Learning Support System for Providing Page-wise Recommendation in e-Textbooks

Keita Nakayama
Kyushu University, Japan
nakayama@limu.ait.kyushu-u.ac.jp

Masanori Yamada
Kyushu University, Japan
mark@mark-lab.net

Atsushi Shimada
Kyushu University, Japan
atsushi@ait.kyushu-u.ac.jp

Tsubasa Minematsu
Kyushu University, Japan
minematsu@limu.ait.kyushu-u.ac.jp

Rin-ichiro Taniguchi
Kyushu University, Japan
rin@kyudai.jp

Abstract: Thanks to the increase in the amount of information on the Internet and the spread of ICT-supported educational environments, much attention has been paid to learning support based on smart recommendation technologies. In this research, we support learning by recommending detailed information about each page of e-textbooks. We developed a recommender system that extracts important keywords on each page of an e-textbook and retrieves related websites to support students' understanding of lecture contents. In this paper, we explain the details of the recommender system and report the experimental results.

Introduction

The increasing amount of information on the Internet, that is, the enormous quantity of open knowledge objects such as Wikipedia, iTuneU, and repositories managed by universities all over the world seems to be useful to support learning. To use the information effectively and efficiently, it is necessary to build a system for the appropriate use of this vast knowledge. The recommender system for learning support is a common and useful system in this era of vast information (e.g., Erdt et al., 2015). Learners must find suitable and useful knowledge online (e.g., news articles, academic papers, and blogs) from a mass of search results. Thus, a recommender system provides an appropriate learning environment that helps learners find appropriate knowledge objects within the vast amount of information on the Internet. Researchers have been applying these resources to learning systems to clarify the effects of recommender systems for learning (e.g., Liang et al., 2006; Yamada et al., 2014). Chen (2007) developed and evaluated a recommender system integrated with an e-Book on a learning system. The recommender system recommended references from knowledge repositories such as dictionaries and discussion forums, based on learning inquiries in a web-based learning system. The results indicated that students tend to use recommender systems to prepare for examinations.

As a learning support system for classroom settings, recommendation of information seems to be effective for the enhancement of important concepts taught in class. There are several studies that recommend related materials according to the student's learning situation (e.g., Chen et al., 2014; Tarus et al., 2018; Wan & Niu, 2018), the student's knowledge gaps in class (e.g., Bauman & Tuzhilin, 2018), or the student's query (e.g., Mbipom et al., 2018). These approaches basically recommend some learning materials which contain the student's demand as a topic in the material. On the other hand, it is also important to recommend extra materials for deep understanding and/or extension of the individual topics, automatically. To provide students with detailed support, it is desirable to recommend not only material-wise but also page-wise information, i.e., information that aids in the understanding of lecture contents and/or in the extension of topics on each page. This research aims to develop a recommender system integrated in the e-Book viewer. The system extracts important keywords of each page in an e-textbook and explores related websites. In this paper, we explain the details of the recommender system and report the experimental results.

Recommender system

System overview

The purpose of this research is to develop a system to support learning by recommending supplementary teaching materials (STMs) as detailed information about each page of e-textbooks used in a lecture. The overview of the system configuration is shown in (Fig. 1). The system consists of an e-Book system, three databases, and a recommender system. In each database, e-textbooks used in lectures, recommendation information of STMs, and e-Book event logs are stored. The system flow is described as follows. First, teachers register e-textbooks in the database via the e-Book system. Next, the recommender system analyzes the e-textbooks and identifies STMs corresponding to each page. The recommendation information is stored in the database. A teacher conducts a lecture using the e-Book system. Students open the e-textbook during the lecture and access the recommended STM information as necessary. (Fig. 2) shows the user interface of the e-Book viewer. Students can access the recommended STMs by clicking the upper right button of the viewer. When a student uses any of the functions of the e-Book system (such as *open an e-textbook*, *go to the next page*, *highlight*, *click a link to an STM*, etc.), e-Book event logs are automatically recorded in the database.

To acquire STM information for each page of the e-textbook, we take three analytics steps, as follows.

1. Word extraction from e-textbook;
2. Calculation of importance of extracted words;
3. Determination of STMs based on important words.

The details are explained in the following sections.

