Agent of Semantic Annotation of Educational Resources Based on Ontology

Aziz Oriche, Abderrahman Chekry, Md. Khaldi

Abstract— Our goal is to describe the content of learning objects (LO) by semantically annotating with unambiguous information to facilitate the exploitation of these resources by software agents, these resources are defined by XML tags that structure is a tree DOM (Document Object Model) and on the same domain. We assume to have a domain ontology defined concepts, relations between these concepts and properties. We defined the intelligent agent AGSA featuring an annotation module representing a set of declarative rules to annotate nodes and their relationships. Metadata for this module are concepts expressed the father/son relationships type concepts/sub concepts between LO.

Index Terms— Semantic Annotation, Metadata, Multi-agent systems, Ontology, learning object (LO).

I. INTRODUCTION

Distance learning is based on the distribution of educational resources through a network (Internet or Intranet) or other electronic media. Educational resources can be used, reused or referenced are called learning objects (LO). However, with the exponential growth in the number of available resources, it is increasingly difficult for learners and teachers find educational content that best meets their needs.

To meet the ever increasing number of resources, the search engines should be able to provide more accurate answers and handle more complex queries integrating knowledge of the user [1]. Given the need for a formal description of the content of teaching materials online, we see that it is interesting to integrate software agents able to label semantically textual and multimedia resources through metadata on the one hand and to exploit to improve information retrieval, on the other hand. The formalization of existing web pages is defined as the task of semantic and automatic annotation.

The semantic annotation of learning resources using metadata allows a better interpretation of their content [6]. When these annotations are available, it is possible to formulate queries based on ontology vocabulary, reasoning about these annotations, and collect responses combining data from different documents.

In this paper, we propose an approach to semantic annotation of educational object guided by anontology. Our work consists to define the concepts of the ontology we followed, then we define the annotation of the intelligent agent we have developed for the task of annotation module. Finally, we illustrate the results of experiments of our approach.

Manuscript Received December, 2013.

Aziz Oriche, Université Abdelmalek Essaâdi, Laboratoire, LIROSA & LNTE, BP: 93002, Tétouan, Maroc.

Abderrahman Chekry, Université Abdelmalek Essaâdi, Laboratoire, LIROSA & LNTE, BP: 93002, Tétouan, Maroc.

Mohamed Khaldi, Professor NUS (normal upper school) martil, Tetouan, BP: 209 Morocco.

II. RELATED WORKS

With the emergence of the Semantic Web, semantic annotation has been the focus of many projects and applications. Since the availability of annotated content on the Web today has become one of the key challenges for the realization of Semantic Web. We will review the semantic annotation tools dedicated to e-Learning applications.

[8] Propose a learning information retrieval system based on a semantic annotation process with Contextual Exploration. The system Use an approach discursive organizations of natural language texts to define another kind of learning objects retrieval. It is a complement system able to use morph syntactical extensions for learning objects terms. This approach explains how a new kind of learning objects retrieval system is implemented by using semantic and discourse automatic annotation of learning objects according to their types (Definition, Example, Exercise, etc.). The learning information presented by the author of a document is captured and the learning or the teaching process for the student or the instructor is respectively facilitated. Extracting learning information enables a person to combine multiple objects and compose personal lessons for an individual learner. a machine learning technique is applied to sort the results according to their similarity with the request term.

[9] Proposed a semantic annotation of omnipresent learning environments. Skills-based learning environments are utilized to endorse the attainment of practical skills, decision making, communication, and problem solving. It is significant to provide feedback to the students from these sessions and remarks of their actions may inform the estimation process and help researchers to better comprehend the learning process. After a series of prototype demonstrators, they have explored the use of semantic gloss in the recording and consequent understanding of such simulation environments. Their Semantic Web approach is outlined and conclusions drawn as to the appropriateness of diverse annotation techniques and their combination with ubiquitous calculating techniques to provide new mechanisms for both student feedback and increased understanding of the learning environment.

