Yang, J. C. et al. (Eds.) (2018). Proceedings of the 26th International Conference on Computers in Education. Philippines: Asia-Pacific Society for Computers in Education

Automatic Generation of Contents Models for Digital Learning Materials

Brendan FLANAGAN[a*], Gökhan AKÇAPINAR[a, b], Rwitajit MAJUMDAR[a] & Hiroaki OGATA[a]
[a]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
[b]*Department of Computer Education & Instructional Technology, Hacettepe University, Turkey*
*flanagan.brendanjohn.4n@kyoto-u.ac.jp

Abstract: There has been much research that demonstrates the effectiveness of using ontology to support the construction of knowledge during the learning process. However, the widespread adoption in classrooms of such methods are impeded by the amount of time and effort that is required to create and maintain an ontology by a domain expert. In this paper, we propose a method to automatically generate a contents model by analyzing learning materials with the aim of supporting the construction of knowledge structures. A map of the keyword nodes is constructed by applying text mining techniques to find the important words and phrases and their relations contained within the learning materials. The process retains links between the nodes and the original learning materials, and it is therefore possible to recommend and rank sections that cover a concept contained within the contents model map.

Keywords: contents model, text mining, learner knowledge

1. Introduction

It has been well documented that learners can benefit from the use of maps to represent the key concepts of knowledge (Lee & Segev, 2012). However, the process of creating and maintaining these maps often involves a domain expert manually creating the map based on their experience and previous knowledge (Wang, Flanagan, Ogata, 2017). To overcome these problems, we propose a method that can automatically extract a simple contents model based on the structure of keywords that occur in digital learning materials with minimal time and effort. The proposed method will be integrated into the upload process of the e-book reader, BookRoll, which is a part of a Learning Analytics platform currently in use at Kyoto University (Flanagan & Ogata, 2017). The BookRoll system logs the reading behaviors of learners that are viewing learning materials on the system. By generating contents models of learning materials and analyzing the reading behaviors of students, it is anticipated that a model of the knowledge acquired by the student can be estimated.

There are many researches into the generation and use of ontologies, concept maps, and mind maps in education to show and create knowledge frameworks. Association rules and other data mining techniques have been used to construct concept maps based on the results of test and quizzes to show the relation between knowledge that was tested (Hwang, 2003; Tseng, Sue, Su, Weng, Tsai, 2007; Chen, Bai, 2010; Chen, Sue, 2013). While this technique is applicable to the structured format of tests, it is difficult to apply similar techniques to unstructured contents that is contained in e-books.

In previous work by one of the authors, a similar algorithm that is extended in the present paper was used to generate mind maps. In Flanagan et. al. (2013), mind maps were automatically constructed from a user's twitter texts for assisting communication between two or more foreign language learners in mutual language exchange settings.

Wang et al. (2014) investigated and evaluated the use of a manually constructed course-centered ontology to support language learning. They created a system to show and manipulate an ontology through a visual representation of key knowledge points. Contents models created by the method proposed in this paper could be manipulated and viewed using such a system so students can further reinforce the understanding of knowledge learnt while reading the learning material.

2. Automatic Contents Model Construction

A contents model can be thought of as a graph of key points that are contained within the text contents of the learning material that it represents. The relation between nodes of this graph are expressed as a weighted edge representing the strength of the relation between two key points that are in the contents.

A full representation of such relations is a co-occurrence graph of words from the contents. Co-occurrence boundaries may take on different forms, such as co-occurrence within: separate learning materials in a course, chapters, sections, pages and sentences, which we will refer to as a document in the remainder of this paper.

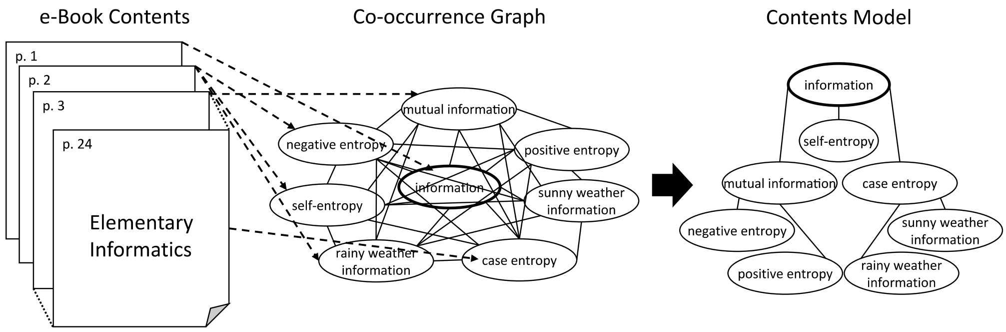

Figure 1. An overview of the proposed contents model generation process.

To reduce the complexity of the graph to the relevant key points, we search for an optimal spanning tree that only selects the strongest relations between nodes, as shown in the overview in Figure 1. This method is an extension of a method previously proposed by the authors (Flanagan et al., 2013).