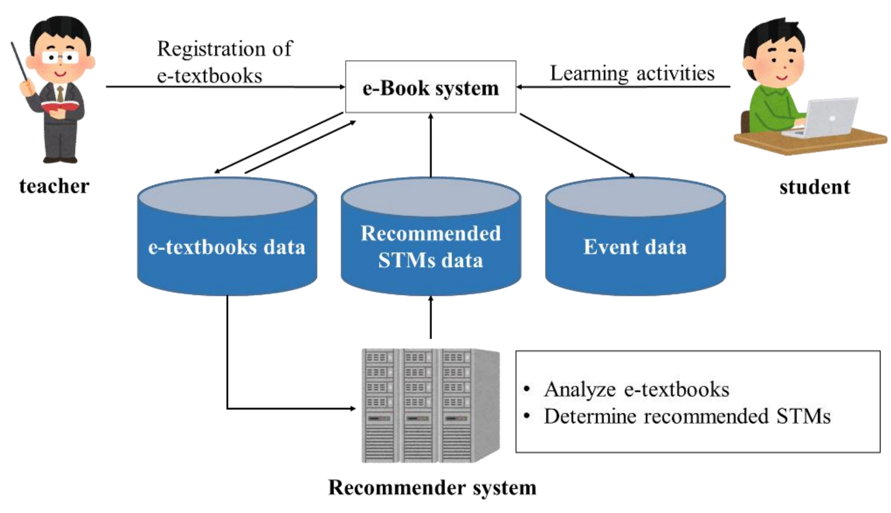

Figure 1: Overview of system configuration

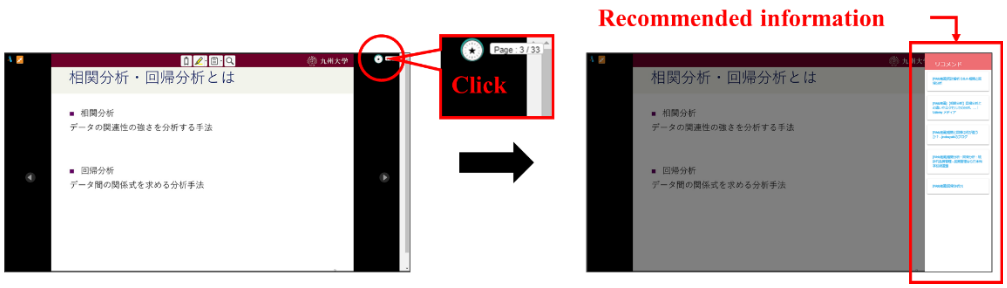

Figure 2: E-textbook viewing screen on the e-Book system (when the student clicks the upper right button of the viewer, the recommended information is displayed.)

Word extraction from e-textbook

To recommend STMs according to the contents of e-textbooks, we first extract nouns in the e-textbooks. In the developed system, e-textbooks registered by teachers are stored as PDFs, and text information about sentences in registered e-textbooks is extracted from the PDFs automatically. Therefore, we can utilize the text information to extract sentences and words from e-textbooks. In this research, we focus on Japanese e-textbooks. In contrast to languages segmented by spaces such as English, Japanese is a non-segmented language. Therefore, it is necessary to divide the sentences to extract words. In this research, we apply MeCab morphological analysis (Kudo et al., 2004). Morphological analysis is the technique of dividing natural language into morphemes, the smallest units of words. When extracting a noun from a sentence through morphological analysis, the sentence is sometimes over-split, so that words composed of plural nouns cannot be correctly extracted. For example, words like "デジタル画像処理" (digital image processing) are divided into three words through morphological analysis: "デジタル" (digital), "画像" (image), and "処理" (processing). However, it is thought that students who see the term "デジタル画像処理" (digital image processing) recognize it as having one meaning, not three. For this reason, words are extracted by performing the following modification on the morphological analysis result.

- Continuous nouns are extracted as one compound noun;
- When nouns in English and Japanese are consecutive, they are extracted as compound nouns;
- Do not extract words consisting only of Hiragana;
- Do not extract words consisting only of numeric characters;
- Do not extract words consisting of one character.

(Fig. 3) shows an example of the extraction results.

