Observing Knowledge Clustering for Educational Resources Using a Course Ontology

Javed Khan
Computer Science Department
233 MSB, Kent State University
Ohio, USA

javed@kent.edu

Manas Hardas
Computer Science Department
255 MSB, Kent State University
Ohio, USA

mhardas@kent.edu

ABSTRACT

In this paper we present a method to represent knowledge associated with a course and observe clustering of concepts for educational resources associated with that course. Course knowledge can be represented in the form of hierarchical prerequisite relation based weighted ontology. A novel approach for selectively processing relevant parts of the ontology is given. Design and analytical information extraction from educational resources is possible using this approach. The applicability of this method is not limited to the domain of education and can be used to analyze artifacts from any domain in which knowledge can be represented as a structured hierarchy.

Categories and Subject Descriptors: H.4.m [Information Systems Applications]: Miscellaneous

General Terms: Experimentation, Measurement.

Keywords: Course knowledge, Educational resources.

1. INTRODUCTION

There have been many organized attempts to create large digital courseware libraries to promote courseware sharing like, ACM Professional Development Center, MITs Open Course Ware (OCW) project with more than 1000 course materials freely available, Universia maintains translated versions of OCW courses in 11 languages, China Open Resources for Education (CORE) with Chinese versions of the OCW, etc. However most courseware today is not accompanied with a conceptual design framework. This makes working with educational resources very difficult. In this paper we present an approach to represent course knowledge using ontology in an expressible and computable format using
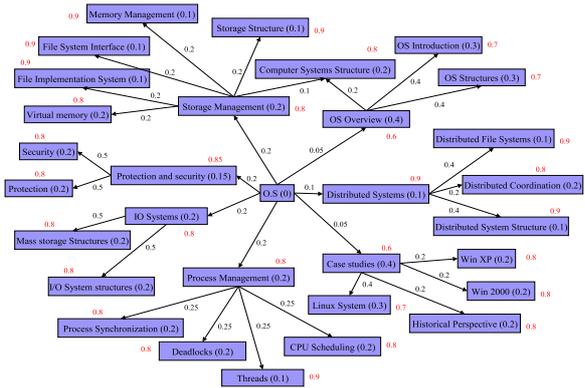
Figure 1: Partial view of "Operating Systems" course ontology 2 levels deep

has-prerequisite relationships where concepts involved in teaching a course are arranged in an hierarchical order of learning. Using a technique called *CSG extraction* [1] for specifically pointing out areas in ontologies of maximum relevance we analyse some educational resources and observe the clustering of concepts in the course ontology based on the knowledge associated with the resources.

2. COURSE ONTOLOGY

In a course ontology, the objects are *concepts* involved in the course and the relations between the concepts are that of *has-prerequisite*. The *has-prerequisite* semantics refers to the prerequisite understanding of the child node needed to understand the parent node. A node is characterized by two values namely, *self weight* and *prerequisite weight*. The self-weight is the amount of knowledge required to understand the concept in itself. To understand the concept entirely, knowledge of the prerequisite concepts is also required, which is given by the prerequisite weight of the node. Another value which characterizes the course ontology is the *link weight*. The link weight is the numerical value for the semantic importance of child concept to the parent concept. Child concepts imperative in the understanding of parent concepts will have a greater link weights than

the others. Thus the course ontology representation is a collection of concepts nodes with self weights and prerequisite weights and has-prerequisite relationships linking these nodes with a value attribute given by the link weight.

2.1 CSG extraction

A generalized CSG can be vast and there needs to be a way to efficiently process the relevant information in these ontologies to give optimum results in minimum time and complexity of computation. Therefore we define a pruned sub-graph called as *projection graph* which cuts the computation based on a limit on propagated semantic significance. *The process of selecting projection graph nodes from the concept space graph for a particular educational resource is called as CSG extraction.* This is achieved by pruning the ontology by introducing a variable called as the *threshold coefficient* (λ). By varying the threshold coefficient the size of the computable projection graph can be varied and thus the semantic significance. Threshold coefficient can be thought of as a parameter which can set the depth to which the topic has been taught. If a topic is not taught in detail, a greater coefficient is assigned so that the depth of the projection graph will be less, and vice versa.Threshold coefficient determines the limit to the quality of understanding of a particular concept.

3. OBSERVING CLUSTERING

Here we try to make interesting inferences by observing the clustering of concepts in the ontology because of the calculated projection graphs for several resources. For this purpose, we consider a specific type of courseware resource, a *test question*, which is nothing but a random question asked in a random test for a random course. Test questions were choosen as resources because concept knowledge associated can be easily identified for themselves as well as their solutions. For the experimental setup we created a course ontology comprising of around 1500 concepts, for the graduate level course of "Operating Systems" and observed the clustering of knowledge in the ontology for a set of questions composed as a part of this course curriculum. We separate out the questions with high and low average scores and observe the clustering in the ontology. Observations: 1. For concepts between 750-1000 density of questions with high scores is more than questions with low scores meaning that students understand the concepts well, or the problems based on these concepts were fairly easy to answer. 2. It is observed that in questions with low scores, concepts are more dispersed (not clustered) around the ontology as compared to those with high inverse correlation. 3. The small clustersing signifies a projection of a concept. It means that questions usually ask concepts near and around a primary concept. 4. Concepts around 200-400 and 750-1000 are frequently asked among the questions with high and low scores equally. This means that the tests were based
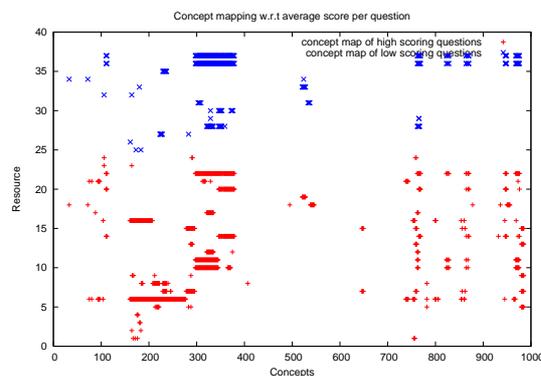
Figure 2: Concept mapping scatter plot w.r.t. average score

on those concepts. All the observations made are specific to the domain of *course knowledge* and experimental setup for *test questions*. The observations and inferences will change in case of different domains. We believe that using information from "artifacts" from a particular domain, like "test questions" in the domain of education, a great deal of information like observed clustering of knowledge, can be extracted. We present an approach to enable making such observations and inferences about the clustering and behaviour of knowledge in different domains.

4. CONCLUSION

We propose a technique for representing hierarchical structured knowledge using weighted ontologies and demonstrate it in the domain of courses (education). A technique called *CSG extraction* is given to select relevant information from course ontology depending upon the desired semantic significance. Using this approach we observe clustering of knowledge for educational resoures in the course ontlogy As future work we are trying to access the ways in which this method of representation and extraction can be applied to classical learning theories which require knowledge to be represented as prerequisite concept structures.

5. REFERENCES

[1] Javed Khan and Manas S. Hardas, A Technique for Representing Course Knowledge Using Ontologies and Assessing Test Problems, pp. 174-179, 5th Atlantic Web Intelligence Conference 2007, June 25-27, 2007 - Fontainebleau, France.

[2] Trent Apted, Judy Kay, "Automatic Construction of Learning Ontologies," icce, p. 1563, 2002 International Conference on Computers in Education (ICCE'02), 2002.

[3] Judy Kay, Sam Holden, "Automatic Extraction of Ontologies from Teaching Document Metadata," icce, p. 1555, 2002 International Conference on Computers in Education (ICCE'02), 2002