Building Domain Ontologies for Hyperlinked Multimedia Pedagogical Platforms

Zenun Kastrati, Ali Shariq Imran, and Sule Yildirim Yayilgan

Department of Computer Science and Media Technology Gjøvik University College, Norway
{zenun.kastrati,ali.imran,sule.yayilgan}@hig.no

Abstract. This paper examines building of the course ontology for describing and organizing hyperlinked pedagogical content. The ontology is used to structure and classify multimedia learning objects (MLO) in hyperlinked pedagogical platform called HIP, and to assist students to search for lectures and other teaching materials in a reasonable time and more efficiently. In addition, this paper proposes a new approach to improve the classification performance by enhancing the information representation model using concepts from the pedagogical course domain ontology. The model will automatically estimate weight of concepts within the ontology, and it will combine the weight with concepts' importance which is calculated using Term Frequency Inverse Document Frequency – *tf*idf* algorithm. This paper is a work in progress. We are in process of creating and implementing the course ontology and an experiment will be conducted to evaluate the classification performance in terms of efficiency and effectiveness for the approach proposed in this paper.

Keywords: Ontology, concept vector space, Markov chain model, HIP.

1 Introduction

Recently, the rapid developments in technology and the increasing usage of computer and other electronic devices have made other forms of education possible, which are different from the traditional ones. The distance learning is one amongst them where students can access the information without being present in class. To improve distance learning process, new means of communication and studying were introduced and new frameworks were created [1]. One such eLearning platform is Hyper Interactive Presenter – HIP [2].

HIP is a technology-rich pedagogical platform that uses combination of media elements to deliver the learning objects. The elements comprise of electronic documents such as wiki pages and PDF documents, presentations, lecture videos, an intelligent pedagogical chat bot. In addition, it also provides navigational links, tagged keywords, and frequently asked questions (FAQ).

A huge amount of information in HIP platform comes from different media modalities that need to be organized, structured and hyperlinked. Structuring and organizing such a wealth of information is labor intensive, prone to errors and a cumbersome task. Structuring information is required due to the need for classifying various

C. Stephanidis (Ed.): HCII 2014 Posters, Part II, CCIS 435, pp. 95–100, 2014.

learning objects with different content into certain predefined classes. In this regards, an automatic classification plays a key role in organizing these massive sources of unstructured information into a structured format. Therefore, we propose a new approach to automatically organize the pedagogical multimedia content using an automatic classification based on the pedagogical course domain ontology.

The course domain ontology consists of a set of concepts in the domain of teaching and associated relations. Concepts are generally expressed through natural languages and most concepts in ontology will be represented as clusters of relevant terms. Specifically, each concept in the ontology will be formed by a list of synonym terms. This ontology development process is known as bottom-up approach ontology learning [3]. This representation is important for the use of the ontology, as it will make it easier to link the concepts in the ontology to the learning objects in actual HIP platform.

Ontology based classification approach represents semantic aspects of information coming from different media modalities through entities defined within the course domain ontology. The learning object using the domain ontology is represented as a vector where the vector components represent concepts. Concepts are extracted from ontology and their importance is calculated from the corpus using statistical measures used in traditional information retrieval tf*idf.

The contribution of ontology concepts in classification process depends on the position where they are depicted in the hierarchy and this contribution is indicated by weight. The hierarchy consists of concepts such as classes, subclasses and instances that may have different weights to represent the concepts' importance.

Furthermore, this paper also proposes a new model to enhance the information representation by automatically estimating the weights of the concepts in the ontology. Thus, in addition to enhancing the representation model, we will improve the classification performance in terms of efficiency and effectiveness.

In the rest of this paper, an overview of previous research work with respect to implementation of course domain ontology on eLearning platforms is given, followed by a section which describes the new proposed approach. A short description of implementation is given in section 4. The paper will conclude with a short section describing the conclusions and future work.

2 Related Work

This section describes research that has been carried out by researchers in the area of domain ontologies for intelligent eLearning architectures and systems, and it examines the key difference between our work and other research.

