is given explicit meaning to machines facilitating the processing and integration of information on the Web.

The Semantic Web is built on the ability of XML to define customized tagging schemas and the flexibility of RDF to represent the data. If the machines are supposed to do useful reasoning tasks on these documents, the language must go beyond basic semantics of RDF Schema. OWL has been designed to meet that need. OWL is part of a scalable stack of W3C recommendations with respect to the Semantic Web [1].

Fig. 1 shows the standardized technologies of the Semantic Web, which is cited below:

1) *XML: provides a syntactical surface for structured documents [2]., achieves interoperability of data to different platforms or systems using different languages. XML has been designed for document exchange, and define the data structure [3].*

2) *XML Schema: is the substitute of the DTD (Document Type Definitions). Itself uses XML syntax, but more flexible than DTD, provides more data types, services for XML documents more efficiently. If the XML is standardized data format then the XML Schema defines the data types of an XML document [3].*

3) *RDF is a standard for describing Web resources proposed by the W3C, as its name implies, RDF (Resource Description Framework) is a metalanguage for resource description framework, to make the information necessary to search engines more "structured" and, more generally, to all necessary tool for automated analysis of web pages [6]. RDF uses a particular terminology to indicate the various parts of statements. Specifically, the part that identifies the thing mentioned in the statement is called the subject (subject), the part that identifies the property or characteristic of the subject is called the predicate (predicate), the part that identifies the value of this property is called the object (object). Thus RDF statements are in the form of a triplet <Subjet, Prédicat, Object>. Graph model is used to query and process RDF [3]. In this model, a statement is represented by:*

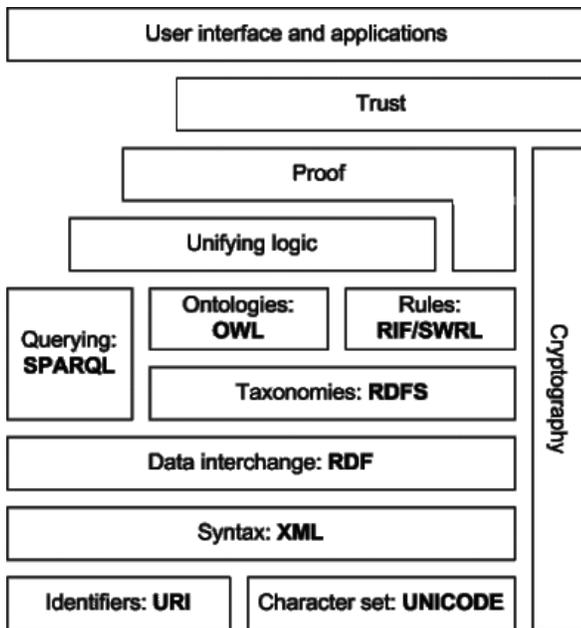


Fig.1. Semantic Web Layers

- a) a node for the subject.
- b) a node for the object.
- c) an arc, directed from the node subject to the node object, for the predicate.

So any RDF statement could be represented by the graph shown in Fig.2.

4) *RDF Schema: is a vocabulary for describing properties and classes of RDF resources [1]. It uses a kind of understandable system by the computer to define vocabularies resources described [3].*

5) *Ontology: RDF provides only basic relations of description, reasoning ability is limited. The context of the Semantic Web develops on the ontology layer and the layer of logical reasoning based on RDF to support knowledge representation and reasoning based on semantics. Ontology is a philosophical term introduced in the nineteenth century that characterizes the study of beings in our universe. In computer science, an ontology is a structured representation of domain knowledge in the form of a network of concepts linked by semantic links. The ultimate goal of the ontology is to show the implicit information accurately. W3C has defined OWL (Ontology Web Language) as the standard recommended language for describing ontologies.*

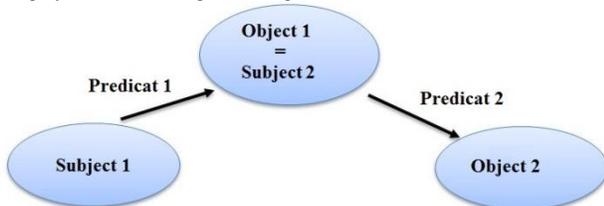


Fig.2. RDF statement in graph model.

