Semi-automatic construction of ontology based on data mining technique

Jingyun Wang  
Research institute for information technology  
Kyushu University  
Fukuoka, Japan  
warmplam@gmail.com

Brendan Flanagan  
Academic Center for Computing and Media Studies  
Kyoto University  
Kyoto, Japan  
bflanagan.academic@gmail.com

Hiroaki Ogata  
Academic Center for Computing and Media Studies  
Kyoto University  
Kyoto, Japan  
hiroaki.ogata@gmail.com

Abstract— In this paper, we present a method to extract the possible relationships between knowledge points by analyzing e-book log and mining quiz data and mining Wikipedia articles. This method will be implemented in an ontology-based visualization support system to support the instructor to construct course-centered ontologies semi-automatically.

Keywords— Ontology; learning data analytics; data mining; association rules

I. INTRODUCTION

The potential benefits of maps, which have nodes as key concepts and links as relationships between key concepts [1], have been well documented in academic literature. Since the working memory's processing capacity limits the transformation of unrelated concepts into long-term memory, knowledge gets incorporated into the human brain more effectively when it is organized into hierarchical frameworks. Learning approaches that facilitate this kind of organization significantly increase the learning capability of learners [2] [3]. Ausubel's learning psychology [4] [5] define this effective assimilation of new knowledge into existing knowledge framework as the achievement of “meaningful learning”. Therefore, how to help learners to efficiently develop their conceptual framework becomes the main issue for fostering meaningful learning in the e-learning field.

From this point of view, maps can serve as a kind of scaffold to help learners to organize knowledge and structure their own knowledge framework [6]; this facilitates meaningful learning. Organizing knowledge concepts in a map structure, e-learning systems can present/provide progressively more explicit knowledge to help learners to slowly develop conceptual frameworks; learners also can clearly understand large general concepts before learning more specific concepts and incorporate new knowledge into their prior knowledge frameworks to foster meaningful learning. In addition, when learners have different levels of prior domain knowledge, using maps they can jump directly to a specific chapter that they are interested in. Although a search engine in the LMS can also be used to look for the information on a certain concept, the inquiry results without relation information between KPs limit its usage.

Ontology is one of the main techniques which are adopted in maps for knowledge representation. “An ontology is a formal explicit specification of a shared conceptualization” [7]. Common vocabularies are defined by ontology for the users (such as instructors, learners and researchers) who need to share information in a domain [8]. E-learning systems using maps to support learning activities, such as the concept map learning system of Chu et al [9], intended to help reduce the user's cognitive load, TM4L[10], a specialized environment for creating, maintaining and using “TM-based” learning repositories, mostly depend on ontology-based engines, or CLSS [11], the customizable language learning support system which manipulates a course-centered ontology to provide an interface displaying the visual representation of knowledge points and their relations for the learning objects arrangement. However, the construction and maintenance of ontology is quite time consuming. In this paper we present a method for constructing ontology semi-automatically.

II. AN ONTOLOGY-BASED VISUALIZATION SUPPORT SYSTEM

Nowadays, e-book systems are widely used in education field. In Japan, the education ministry panel is urging schools (including K-12 and higher education) to use digital textbooks for supporting daily classroom teaching from 2020. Since 2014, Kyushu university of Japan started to use e-book systems, together with learning management systems Moodle and e-profile system Mahara, for supporting daily classroom teaching. E-book systems can provide a platform for instructors to easily upload the teaching materials, and learners can conveniently view those files and even make markers or put comments on the files. These systems can also record a learners’ learning activities and facilitate reporting to instructors. However, in e-book systems it is also difficult for the learners to identify the knowledge they possess before and after a learning activity. Furthermore, existing e-book systems (even other e-learning systems) cannot encourage learners to
compare new knowledge with the relevant acquired knowledge, and cannot effectively support the construction of their knowledge structure.

Therefore, an ontology-based visualization support system (VSSE) has been developed to provide a meaningful learning environment to help e-book learners to effectively construct their knowledge frameworks [12][13]. Four main functions are provided in VSSE. (1) After setting the time period (for example, from Jan. 3rd 2016 to Jan.4th 2016), the learner will get a concept map that shows the knowledge points involved in the pages that she/he read during that time period and the relevant knowledge points in different colors. (2) Users can open all the concepts in a tree structure until reaching the knowledge point they are seeking. Each knowledge point's location in the -Book system and its related knowledge will be shown in a concept map. (3) After setting one e-book and its page range, the learner will get a concept map that shows the knowledge points in those pages, their related knowledge (in different colors), and the relation between them. (4) Instead of showing directly all the information of a concept map as a function (1-3), the system provide a discovery learning environment (called “cache-cache comparison” mode) that enable e-book users to seek or create hidden relations or concepts in a concept map based on their log data. This design is intended to encourage learners to actively locate new knowledge in their knowledge framework and check the logical consistency of their ideas.

