Ontology based semantic metadata extraction system for learning objects

Ramzi Farhat, Baraâ Jebali and Mohamed Jemni

Research Laboratory of Technologies of Information and Communication & Electrical Engineering (LaTICE)
University of Tunis, 5 Avenue Taha Hussein, B.P. 56, Bab Menara, Tunis, Tunisia
jebali.baraa@gmail.com,ramzi.farhat@esstt.rnu.tn,
mohamed.Jemni@fst.rnu.tn

Abstract. Educational metadata play a crucial role in enabling learning objects' discovery for an efficient use. Consequently the e-learning community has developed several educational metadata schemas (e.g. IEEE LOM standard). Unfortunately to implement advanced tools it has been found that those metadata are not sufficient. Thus, many research works suggest the use of educational semantic metadata. However the main barrier is the fact that providing manually semantic metadata still a hard and complex task for authors. Consequently we propose an ontology-based approach allowing the automatic extraction of semantic metadata from a specific sub-set of IEEE LOM metadata. Experimentations results are presented and discussed.

Keywords: Semantic metadata, learning objects, ontology based information extraction, LOM.

1 Introduction

The educational content is delivered, in most of cases, as learning objects. Tedious problems related to their accessibility and to their indexation arise. Consequently educational metadata is proposed as a solution to smooth over these problems. Within the past years, many metadata standards and norms were proposed, and the IEEE-LOM standard has being widely adopted by the e-learning community [1]. The learning object's author has to fulfill metadata manually. In addition to the fact that this is a tedious process, metadata are not designed to be understandable by a machine. Consequently implementing intelligent services - such as semantic search engines and content personalization services - is not yet possible.

On reflection, it is critically important to add a semantic metadata layer for the learning objects to ensure its common comprehension by humans as well as machines. For the same purpose semantic metadata was introduced by the W3C with the appearance of the semantic web. It is defined as a metadata that refers to a given

G. Chen et al. (eds.), *Emerging Issues in Smart Learning*,
Lecture Notes in Educational Technology, DOI 10.1007/978-3-662-44188-6_34

domain ontology which presents a common vocabulary to describe domain's concepts and relationships between them [2].
In e-learning, research projects related to semantic metadata have few impacts in the practice. This can be explained in part by the fact that adding semantic metadata to a learning object is a hard and complex task for authors. Moreover, many learning objects already exist and enriching them by semantic metadata will be a tremendous and a time consuming task. Moreover semantic metadata are not universal because they depend on domain ontologies which vary in most cases from a community of practice to another. Consequently the task must be repeated each time the learning object is used in a new context. Thus we can say that the creation of semantic metadata is not suitable to be done by a human.
On this account, we proposed to generate semantic metadata automatically. We proposed an ontology based information extraction approach in order to guarantee the consideration of the use context, always modeled by ontologies. The input of our system is a set of LOM metadata and a domain ontology. The output of our system is a set of semantic metadata presented in the RDF formalism.
The paper is organized as following. The next section gives an overview of our approach. The third section contains experimental results. The last section is a conclusion.

2 Related works and our approach overview

In our point of view, semantic metadata is not supposed to replace classic metadata. In fact, we are for an evolutionary approach (vs. a revolutionary approach) which is compatible with existing tools and existing practices. Moreover, in our opinion the semantic metadata must not be integrated to the learning object's delivery package to conserve its interoperability. In fact they are specific to a domain ontology which is not necessary shared by all its users.
In addition to that, we suppose that semantic metadata must contain relevant information about the content with author's imprint and it must be expressed in a machine "understandable" and processable way. For that, we proposed to use LOM metadata as input in our approach which is summarized information -provided by the author- about the learning object educational content. . After these considerations, we have designed an approach to extract automatically semantic metadata. Firstly we have to choose relevant information from the LOM metadata to take it as input to our system [3, 4]. The domain ontology is also supposed as input. Then we have investigated how to ensure an automatic semantic metadata extraction form the available inputs. Here, we are strongly influenced by the ontology based information extraction (OBIE) domain, which is a subfield of information extraction. In fact, an OBIE system is a system that processes unstructured or semi-structured natural language text through a mechanism guided by ontologies to extract certain types of information and presents the output using ontologies [5]. Clearly, the output of such

systems is the basic element to build semantic metadata. For that, we proposed a customized ontology semantic metadata extraction system (OBSemE) architecture presented in [6]. Our system is composed of three main parts: ontology parser, pre-processor and Information extraction module. The ontology parser is used to extract the concepts of the ontology and the relations between them. The pre-processor prepares the LOM metadata to be suitable as input for the information extraction module. The pre-processing module is based on natural language processing technologies. The information extraction is explained in the following section.