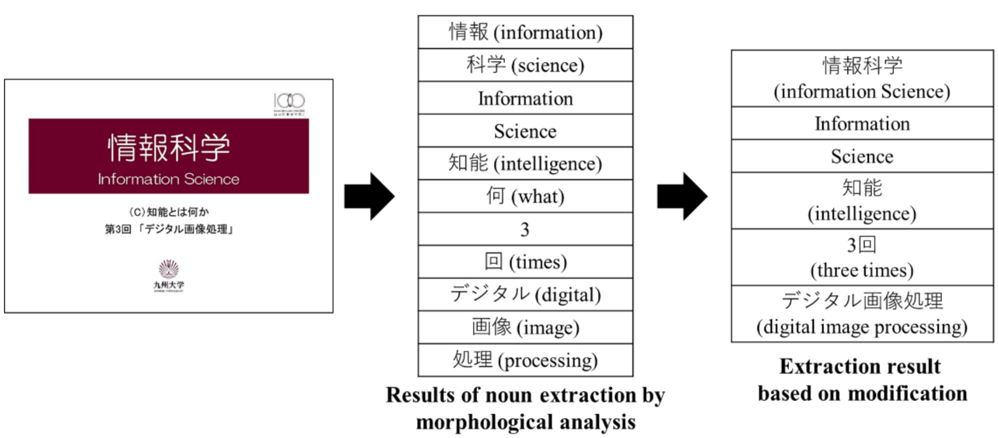

Figure 3: Word extraction result

Calculation of importance of extracted words

In this research, we recommend STMs corresponding to each page of an e-textbook. Therefore, to decide the STMs, we estimate the importance of the extracted words on each page. (Fig. 4) shows an overview of the process flow to estimate the importance of extracted words. The estimation procedure involves two steps: 1) dividing each page of an e-textbook into a subset of pages (we refer to them as "segments" in the following subsections) in terms of similar topics, and 2) calculating the importance of the words on each page within each segment. The detailed procedures are explained in the following subsections.

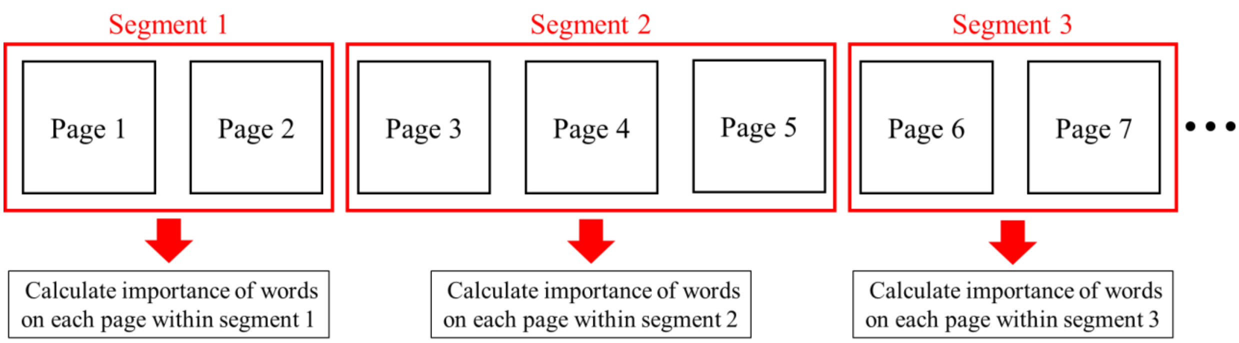

Figure 4: Overview of calculating importance of extracted words

Page grouping based on topic

Since an e-textbook is generally composed of multiple topics, it may be considered that there is a difference in importance depending on the topic, even for the same word. Therefore, in this research, whole pages of an e-textbook are divided into subsets of pages (segments) according to the topic of the contents, and the importance of each word is calculated for each segment. The procedure for obtaining segments is as follows.

1. From the extraction result, the recommender system obtains a set W of words included in the e-textbook and the appearance number N_w of each word $w \in W$ in the e-textbook.
2. Based on the words on each page of the e-textbook, a feature vector $\boldsymbol{c}_i = (c_{i,w_1}, \ldots, c_{i,w_l}, \ldots, c_{i,w_n})$ of page i is generated. When there is a word w_l on page i, the value of element c_{i,w_l} is N_{w_l}; otherwise, it is 0.
3. The cosine similarity of feature vectors between adjacent page i and j is calculated using (Eq. 1).

$$sim(\boldsymbol{c}_i, \boldsymbol{c}_j) = \frac{\sum_{k=1}^{n} c_{i,w_k} \cdot c_{j,w_k}}{\sqrt{\sum_{k=1}^{n} c_{i,,w_k}^2} \cdot \sqrt{\sum_{k=1}^{n} c_{j,,w_k}^2}} \tag{1}$$

4. Adjacent pages are assigned to the same segment if the cosine similarity is larger than the threshold.

Calculating importance of words

The importance of words in each segment is estimated using the TF-IDF method (Salton & Buckley., 1988). The TF-IDF method is a method for evaluating the importance of words in a document using two indicators: term frequency (TF) and inverse document frequency (IDF). Based on the TF-IDF method, the importance of word w_l is expressed by the following equation,

$$tfidf_{w_l} = tf_{w_l} \times idf_{w_l} = \frac{N_{w_l}}{N} \times log\left(\frac{S}{S_{w_l}}\right) \tag{2}$$

where N_{w_l} is the number of word w_l in the segment, N is the number of all words in the segment, S is the number of all sentences in the segment, and S_{w_l} is the number of sentences containing a word w_l in the segment.