6) *OWL: like RDF, OWL is an XML language enjoying the universal syntax of XML. It adds the ability to make comparisons between properties and classes: identity, equivalence, contrast, symmetry, cardinality, transitivity, disjunction, etc. The W3C provides OWL with three sublanguages with increasing capacity of expression, and it is as necessary that we choose the appropriate language.*

d) *OWL Lite is the simplest sublanguage of OWL, it is intended to represent hierarchies of simple concepts.*

e) *OWL DL is more complex than the previous one, it is based on description logic, hence its name (OWL Description Logics). It is adapted to reasoning, and ensures the completeness of reasoning and its decidability.*

f) *OWL Full is the most complex version of OWL, designed for situations where it is important to have a high level of ability to describe, even if they cannot guarantee the completeness and decidability of calculations related to the ontology [7].*

III. ONTOLOGIES AND E-LEARNING

A. Metadata :

We can define metadata as "data about data" treatable by a machine [8], in the case of pedagogical documents, document content is data and information on the authors, their interests and their pedagogical goals are metadata.

B. Metadata based on ontologies

In the context of the Semantic Web, ontologies provide specifically rich semantics, better than any other method of knowledge representation known. In a research topic of educational content on a platform of education, basing on the conceptual vocabulary defined in ontology may help to improve the accuracy of this research by avoiding ambiguities in terminology and allowing inferences decreasing noise and increasing relevance.

C. Ontologies for e-learning

In June 2000 the European Commission defines e-learning as "the use of new multimedia technologies and the Internet to improve the quality of learning by facilitating access to resources and services, as well as exchanges and remote collaboration". E-learning, and as other Web services, can benefit from the new vision of the Semantic Web while relying particularly on the potential of ontologies.

1) Need of e-learning systems

The different needs of e-learning system that ontologies play a role to fulfill, can be summarized in:

a) *Need for archiving and information research: An e-learning application is put online through the use of Web. Given the diversity and the exponential growth of pedagogical resources used in an e-learning type of education, it is increasingly difficult to find relevant pedagogical materials.*

b) *Need to share: problem with keywords to use to search for learning materials.*

c) *Need for reusing of pedagogical objects: Given the volume of increasingly growing pedagogical materials*

available on the net, just a small number of pedagogical objects are reusable. The search and selection of relevant text fragments, figures, exercises, from a document with the aim of their reuse in a new document has become almost impossible.

d) *Need for customization and adaptation*: A system of e-learning is for a community of users who do not have the same expectations, knowledge, skills, interests, etc. They are not able to understand or accept documents, except those of the organization, content and presentation are adapted to their needs.

IV. CONCEPTION AND IMPLEMENTATION

A. Construction process

The process of building ontology exploitable within a computer system is based on two steps: ontologization and operationalization. The ontologization consists of building a conceptual ontology. This means that we intend to provide a description of the target world, faced with this task we take into account the various sources of knowledge: glossary of terms, other ontologies, texts, interviews with experts, etc. The operationalization consists in encoding conceptual ontology obtained, using an operational language of knowledge representation (provided with mechanisms of inferences). It should be noted that this process is not linear and that many trips are a priori necessary to develop an ontology adapted to operational needs.

B. Conception of the application ontology

1) Choice of a construct methodology

To build the application ontology, there are different methods of construction, and the choice between these methods is performed according to our needs. The method developed by [Bernaras et al, 1996] was used in this work, it is based on three steps:

- Specify the application based on the ontology, in particular terms to collect and tasks to execute using this ontology.
- Organize terms using meta-categories: concepts, relations, attributes, etc.
- Refine and organize the ontology according to the principles of modularization and hierarchical organization.

This choice can be justified by two reasons:

- This method is suitable for the application ontologies rather than domain ontologies.
- It is structured around a set of terms that must be transformed into an ontology.

2) Construction principles

a) *Clarity and objectivity* [9]: all terms used in this ontology have been associated with definitions.

b) *Completeness* [9]: to respond to this principle definitions of concepts and relations of our ontology have been associated with conditions, others have been associated with necessary and sufficient conditions, but of course depending on the possibility to define these conditions.

c) *Maximum ontological extensibility* [9]: the definition of a term explains just the term itself, its definition cannot be the same except for a more general term, or a more specialized term.

d) *Principle of ontological distinction* [10]: the concepts in the ontology are sufficiently disjoint.

e) *Minimum semantic distance* [11]: there is a minimum distance between the concepts children of the same parents.

3) Representation of concepts :

Fig. 3 shows a hierarchical representation of the concepts used to model the pedagogical universe in our ontology.

