Annotating the domain ontology of a course with its syllabus and learning objectives

Rekha Ramesh
Department of educational Technology,
Indian Institute of Technology Bombay,
Mumbai, India
e-mail: rekha.ramesh@iitb.ac.in

Sasikumar M.
CDAC Bombay,
Mumbai, India
e-mail: the.little.sasi@gmail.com

Sridhar Iyer
Department of CSE,
Indian Institute of Technology Bombay,
Mumbai, India
e-mail: sri@iitb.ac.in

***Abstract*—In this paper, we discussed an approach towards integrating both the contents and cognitive level information extracted from LOs of a course into the domain ontology. We call this as LO annotated ontology (LAO). This can form a first step towards building an automated system to measure the alignment of assessment instrument (AI) to course LOs. The effectiveness of this approach can be tested by comparing the manually generated results by the experienced teachers to the system generated results.**

Keywords— Learning Objectives; Syllabus, Alignment; Ontology based knowledge Representation

I. Introduction

An assessment instrument (AI) should be properly aligned with the learning objectives (LOs) of the course [1][2]. Today, teachers have to spend a lot of time and effort in manually ensuring this alignment. So, there is a need for an automated system to measure the alignment of an AI of a course to the set of LOs of that course. In order to build such a system, we need to capture the relevant knowledge from syllabus and LOs and map it into a knowledge representation which is in a machine parsable form. Ontology is one such mechanism [5][6]. This paper discusses such an approach and generates an LO annotated ontology (LAO). When the knowledge from items in AI are also extracted and annotated into such an ontology, it can form as a representation mechanism for building an automated system to measure the alignment of AI to course LOs. Our proposed approach includes the process of extracting concepts and cognitive level from an LO using NLP techniques and the complex process of mapping these to the nodes of the ontology. Color coding is introduced to capture the result of the mapping.

II. The Design of our proposed approach

The overall process of generating an LAO is shown in Fig. 1. The main component of the system is the LO annotator which takes domain ontology, syllabus and LOs as input and outputs LAO. The domain ontology contains all the concepts related to a particular domain and relationship between them [4]. For example, for the domain of Data Structures, it will contain concepts relating to data structures including various known data structures, their representation and applications and operations on them [4]. Fig. 2 shows part of domain ontology. Every node in the ontology represents a concept/topic from the domain. The dependencies/ relationships between the concepts are shown using links. In our ontology, we are assuming the links such as 'hasSubClass', "hasRepresentation", "hasOperation", "isA" "hasApplication" and "includes". The links are used to traverse the ontology to locate the neighborhood nodes which are relevant in the ontology. The type of links decides what nodes are to be included for mapping.

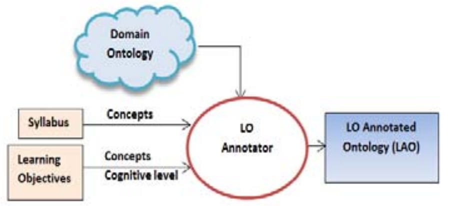

Fig. 1. The overall process of generating LAO

Every university can have their own syllabus and LOs which can be viewed as a subset of domain ontology. This subset can be indicated in domain ontology using some color coding. This is called (LAO). The annotator assumes that in initial ontology all the nodes are colored as white. When the syllabus is mapped to it, the matching nodes will be colored as black and when LOs are loaded, the matching nodes will be partially colored as red. Different cognitive levels will be indicated by varying shades of red. The shade/intensity of the red color is dependent on the cognitive level of LO involving those concepts. Higher the level, darker is the shade of the color. Fig. 2 shows how the LAO will look after coloring all the relevant nodes.

III. LO ANNOTATOR

The annotator takes the input from the syllabus and LOs and annotates the domain ontology as described above. Every course consists of a set of predefined LOs (l_1, l_2, l_3,.. l_n) covering the entire syllabus. Every LO contains 2 attributes: a

978-1-5090-2504-6/16 $31.00 © 2016 IEEE
DOI 10.1109/LaTiCE.2016.45

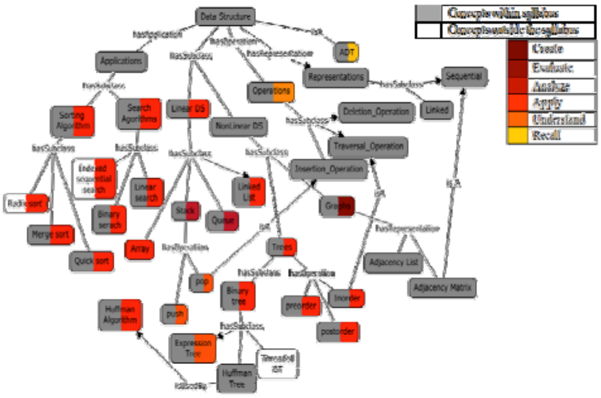

Fig 6. The Domain ontology with the example LOs mapped

set of topics/concepts (c_1, c_2, ..,c_l) from the syllabus addressed by that LO and cognitive level defined by Bloom's Taxonomy[4]. The key design element is how to automatically extract relevant information (concepts and cognitive level) from the LO text. This requires some amount of NLP techniques such as tokenization, lemmatization, POS Tagging, etc. Basically, each LO text can be first parsed to separate into a set of words or tokens.