[10] Has proposed a technique named charaParser, a software application for semantic for morphological descriptions. A morphological description contains a set of descriptive sentences, describing the physical characteristics of an organism. A descriptive sentence in morphological descriptions is a string of text set off by a semicolon or a period and describes the characteristics of a structure (an organ or a part of an organ). Historically, to make morphological descriptions compact, function words that do not carry meanings (such as articles, a, an, the, and auxiliary verbs, is, are) are often omitted,

making descriptive sentences more like fragments or phrases than standard English sentences. This sentence describes the leaves of a plant and contains “subsentences” describing bases and margins of the leaves. The goal of CharaParser is to convert description sentences into annotations in XML in such a detailed manner that the annotations themselves carry all the information stated in the original sentences. Although the native format for the annotation is XML, the annotations can be reformatted in a number of semantic aware data formats, such as Resource Description Framework/Schema (RDF/S and, for the biosystematics domain, Structured Descriptive Data (SDD) recommended by Taxonomic Database Working Group (now Biodiversity Information Standard). SDD is used by key generation software such as Lucid to generate organism identification keys. We have developed software to reformat the XML annotations into these formats.

QBLS [11] is a learning system for instructors and students. It proposes annotations using an RDF description. The course is structured referring to a pedagogical ontology constituted of cards (definition, example, procedure, solution, etc), then the pedagogical resources are created (course, topic, concept, and question). These resources deduced from the initial course are stored with their respective annotations in “a database of pedagogical knowledge”. Students can thereafter practice how to resolve some questions, or learn more details about a definition, etc. When the user formulates a request, the search engine CORESE is activated to search the pedagogical cards as response to the user's query.

[12] Proposed a keyphrase extraction tool for semantic metadata footnote of learning materials. Keyphrases play a significant role in depicting a document. In learning management systems they lead to enhanced information retrieval. On the other hand, comparatively few learning documents have key terms assigned and therefore finding techniques to computerize the extraction is desirable. The goal of their proposed technique is to depict the generation of list of keyphrases from a document using part of speech tagging and ranking them by means of formatting features implemented on them by the author rather than trusting only on statistics (such as term frequency).

We also denote the SOAF system [13] which proposes architecture to extract semantic descriptions of multimedia learning resources automatically. It is based on Latent Semantic Indexing using the representation of the resources in a vector space through their visual features. SOAF considers three types of metadata that might describe a learning object:

- Low-level features which generate automatic semantic indexing.
- High level descriptors provided by authors (title, date of creation, etc.).
- Collaborative annotations that are given by users.

III. DEFINITION OF LEARNING OBJECT (LO)

The concept of learning object manages to unite the efforts made by the various interest groups because it brings together a number of assets recognized at different levels: economic, educational or technical.

Different interest groups, educational institutions, training companies and standardization organizations fail to unite one general definition of learning object. In [2], a pedagogical document (PD), or Learning Object (LO) is defined as "any entity, digital or not digital, can be used for learning, education or training".

A learning object can be reused for different purposes, different platforms, or different public [3]. LO can take the form of any learning material involved in education to help the learner in his course or in monitoring its travels in different platforms of online learning.

In our case, the learning object is the content of a node in the tree DOM (Document Object Model, a W3C Recommendation). Each node of the tree represents a structural element (tag) and has a son node containing the text. The nodes are related by links father/son who represents their nesting tree. The annotation of a node is defined by the presence of one or more instances of learning objects in this node and concepts with which they were aligned. A node is annotated an instance of one or more ontological concepts.

IV. ONTOLOGY EXTENDED AN ANNOTATION CONCEPTS

Ontology is a structured set of terms and concepts representing the meaning of a field of information through metadata. The ontology is itself a model representative of a set of data concepts in a domain and the relationships between these concepts. In 1993, Gruber proposed definition remains until now the most cited definition: "An ontology is an explicit formalization of a common shared conceptualization" [5].