4) UML class diagram

The class diagram shown in Fig. 4 illustrates concepts, attributes and relations linking concepts together:

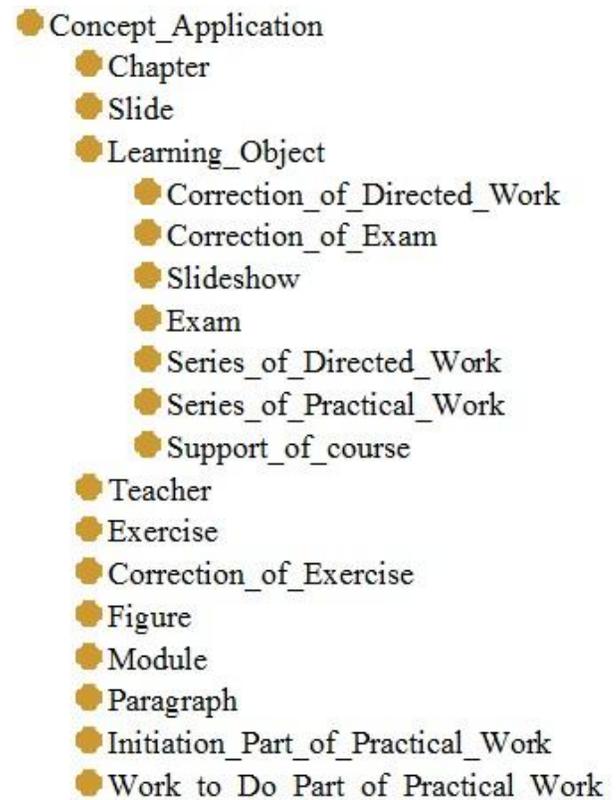


Fig.3. Hierarchical representation of the ontology

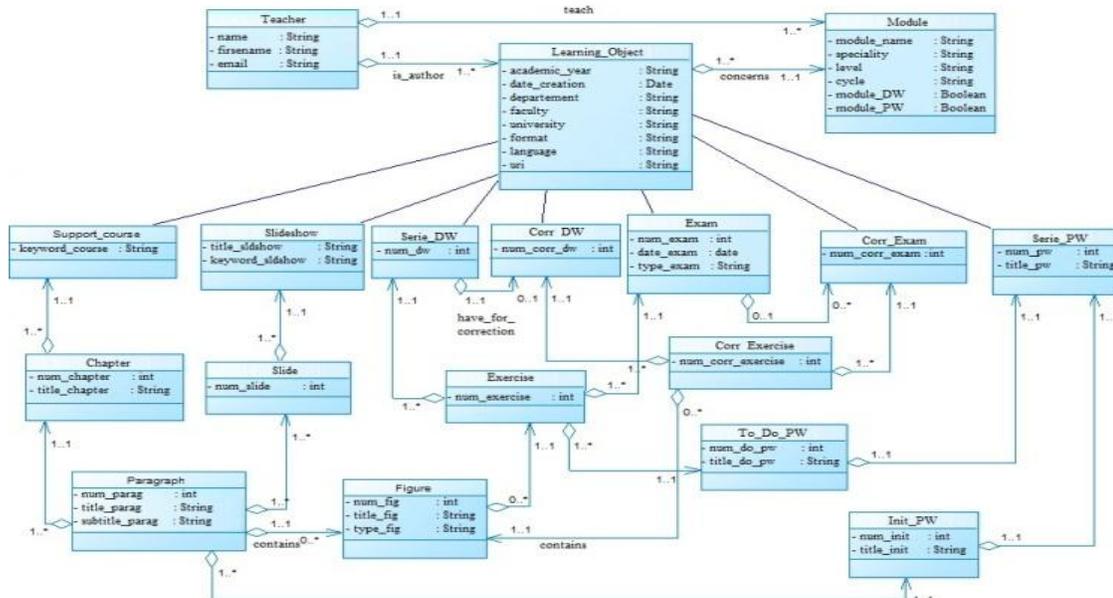


Fig.4. UML class diagram of the ontology

C. Implementation and use of the ontology:

The ontology editor « Protégé version 3.1.1 », was used to edit our ontology with the aim to automatically generate the OWL code corresponding as well as to generate the HTML documentation. A fragment of OWL code generated is illustrated in Fig. 5.