3 Extraction method and experimental results

The core of our approach is based on the OBIE principles and the semantic web concepts in order to extract automatically semantic metadata from a preselected subset of LOM data elements. We proposed the use of an OBIE algorithm called Hieron [7] based on the published promising experimental results [8]. The use of this algorithm leads to tow steps of processing. A training step, in which the system learns how to classify the terms using the ontology concepts and a classification step in which the system begins his automatic processing.

To validate our approach we have developed a system called OBSemE (Ontology Based Semantic Metadata Extraction). After the implementation of our system, we moved to the experimentation step. To do that we have prepared a data set composed of computer science domain ontology and a set of LOM files.

In the aim of evaluating the obtained results [6] we used ontology based information extraction metrics called augmented precision and augmented recall based on a hierarchical error metric called BDM [9]. First results are presented by the curves in Fig. 1.

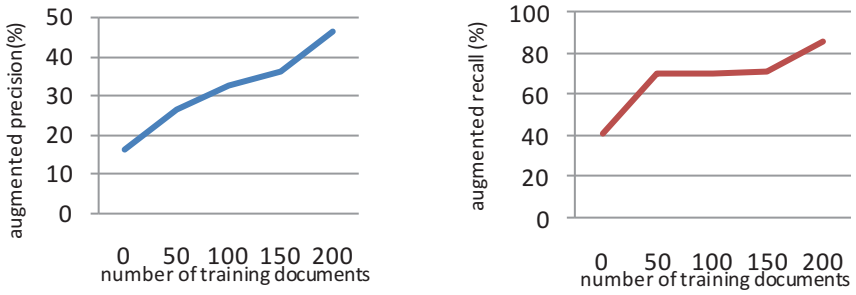

Fig. 1 Augmented precision and augmented recall curves

Obviously, more the training step is rich more the system's precision and recall are higher. Both, the number and the quality of the LOM files affect the precision remarkably. In fact, semantic metadata generation depends on many factors else. It

depends on the quality of the information contained in the documents and their coherence with each other and with the ontology [10].

4 Conclusion

In this paper we have presented an automatic ontology based semantic metadata extraction approach. The input of our approach is a subset of LOM metadata which is describing the educational content and is highly available. The output of our approach is a set of semantic metadata when each one contains a concept and its degree of relevance. The extraction process is done throw the execution of an adapted version of the Hieron algorithm. In order to validate our approach we have developed a prototype called OBSemE. The results of the experimentation proved that it is possible to talk about automatic generation of semantic metadata with existing technologies.

References

[1] Duval, E., Hodgins, W., Sutton, S., and Weibel, S.L.: Metadata Principles and Practicalities. D-Lib Magazine, vol 8 (4), ISSN 1082-9873, pp. 13 (2002),

[2] Scerri, S., Abela, C. and Montebello, M.: semantExplorer: A Semantic Web Browser. In : IADIS International Conference WWW/Internet, Lisbon, Portuga (2005)

[3] Farhat, R. and Jebali, B.: OBSemE: An Ontology-based Semantic metadata Extraction System for Learning objects. In: The Fourth International Conference on ICT & Accessibility, ICALT 2013, 24-26 October, Hammamet, Tunisie (2013)

[4] Jebali, B. and Farhat, R.: Ontology-based semantic metadata extraction approach. In: International Conference on Electrical Engineering and Software Applications, Tunis: Hammamet (2013)

[5] Ochoa, X., Klerkx, J., Vandeputte, B. and Duval, E.: On the Use of Learning Object Metadata. In: The GLOBE Experience. EC-TEL ECTEL: European Conference of Technology Enhanced Learning, Palermo, Italy, 20-23, Lecture Notes in Computer Science, volume 6964, Springe, pp. 271-284, September (2011)

[6] Al-Khalifa, H. and x Davis, H.: The Evolution of Metadata from Standards to Semantics in E-Learning Applications. In Proceedings of Hypertext'06, Odense, Denmark (2006)

[7] Dekel, O., Keshet, J. and Singer, Y.: Large Margin Hierarchical Classification. In: The 21stInternational Conference onMachine Learning ICML-2004, Canada (2004)

[8] Li, Y., Bontcheva, K. and Cunningham, H.: Using Uneven Margins SVM and Perceptron for Information Extraction. In: Proceedings of Ninth Conference on Computational Natural Language Learning, CoNLL-2005 (2005)

[9] Maynard, D., Peters, W. and Li, Y.: Metrics for evaluation of ontology-based information extraction. In: Proceedings of the WWW 2006 Workshop on Evaluation of Ontologies for the Web, ACM, New York (2006)

[10] Lei, Y., Uren, V. S. and Motta, E.: SemRef: An Evaluation Reference Model for Semantic Metadata. In: The 4th European Semantic Web Conference (2007)

[11] Farhat, R., Defude, B. and Jemni, M.: Authoring by reuse for SCORM like learning objects. In: The 9th IEEE International Conference on Advanced Learning Technologies, Riga, Latvia , July 14-18 (2009)