In our work, we consider ontology is defined by a set of concepts, relationships between concepts, objects belonging to these concepts and properties. It allows to implement rich and varied relations between these objects. This lexicon is enriched by the information extraction agent defined in [1]. We defined three sets of concepts for this ontology:

- A set of concepts to represent an LO
- A set of concepts describing instances of LO
- A set of concepts describing the father / son relationships type concept / sub concept between LO

The ontology that our system uses for the annotation, defines two sets of concepts: a set of concepts to represent an LO, and a set of concepts describing instances of LO. All instances of concepts are automatically enriched by the information of extraction agent defined in [1] and the set of concepts is enhanced manually by the expert validates the words by changing or deleting some or linking to a more specific or more general concept.

In addition to metadata whose semantics is defined in the ontology, we defined a specific set of metadata for the task of annotation. This metadata express the father/son relationship type concept/sub concept between instances of concepts identified in the educational documents. More the concepts are close to the root, more their meaning is general and, conversely, the closer they are leaves; their meaning is more specific to the field concerned.

Metadata annotation, defined concepts and relationships to describe instances of concepts in our training corpus and their nesting tree, we have defined the following seven concepts: Object, Element, LearningElements, LearningObject, LearningObj o, LearningElem e and LearningNeutral, all of these concepts are enriched automatically by the annotation module of our AGSA agent.

V. ARCHITECTURE OF AGENT ANNOTATION

We choose a module architecture (Fig. 1) by achieving our agent. This architecture consists of four modules: the knowledge base module, the processing module, the communication module and the annotation module. Since the objective of our agent is the annotation of educational resources, the annotation process is the task of annotation module described in detail in Section D. for the sake of interoperability; we have respected the standard FIPA-ACL [4] for inter-agent communication module.

A. Knowledge module

This module contains all the knowledge of the agent. These skills are composed of concepts and rules OWL ontology we have defined (classes, instances, properties and operations) and RDF metadata annotation. The terms of the knowledge base of the agent are directly operated; no transformation of knowledge is required.

B. Processing module

The processing module is the basic control structure of the agent; it includes techniques for interacting with the knowledge module for different actions (updating the knowledge base, triggering the behavior of the agent; verification the environment ...) and manages calls with the communication and annotation module if actions should be performed.

C. Communication module

The communication module organizes the queries in the form of a message from the other agents of the learning environment (extractant agent and treatment agent [1]) after that, he called the processing module. It also receives requests for messaging annotation module. These applications transmissions are replies received requests.

D. Annotation module

To generate the annotation concepts of ontology we adopted the following definition: "a node n annotated by concept e" and "a node n' annotated by concept é", if n is a descendant node of n' (conversely n' is the ascending node of n) then e is sub concept of é (inversely é super concept of e). The concepts of annotation are defined as follows:

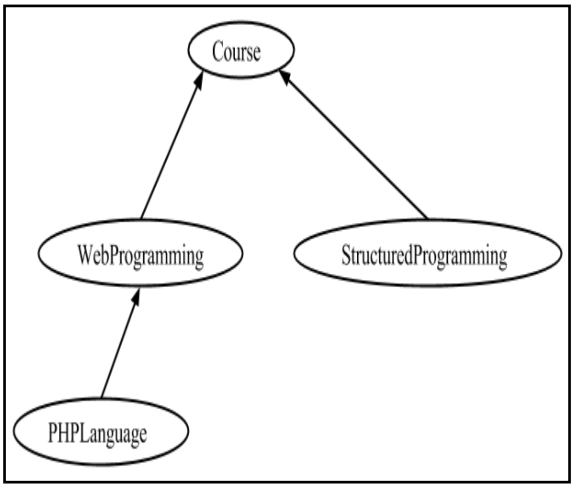

Figure 1: Convergent concepts

- Element: Is defined as a direct super concept of all concepts of ontology.