The proposed method has been implemented to generate contents models with minimal time and effort by the teacher. It is fast enough to generate the contents model as a part of the upload process of e-books to the BookRoll system. Integration with existing ontology-based systems for visualizing a students' knowledge framework, such as the system proposed in Wang et al. (2014) could also be used for the implementation and evaluation of the effectiveness of contents models in classroom settings.

3. Case Study: Elementary Informatics Lecture Contents Model

In this section, we introduce a contents model that was generated by analyzing the slides from a presentation that were used in a lecture on Elementary Informatics. The lecture slides were written in Japanese and consist of 24 pages in total. The slides had been uploaded to BookRoll as a PDF file, and pdfminer[1] was used to extract the contents text and location of 219 bounding boxes, however the location information was not analyzed in this paper. The extracted Japanese text was then parsed using MeCab[2] which performs morphological analysis to separate the individual words and their parts-of-speech from a sentence. The text was then preprocessed to generate n-grams of length from uni-grams to 4-grams using only nouns, and resulted in 883 tokens that were indexed using the GETAssoc[3] search engine.

The contents model generated from the sample lecture slides is shown in Figure 3. During initial analysis, it was found that many words of low frequency were included in the contents model, so the number of nodes was restricted based on frequency threshold of 10 or more occurrences. It can be seen that the key concepts of "mutual information" and "case entropy" are captured to a degree. In the instance of "case entropy" the child nodes "rainy weather information" and "sunny weather information" are references to examples that were used to explain the concept of "case

---

[1] https://euske.github.io/pdfminer/
[2] http://taku910.github.io/mecab/
[3] http://getassoc.cs.nii.ac.jp

entropy”. The use of a mask based on dictionaries or thesauruses might help to narrow the selection of concept nodes to domain relevant concepts.

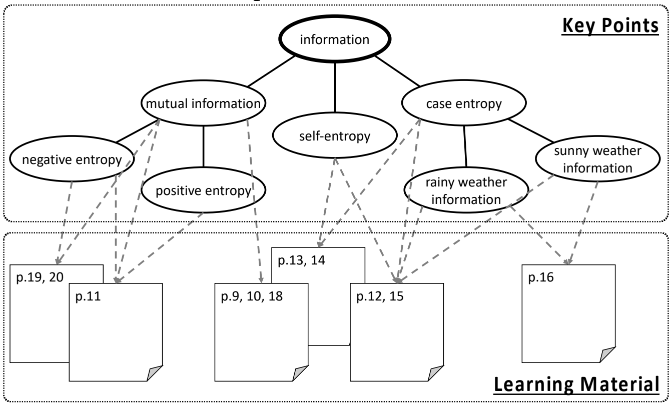

Figure 3. Example contents model generated from the analysis of slides from an Elementary Informatics lecture.

4. Conclusion

In this paper, we propose the automatic generation of contents models from the analysis of uploaded digital learning materials to an e-book reader. The proposed method was used to construct a contents model of the slides from a lecture on Elementary Informatics that was taught in Japanese.

There are some limitations of the proposed method that should be mentioned: the generation of the graph is dependent on the input text and therefore may create off topic nodes from examples, and the detail level of the output contents model requires supervision from the domain expert. In future work, we plan to apply the masking of concepts based on existing ontologies and thesauruses. The construction of the model based on the structure of existing knowledge models should also be investigated.

Acknowledgements

This research was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (S) Grant Number 16H06304.

References

Chen, S. M., & Bai, S. M. (2010). Using data mining techniques to automatically construct concept maps for adaptive learning systems. *Expert Systems with Applications*, *37*(6), 4496–4503.

Chen, S. M., & Sue, P. J. (2013). Constructing concept maps for adaptive learning systems based on data mining techniques. *Expert Systems with Applications*, *40*(7), 2746–2755.

Flanagan, B., Yin, C., Inokuchi, Y., & Hirokawa, S. (2013). Supporting interpersonal communication using mind maps. *The Journal of Information and Systems in Education*, *12*(1), 13–18.

Flanagan, B., & Ogata, H. (2017). Integration of Learning Analytics Research and Production Systems While Protecting Privacy, *Proceedings of the 25th International Conference on Computers in Education* (ICCE2017), 333–338.

Hwang, G. J. (2003). A conceptual map model for developing intelligent tutoring systems. *Computers & Education*, *40*(3), 217–235.

Lee, J. H., & Segev, A. (2012). Knowledge maps for e-learning. *Computers & Education*, *59*(2), 353–364.

Salton, G., & McGill, M. (1983). Introduction to modern information retrieval. McGraw-Hill, New York (1983).

Tseng, S. S., Sue, P. C., Su, J. M., Weng, J. F., & Tsai, W. N. (2007). A new approach for constructing the concept map. *Computers & Education*, *49*(3), 691–707.

Wang, J., Mendori, T., & Xiong, J. (2014). A language learning support system using course-centered ontology and its evaluation. *Computers & Education*, *78*, 278–293.

Wang, J., Flanagan, B., & Ogata, H. (2017). Semi-automatic construction of ontology based on data mining technique. In *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 511–515). IEEE.