To facilitate the visualization supports of meaningful learning, the descriptions of the information about all the knowledge points and their relations are required for VSSE. In this research, a knowledge point (KP) is defined as “a minimum learning item which can independently describe the information of one certain piece of knowledge in a specific course”; a learner can understand a KP by its own expression or can acquire it by practice. Those domain knowledge needed by the system is suggested to be automatically extracted from an ontology, which is designed and developed based on the e-books.

For a mock up demo, firstly we adjust and apply the ontology design method described by Wang et al. [11] to develop a course-centered ontology of an existing computer science course (called COCS). COCS includes about 100 KPs and 20 kinds of relations extracted and defined based on the analysis of all the e-books of this computer science course. Figure 1 illustrates some KPs and their relations in COCS.

For the construction of COCS, the instructors of the course were required to record KPs and the relation between KPs in the computer science course manually. For the Kikan Education in the Arts and Science Department, there are 3730 courses registered. Recording only KPs in a course is not difficult for the instructor. However, preparing all the relations between KPs in the same course and cross two courses is quite time-consuming. Therefore, we present a method to support the record of the relations using data mining techniques.
III. A PROCESSING MECHANISM FOR SEMI-AUTOMATICALLY CREATING AN ONTOLOGY

Firstly, instructors of a course still need to describe all the KPs based on E-books of that course. Then a tool developed by authors, which can automatically identify the location (including the file ID and the page number) of the KPs in the E-Book system, will be used to add the location information details into the ontology automatically.

Secondly, the user logs of E-book system will be analyzed to extract the possible relations between KPs. The E-book logs were collected from students of Kyushu University who attended the course during the past years (since 2015). Each access log contains the student's ID, the section of the course, the page number of the slides, the action which the student took, access date, access time and the duration of the reading time of the page. There are 20 kinds of action, such as open, close, next, previous, zoom and jump. Based on the analysis of the log data, we expect to find out which pages have KPs that are highly related.

In Fig.2 we show an example of the possible input and output of our proposed mechanism. The left-hand side shows a graph of all the e-book logs that represent the use of the e-book system by learners. The pages of the e-book are shown as circles, which may contain KPs that are indicated by triangles. A learner's interaction between the pages of the e-book are shown as directed edges of the graph. The right-hand side is a representation of the KP relations that have been extracted by analyzing the e-book logs with a processing mechanism based on an algorithm that was proposed in [14]. In this representation, KPs are indicated by triangles in the lower half and round rectangles indicate upper concepts of the ontology.
We propose applying the minimum spanning tree algorithm to the complete graph representation of possible relations to extract the relations of the KPs in the ontology. The algorithm searches for a tree in the complete graph that contains the strongest possible relations between KPs. It achieves this by greedily selecting the strongest links that have at least one node that is not connected to a previously selected edge. The process of transforming a complete graph representation of the possible relations to a minimal tree of KPs is shown in Fig. 3. In step A, a complete graph can be created from the interactions between the pages of the e-books and the KPs that are contained within those pages. The directed weighted edges of the complete graph represent the number of interactions between the pages in which the KPs are contained. Overlapping edges of the graph are then merged and the weight of the resulting edge is the sum of the merged edge weights as shown in step B. The weights of the edges are then inversed in steps C so the minimum spanning tree algorithm can be applied directly to calculate the graph as shown in steps D. This effectively acts as a threshold that would automatically remove very weak links between KPs, while ensuring that the resulting graph is a tree and doesn't contain any cycles or loops between KPs. Then we propose that clustering analysis of the KPs can be used to find the upper concepts of the ontology. However, as clustering does not automatically label the nodes of the cluster tree, the instructors need to modify the names of the upper concepts. Further investigation is required for the automatic naming of upper concepts.

Thirdly, we will also mine the quizzes contents of Moodle system and the content of Wikipedia articles including the hyperlink to infer the relations especially the prerequisite relationship between KPs. Wikipedia has been used as a knowledge source to infer semantic relationship by Several studies [15] [16] [17]. Moreover, the quizzes contents including the reference answers in Moodle will also analyzed. KPs, which shows together in one question item or the answer of one question, could be considered highly related. Methods, which consider mining the association rules from historical testing records, have been presented by serval researches [18] [19] [20]. Finally, based on the analysis result of the previous two step, each KP will have a list of possible related KPs. When the instructor click one KP from the tree structure as shown in Fig.2, the system will show the location information of this KP and its possible related KPs with relation names. The instructor will be enable to delete the unrelated KPs from the list or further modify relation name of a related KP.
IV. CONCLUSION

In this paper, we present a method to extract the possible relationships between KPs of course by analyzing e-book log and mining quizzes data and Wikipedia articles. In future work, we will implement this function in VSSE and use it to support the construction of course-centered ontologies for daily teaching. Its effectiveness will be evaluated based on the rate of possible KPs which are deleted by instructors and the rate of possible relation name which are modified by instructors.

ACKNOWLEDGEMENTS

The research are supported by KAKENHI Grant Number 17K17936, the Research and Development on Fundamental Utilization Technologies for Social Big Data (No. 178A03), and the Commissioned Research of National Institute of Information and Communications Technology, Japan. This work was also supported by JSPS KAKENHI Grant Number 16H06304.

REFERENCES