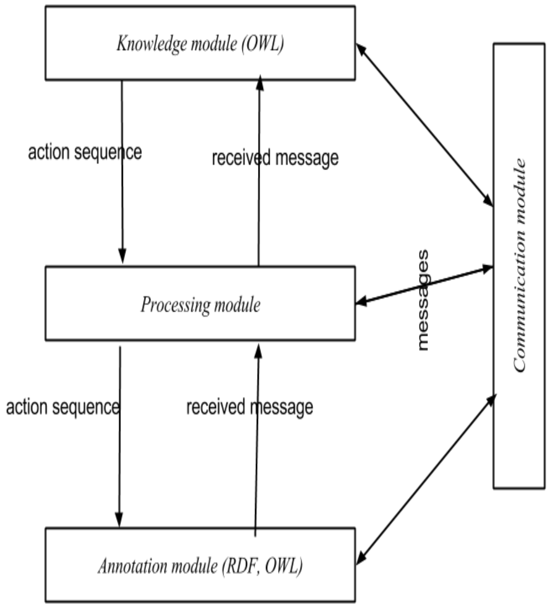

Figure 2: Architecture of Agent

- LearningObjects: Is defined to annotate a node of document whose text contains several learning objects of different concepts. An instance of LearningObjects type is related to concepts of learning objects identified in this node via isPointing property.

For example, a node containing the text "Programming is the comprehensive process that leads from an original formulation of a computing problem to executable programs"(1). It involves activities such as analysis, understanding, and generically solving such problems resulting in an algorithm "(2).

Learning objects are located in this text: (1) aligned with the definition concept and (2) aligned with the concept Explanation. This node is annotated as an instance of concept LearningObjects and is related via the property isPointing Definition and Explanation concepts.

- LearningElem e: Is defined for each e ontological concept, annotate nodes of documents containing learning objects of convergent concepts.

We say that two concepts are convergent if the instances of these concepts in the same domain. In (Fig. 2), the concepts WebProgramming and PHPLanguage are converging cons by the concepts WebProgramming and StructuredProgramming are divergent.

In this example, the node where several LO of convergent concepts PHPLanguage and WebProgramming and their sub concept Course are identified in the text, this node is annotated as an instance of LearningElem Course

Example: Consider a node containing the text " The word object has several meanings depending on the discipline: In law, an object is one of the four basic conditions necessary for the formation and validity of a contract, in physics, its' is anybody that can be detected, in mathematics, an object is a well-defined ideality".

A several terms are located of the concept Definition, this node is annotated by LearningElem Definition.

- LearningElements: Any concept LearningElements e generated such as e is sub concept direct of Element is considered LearningElements sub concept. The LearningElem e concepts generated are descendant concepts of LearningElements.
- Object: Is defined as a direct super concept of concepts: Element, LearningElements, LearningNeutral and LearningObjects. Annotated nodes are all instances of this concept.
- LearningObj o: Is defined for each ontological concept e, annotate nodes documents with learning objects of divergent concepts (Fig. 3).

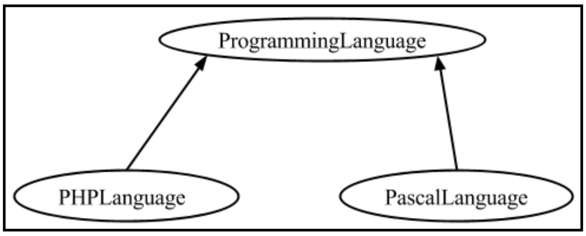

Figure 3: Divergent concepts

In this example, the node where several LO of divergent concepts PHPLanguage and PascalLanguage are identified in the text, this node is annotated as an instance of LearningObj ProgrammingLanguage.

- LearningNeutral: is defined to annotate a node of document containing neutral learning objects. An instance of LearningNeutral type is related to concepts of learning objects identified in this node via noLinks property.

Neutral LO is an object that has no hierarchical relationship whatsoever of convergence/divergence or the family relationship with other objects.