Course domain ontology, as a knowledge for certain course domain, can be utilized at development time to clarify the meaning of the concepts and their properties for a specific domain of interest, particularly to aid understanding, communicating and facilitating system integration. This can be accomplished during the requirements engineering phase. Also, generating an eLearning course using course ontology could assist in defining a systematic approach in developing eLearning course [3], [4].

In the light of above, several researchers analyze and develop domain ontologies for specific topics in eLearning. Researchers in [5] conducted a research on designing and developing an ontology in an area of databases that could be used in the provision of an eLearning course. They created an ontology for database systems course and a textbook was used as the initial corpus for building this ontology.

An ontology for C Programming language course was developed and presented in [6]. The authors analyze the problems in developing an ontology for a course domain and demonstrate the applicability of their proposed approach for the development of eLearning course ontology. Since developing an ontology is an engineering process, they proposed to use the same standard which is used for developing the software project life cycle processes. Moreover, an example of how the course ontology can be utilized in developing the Algorithms Design and Analysis course was presented in [7].

Our work is different from others in terms of its usage and its application to HIP. Besides building course domain ontology, we will also use it as a means to structure and organize the semantic content in HIP pedagogical platform. The others work described uses the ontology engineering process approach for creating a vocabulary for a certain course domain.

Furthermore, our ontology will be implemented in a real pedagogical platform and it will be employed to enhance the information representation model using its concepts' importance and weight.

3 Proposed Model and Methodology

The following section gives an elaboration of the representational model proposed in this paper which is inspired from [8]. The proposed model consist of three subtasks; mapping the domain ontology into a Markov chain model, calculation of transition probability matrix for Markov chain model and calculation of information content for each concept in ontology. The final step is building a concept vector space as an information representation model.

3.1 Modeling of Markov Model by a Domain Ontology

Following the formal definition of the domain ontology, we will adopt a model where the course domain ontology will be presented as a directed acyclic graph in which concepts such as classes and their instances are structured in a hierarchy. This definition will be represented by the tuple $O = (C, H, I, type(i), rel(i))$ [8], where:

- C is a non-empty set class identifiers
- H is a set of taxonomy relationship of C
- I is a potentially empty set I of instance identifiers
- $type(i)$ is an instance type relation that maps each instance in I to a set of one or more classes in C
- $rel(i)$ is an inter-instance relation that maps each instance in I to a set of zero or more other related instances in I.

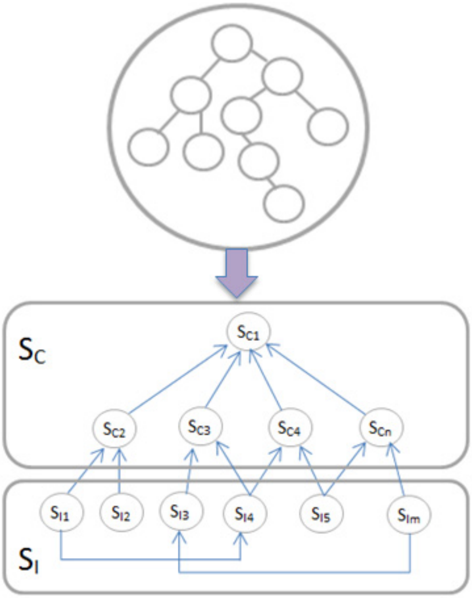

Fig. 1. Mapping of ontology into a Markov chain model

The graphical representation of the ontology will be implemented using the Markov chain model.

The Markov chain model is an equivalent mapping which means that classes in the ontology and the instances of those classes are mapped to states in the Markov chain model and all instance-to-class relations and hierarchical relations between classes are mapped to state transitions. This mapping is illustrated in Fig. 1, where the ontology is represented as a circle at the top and the Markov chain model is below. The latter is partitioned in two disjoined subsets, SC that indicates states of classes and SI indicates states of instances.