The process of ontology building can be integrated into the life cycle of an ontology as shown in Figure 6 [12].

```

<owl:onProperty>
<owl:ObjectPropertyrdf:ID="inverse_of_est_compose_chapitre"/></owl:onProperty>
</owl:Restriction>
</rdfs:subClassOf>
</rdfs:subClassOf>
<owl:Restriction><owl:onProperty>
<owl:DatatypePropertyrdf:ID="num_chapitre"/>
</owl:onProperty>
<owl:cardinality
rdf:datatype="http://www.w3.org/2001/XMLSchema#int">1<
/owl:cardinality>
</owl:Restriction>
</rdfs:subClassOf>
</rdfs:subClassOf>
<owl:Restriction>
<owl:cardinalityrdf:datatype="http://www.w3.org/2001/XMLSchema#int">1</owl:cardinality>

```

Fig.5. Fragment of OWL code generated

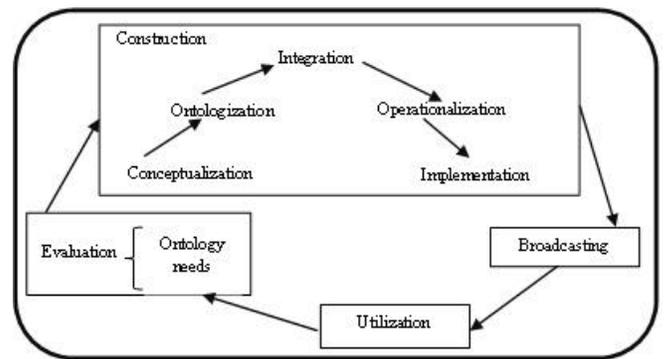


Fig.6. Life cycle of an ontology

D. Application platform:

The platform was developed in JSP (Java Server Pages). The JSP API is part of J2EE (Java 2 Enterprise Edition), it gives to developers the means to develop Web applications in a simple and powerful way. As it allows to separate the programmatic logic (java code) from presentation (HTML tags). It is also essential to use the framework Jena which is a Java framework oriented toward the development of Semantic Web applications, it presents a set of tools (API) open source developed by HP Labs Semantic Web Program to read and manipulate ontologies described in RDFS or OWL, and to apply some inference mechanisms. There are three types of users of this platform: administrator, teacher and student. Different tasks that can perform each type of user are shown in the following use case diagram:

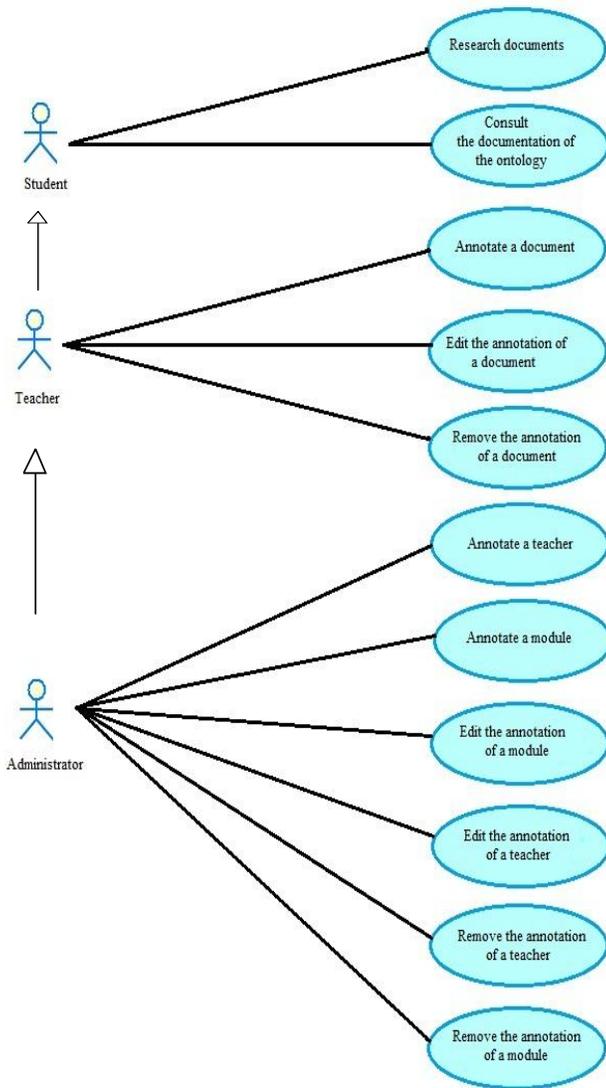


Fig.7. Use case diagram

It should be noted that the approach to annotate the documents consisted of:

- Add the metadata that describe a document to the file OWL that encodes the application ontology.
- Store documents in a specific location on the server.
- Manage access to documents with metadata "URI".