Set of words or tokens from each LO is matched with the node names from the ontology. If the match is found the corresponding nodes are marked in the domain ontology. But the matching process is not direct as concepts may not be explicitly available in LOs. They may be multi-worded or differently worded which can be to some extent handled by annotating the nodes with synonyms. Implicit or hidden concepts will have to be identified by devising an ontology traversal algorithm. In such cases, the following decisions need to be taken by the system as it processes each LO i) Is there a need to find the related nodes? ii) What links to traverse in the ontology? iii) What depth it should be traversed? For example,

LO: Students should be able to demonstrate and implement different methods for traversing trees.

Here, ***traversal operation*** (with synonym 'methods for traversing') and ***trees*** form the explicit nodes and other implicit nodes to be identified are ***preorder***, ***inorder*** and ***postorder*** traversals.

Blooms taxonomy forms the basis for cognitive level identification of an LO. Every level of Blooms taxonomy namely, Recall, Understand, Apply, Analyze, Evaluate and Create is associated with an elaborate set of keywords. These keywords can be stored into a dictionary. The tokens are matched to the keywords in the dictionary and accordingly its cognitive level is identified. If two tokens match with the keywords of two different Blooms level, then the higher level one is chosen as cognitive level of the complete question. The cognitive level of above LO is '***Apply***' as it contains the action verbs '*demonstrate*' and '*implement*' at '*understand*' and '*implement*' respectively. .Fig. 6 shows how the LAO will look after coloring all the relevant nodes.

IV. TESTING

The testing proposed primarily is to check whether the annotator is annotating correctly by giving the right color and right shade of color to the nodes. What is the right color will be decided by giving the domain ontology, syllabus and LOs to the expert teacher and telling them to manually create LAO. The teachers who have teaching experience of more than 5 years and have thorough domain knowledge are considered as expert teachers. The teacher generated LAO will be compared with the system generated LAO in terms of both concepts and cognitive levels. A confusion matrix will be generated which will classify total number of concepts from all LOs into 4 classes: the number of concepts in which (i) both the teacher and system have agreed that the concepts are covered by an LO (True positives). So they color the nodes (ii) both the teacher and system have disagreed that the concepts are covered by an LO (True Negatives). So both did not color the nodes (iii) only the system has agreed to color but teacher did not color (False Positives) and (iv) only teacher has agreed to color but system did not color (False Negatives). Each LO will be further analyzed to determine which ones have contributed to False Positives and False Negatives and what are the characteristics of such LOs. Sometimes the problem may be because of the framing the LOs in a particular way.

The mismatch in cognitive level may be because of the inherent ambiguity at adjacent levels in Bloom's taxonomy. Further analysis can be done to find the degree of mismatch i.e. in how many levels of difference is there between the teacher and system generated results.

V. CONCLUSION AND FUTURE WORK

The problem being focused is integrating both the contents and cognitive level information extracted from LOs of a course into the domain ontology. This can form a first step towards building an automated system to measure the alignment of AI to course LOs. This paper brings forth the need of such a system and proposes an approach to implementing such a system. In this paper only the part up to capturing the syllabus and LOs is discussed. Our proposed approach explains the process of extracting concepts and cognitive level from an LO using NLP techniques and the complex process of mapping these to the nodes of the ontology. Implementation of the system is in progress.

REFERENCES

[1] A. J. Nitko, *Educational assessment of students*. Des Moines, IA 50336-1071, PO Box 11071:Prentice-Hall, Inc., 2001

[2] J. Biggs, "Aligning teaching and assessing to course objectives," Teaching and Learning in Higher Education: New Trends and Innovations, 2, pp.13-17, 2003.

[3] D. R. Krathwohl, " A revision of Bloom's taxonomy: An overview. Theory into practice," Vol. 4 41, pp. 212-218, 2002.

[4] Natalya F. Noy and Deborah L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," Stanford University, 2001.

[5] R. Mizoguchi and J. Bourdeau, "Using Ontological Engineering to Overcome AI-ED Problems: Contribution, Impact and Perspectives, " *International Journal of Artificial Intelligence in Education*, 1-16, 2015.

[6] H. Chung and J. Kim, "An Ontological Approach for Semantic Modeling of Curriculum and Syllabus in Higher Education," IJIET, Vol. 6, No. 5, May 2016.