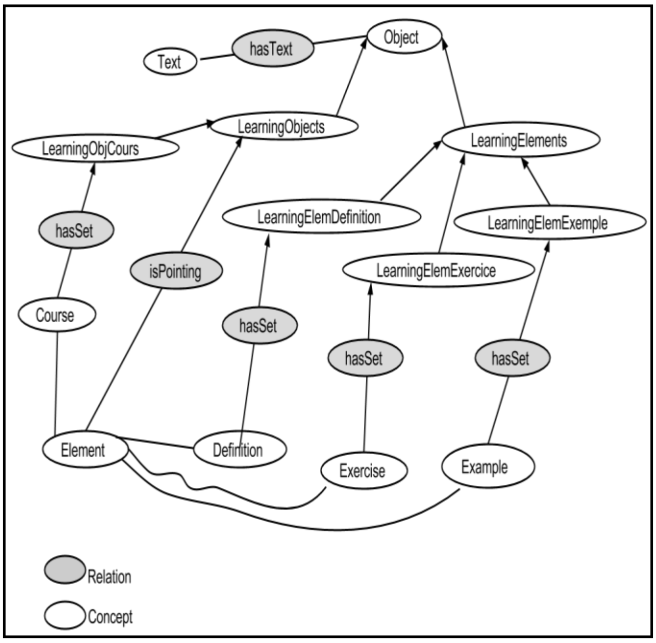

Figure 4: Annotation concepts graph

Relationships and properties of these concepts are depicted in the annotation concepts graph (Fig. 4). For each concept graph annotation, we defined rules to cover all possible forms of learning object. We started with a text example for each concept to generalize all possible structures. This method sets incrementally strong of rules that form the language resources used in our approach base.

VI. EXTRACT RULES OF ANNOTATION NODES

These rules are used to annotate a node as an instance of a concept in the field

1) If a node n contains a single learning object, a concept aligned to e, this node is annotated by e
2) If a node n contains several learning objects of many different concepts, then this node is annotated by the metadata LearningObjects.
3) Any node LearningObjects type is indexed using the relationship isPointing by any concept e related to this node.
4) We generate instances LearningElements for nodes may represent a set of instances of convergent types.
5) If a node contains multiple learning objects and if they are only aligned to convergent concepts, this node is typed by the metadata LearningElem e sub concepts of LearningElements .
6) If a node contains multiple learning objects and if they are aligned to divergent concepts, this node is annotated with metadata LearningObj o.

7)A neutral node is annotated by the LearningNeutral concept.

VII. CHARACTERISTIC OF AGENT

- AGSA is autonomous, it just works without human action to trigger it, it detects if there is a change in the ontological basis or in educational materials for training to update the database associated with this change annotation.
- He is still active throughout training and even proactive insofar as it is able to annotate the educational resources of training, to intervene in the annotation process independently and suggest the following actions according to base of RDF annotation built progressively when learning.
- It is fast and efficient because it can quickly find annotations in a knowledge base RDF, much faster than other annotation tool.
- It is adjustable because its operation depends on the user's profile (reaction time, and content adaptation).
- It is scalable since it is possible for the user to enrich the knowledge base of its own annotations.
- It is sociable as it is able to interact with the extraction agent [1], designed for the extraction of pedagogical concepts.

VIII. EXPERIMENTS AND RESULTS

The objective of this step is to evaluate the performance of our annotation agent. An important indicator is the number of correctly annotated with respect to the number of documents processed concepts. To validate our approach annotation learning objects, we have developed the agent AGSA (**A**gent **G**enerating **S**emantic **A**nnotation) using the platform for developing multi-agent systems *JADE* [7].

Our corpus annotation agent has a set of 30 documents online training covering several concepts. The test corpus was annotated by an expert for each learning object spotted, he says its type.

To evaluate the process of annotation, we refer to the metadata extraction defined by the extraction agent [1] and the concepts of the ontology. The results of the process annotation performed by our AGSA agents are illustrated in the following Table I:

CONCEPTS	NUMBER OF CONCEPTS	ANNOTATED CONCEPTS
LEARNINGOBJECT (CLO)	85	72 (84.7%)
LEARNINGELEMENTS (CLE)	98	88 (89.8%)
LEARNINGNEUTRAL	37	29 (78.4 %)
Learning Object (LO)	630	495 (78.5%)

Tableau I: Results of the step of annotation

With:
CLO: Contain learning objects annotated by LearningObject and LearningObj o.
CLE: Contain learning objects annotated by LearningElements and LearningElem e.
We note that the accuracy of the annotation exceeds 78% for the total set of annotated concepts. We see very clearly that the concepts most marked in the knowledge base are instances of LearningElements. This result is not surprising since the task of recognition of LearningElements in Information Extraction is the one that gives the best results because it relies heavily on existing glossaries and other vocabularies. During the experimental phase, we have also seen that the quality of the annotation is closely linked to the quality of learning objects extraction.