Fig. 8, shows an example of annotation (annotation of a module).

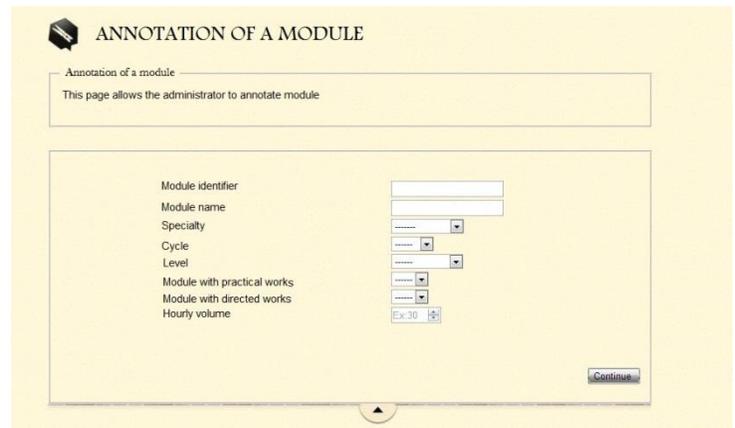


Fig.8. Annotation of a module

Regarding the search for a document, it takes three options, one option depending on document metadata, an option according to the author of the document and the last option as the taught module (Figure 9).

To exploit the ontology in the application, a dedicated query language is essential; SPARQL (his name is an acronym for SPARQL Protocol And RDF Query Language). SPARQL Conducts research on RDF graphs [4]. Thus, SPARQL is a programming interface between applications, an API as a Web services standard, and it opens the way to a universal API for querying structured data, in which the semantics of the query is no longer situated in the API, which limits the possibilities, but in the query itself. [5]



Fig.9. Research for a document

V. CONCLUSION AND PERCPECTIVES

This paper shows what is the Semantic Web is, and on what it is based in terms of standards and languages, and also covers the notion of ontology and the contribution of ontologies in Semantic Web context.

The objective of this paper was the conception of ontology to provide a vocabulary for the annotation and research of documents in a platform for distance education. However, this work is not perfect and can be improved in several areas, such as:

- Develop other ontologies and combine them with those made here to enrich the vocabulary used for annotating and research.
- Reuse this ontology in other platform based on semantic web techniques.
- Add the intelligent agent technologies at the application to provide reactivity with users

ACKNOWLEDGMENT

The authors of this article are grateful to all their colleagues who have contributed to the achievement of this article.

REFERENCES

- [1] S.Boutemdjit, "Web sémantique et e-learning.", cours ITF 6162, University of Montréal, Winter 2004.
- [2] P.Laublet, C.Reynaud and J.Charlet, "Sur quelques aspects du web sémantique", University of Paris-Sorbonne - CNRS (LaLLIC), University of Paris-Sud - CNRS (L.R.I.) & INRIA, University of Paris-X Nanterre, 2002.
- [3] R.Liu and M.Deng, "A research on e - learning resources construction based on semantic web", College of Educational Science, China West Normal University, Nanchong, Sichuan 637000, China. 2012.
- [4] <http://web-semantique.developpez.com/tutoriels/jena/arq/introduction-sparql/>
- [5] <http://www.lespetitescases.net/les-technologies-du-web-semantique>
- [6] <http://www.w3.org/TR/2013/WD-rdf11-concepts-20130115/>
- [7] X.Lacot, "Introduction à OWL, un langage XML d'ontologies web", 2005.
- [8] Y.Goita, "Les applications de XML à la production d'objets pédagogiques interactifs", 2001.
- [9] T.Gruber, "A translation approach to portable ontology specifications". Knowledge Acquisition, 5(2), 199-220, 1993.
- [10] S.Borgo, N.Guarino and C.Masolo, "Stratified ontologies: the case of physical objects". Paper presented at the ECAI96. Workshop on Ontological Engineering, Budapest, 1996.
- [11] J. Arpirez, A. Gómez-Perez, A. Lozano and S. Pinto, "(ONTO)2Agent: An ontology-based WWW broker to select ontologies". Paper presented at the Workshop on Applications of Ontologies and PSMs, Brighton, England, 1998.
- [12] F. Gandon, "Ontology engineering: a survey and a return on experience", rapport de recherche 4396, INRIA, 2002.