3.2 Calculation of Transition Probability Matrix and Information Content

The probability matrix for Markov chain model will be calculated based on the Page Rank algorithm [9]. We will employ this algorithm because our Markov graph owns a feature called irreducible which guarantees the convergence of the algorithm. A graph is irreducible when it is finite and, from every state it is possible to go to every other state, and the probability of transition from a state u to a state v is only dependent on the state u and not on the path to arrive at state v.

The page rank algorithm will be adjusted with two other parameters presented in [8], which are as follows:

- A probability distribution weight, ω, which determines how probabilities are distributed between states representing classes, S_C, and states representing instances, S_I.

- Weighting factor of instance *weight(i)* that maps each instance in *I* to a normalized weight between 0 and 1, and it is calculated as:

$$\mathrm{weight(i)} = \sum_{i=1}^{n} \mathrm{edge\,(i)/totalEdges}$$

Once we get the transition probability matrix, we can calculate the Information content which will give us the importance of each concept relative to other concepts.

3.3 Building the Information Representation Model

The final step of proposed model is building the information representation model using concept vector space. The information representation model will be created using the importance and the weight of concepts. Importance of concepts is calculated using *tf*idf* algorithm while weight of concepts is calculated as described in section 3.2.

4 Proposed Implementation

The main goal of using ontology is to organize MLO objects. Every new unlabeled learning object will be assigned to a predefined category in HIP. This will be done by calculating the similarity between the extracted terms from the learning object and the ontology concepts. The learning object will be assigned to a category having the highest similarity value with respect to that learning object.

5 Conclusion and Future Work

HIP is an eLearning platform that provides different types of media elements to deliver the learning objects. Structuring and organizing huge amount of learning objects comprising of multiple media elements is labor intensive, prone to errors and a cumbersome task. Therefore, the organization of the pedagogical multimedia content using an automatic classification approach based on ontology is described in this paper.

Ontology represents semantic aspects of the learning objects through entities defined within the domain ontology. Each learning object that uses ontology is represented as a vector, whose components indicate the importance of a concept. These concepts vectors are created by concepts extracted from the pedagogical course ontology and their importance and weight. Concepts' importance is calculated from the corpus by using statistical measures from information retrieval while concepts' weight is calculated using the model proposed in section 3.2. The new approach we are proposing, besides keeping information on importance of concepts, it will enrich the representation model by estimating weight of concepts automatically within the ontology.

Further research is required to implement the proposed model in order to have a reliable comparison and to evaluate the performance of the proposed model with the existing classification methods.

References

1. Imran, A.S., Cheikh, F.A.: Multimedia Learning Objects Framework for E-Learning. In: The International Conference on E-Learning and E-Technologies in Education (2012)
2. Imran, A.S., Kowalski, S.J.: HIP - A Technology-Rich and Interactive Multimedia Pedagogical Platform. Accepted in HCII (2014)
3. Happel, H.-J., Seedorf, S.: Applications of Ontologies in Software Engineering. In: Semantic Web Enabled Software Engineering, SWESE 2006 (2006)
4. De Nicola, A., Missikoff, M., Schiappelli, F.: Towards an ontological support for eLearning courses. In: Meersman, R., Tari, Z., Corsaro, A. (eds.) OTM-WS 2004. LNCS, vol. 3292, pp. 773–777. Springer, Heidelberg (2004)
5. Boyce, S., Pahl, C.: Developing Domain Ontologies for Course Content. Educational Technology & Society, 275–288 (2007)
6. Yun, H.-Y., Xu, J.-L., Wei, M.-J., Xiong, J.: Development of Domain Ontology for E-learning Course. In: IT in Medicine & Education, ITIME 2009 (2009)
7. El-Ghalayini, H.: E-Course Ontology for Developing E-Learning Courses. Developments in E-systems Engineering (2011)
8. Frost, H.R., McCray, A.T.: Markov Chain Ontology Analysis (MCOA). BMC Bioinformatics 13(14), 23 (2012)
9. Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. In: Proceedings of the Seventh International Conference on World Wide Web (1998)