IX. CONCLUSION

In this paper, we proposed an approach of annotation learning objects based on an ontology and results retrieval of learning objects extraction defined in [1] agent. Currently, our work has a significant interest in the field of online learning, distance education (e-learning), education, etc. To evaluate our approach, we have developed the agent AGSA which includes a module annotation to express father/son relationships type concept/sub concept between instances of concept. We note, through the evaluation results, our approach allows access to knowledge that is expressed in online training, and bring learning objects contained in these formations, something that cannot do with a system of classical training.

REFERENCES

1. Aziz O, Abderrahman C, Mohamed K. "Intelligent Agents for the Semantic Annotation of Educational Resources", International Journal of Soft Computing and Engineering (IJSCE), Volume-3, Issue-5, November 2013.
2. Learning Technology Standards Committee (LTSC) P1484.12.1-2002 Learning Object Metadata Working Group, "Draft Standard for Learning Object Metadata", Available: http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf, 2002.
3. Sabine G., K. Ramesh R., "Concept Guide on Reusable Learning Objects with Application to Soil, Water and Environmental Sciences", Dec. 2007.
4. Foundation for intelligent physical agents, FIPA ACL message structure specification Available at: http://www.fipa.org/specs/fipa00061/SC00061G.pdf
5. Gruber T. "A Translation Approach to Portable Ontology Specifications", Knowledge Acquisition, pp. 199-220, 1993
6. Mouhamadou T., Nathalie P., Nacéra B., ″Contextual and Metadata-based Approach for the Semantic Annotation of Heterogeneous Documents″, SEMMA 2008 CEUR workshop proceedings, Vol 346, ISSN 1613-0073.
7. Fabio B., Giovanni C., Dominic G., "Developing Multi-Agent Systems with JADE", Series Editor: Michael Wooldridge, Liverpool University, UK, February 2007.
8. Rim F., Boutheina S., Jean-Pierre D., "Relevant learning objects extraction based on semantic annotation of documents," 2nd International conference on web intelligence, Mining and Semantics, No.47, 2012.
9. Mark J., Danius T., Kevin P., David C., Eloise M., Mary G., "Semantic annotation of ubiquitous learning environments," IEEE Transaction on learning technologies, vol. 5, No. 2, 2012.
10. Hong C., "CharaParser for fine-grained semantic annotation of organism morphological descriptions," Journal of the American Society for Information Science and Technology, vol. 63, No. 4, pp. 738-754, 2012.
11. Dehors S., Faron-Zucker C., Kuntz R. "Reusing Learning Resources based on Semantic Web Technologies". In Proceedings of the International Conference on Advanced Learning Technologies, Kerkrade. 2006.
12. Jain.S., Pareek.J., "KeyPhrase extraction tool for semantic metadata annotation of learning materials," International conference on signal processing system, pp. 625-628, 2009.
13. Cernea D., Moral E., Gayo J. "SOAF: Semantic indexing system based on collaborative tagging". Interdisciplinary Journal of E-learning and Learning Objects, Vol 4. 2008

AUTHORS PROFILE

Aziz Oriche, is a young doctoral student at the University Abdelmalek Essaâdi, Faculty of Sciences, Tetouan Morocco, Computer Lab, Operations Research and Applied Statistics, he earned a master's degree in software quality and it has 4 publications as author and co-author.

Abderrahman CHEKRY, is a young doctoral student at the University Abdelmalek Essaâdi, Faculty of Sciences, Tetouan Morocco, Computer Lab, Operations Research and Applied Statistics, he earned a master's degree in software quality and it has 4 publications as author and co-author.

Mohamed KHALDI, Professor NUS (normal upper school) Martil, Tetouan, BP: 209 Morocco.