E-textbooks are divided into title area and body area. The title represents the contents of each page, and words in the title area seem to be important. Therefore, the importance of each word is calculated by weighting words in the title area to the importance of (Eq. 2). The importance I_{w_l} of the word w_l is expressed by the following equation,

$$I_{w_l}=tfidf_{w_l}\times\left(1+T_{w_l}\right) \tag{3}$$

where T_{w_l} is the number of word w_l in the title area.

Determination of STMs to recommend based on important words

In this research, recommended STMs are determined based on the importance of the word. First, let the top n words of importance in each page be the important words representing the content of the page. Next, the top n words are used as the search query to retrieve related websites. Then, the top m of the search results are determined to be the recommended STMs and stored in the database.

Experiment

We conducted experiments to investigate the effectiveness of the proposed recommender system. In the experiments, up to five websites were retrieved as STMs for each page of an e-textbook. The experiments were conducted in two different courses. There were 53 students in course 1 and 99 students in course 2. All students were majoring in Materials Science. In each course, the teacher explained STMs recommendation at the beginning of the lecture. During the lecture, students opened the e-textbook and clicked recommended STMs as necessary. After the lecture finished, the students answered questionnaires about the use of the system. In the following, we report the results of the questionnaire, and analytics results of event logs collected by e-Book system.

Questionnaires

All the students answered questionnaires to evaluate the recommender system. (Tab. 1) shows the questionnaires. Q1 asked about whether the system was used during the lecture, because we did not force the students to use it. Q2 and Q3 asked students for the reason they used the system and about the effectiveness of the system. In Q2 and Q3, students could select multiple choices. Q4 asked whether the recommended STMs matched the content of the e-textbook.

According to the result of Q1, the number of students who used the system during the lecture was 62 (40.8% of all students). (Fig. 5) shows the results of Q2, Q3, and Q4. Sixty-two students who answered "Yes" to Q1 answered Q2 and Q3. In addition, multiple answers were allowed for Q2 and Q3. The result of Q2 show the reasons students used the system. From the result, it can be seen that many students selected "Q2= (1)" and "Q2 = (3)" as reasons for using the system. "Q2 = (1)" means that the system was used for the purpose of supplementing the contents that students could not understand. "Q2 = (3)" means that the system was used experimentally. Because it was the first time all students were informed about the system, many chose "Q2 = (3)." The result of Q3 shows how the system helped learning and how recommended STMs provided various learning effects for the students. Many students selected "Q3 = (1)" and "Q3 = (2)" among the answer items. "Q3 = (1)" means that it helped understanding, while "Q3 = (2)" means that it helped in the acquisition of new knowledge. Based on the result that many students selected "Q2 = (1)," it is assumed that learning by the recommended STMs is useful for students. In addition, there are a certain number of students who selected "Q3 = (3)." "Q3 = (3)" means that students became more interested in the content of e-textbooks by browsing the recommended STMs. The result suggests that recommendation by the system leads to motivation for learning. The result of Q4 shows the subjective evaluation of agreement about the suitability of the recommended STMs for the content of the e-textbook. More than 80% of students agreed with the suitability of the recommended STMs.

	Question items	Answer items
Q1	Did you use the recommender system during the lecture?	(1) Yes (2) No
Q2	This is a question for those who answered "yes" to Q1. Why did you use the recommender system?	(1) There was a part that I could not understand. (2) I was interested in the content of the slide. (3) I became curious about it. (4) Other
Q3	This is a question for those who answered "yes" to Q1. How was the recommender system useful?	(1) I understood the part I had not previously understood. (2) I gained new knowledge. (3) I became interested in the content of the e-textbook. (4) Other
Q4	To what extent do you think the content of the e-textbook slide and the contents of the recommended supplementary teaching material matched?	(1) Match (2) Partial match (3) Partial mismatch (4) Mismatch

Table 1: Questionnaire

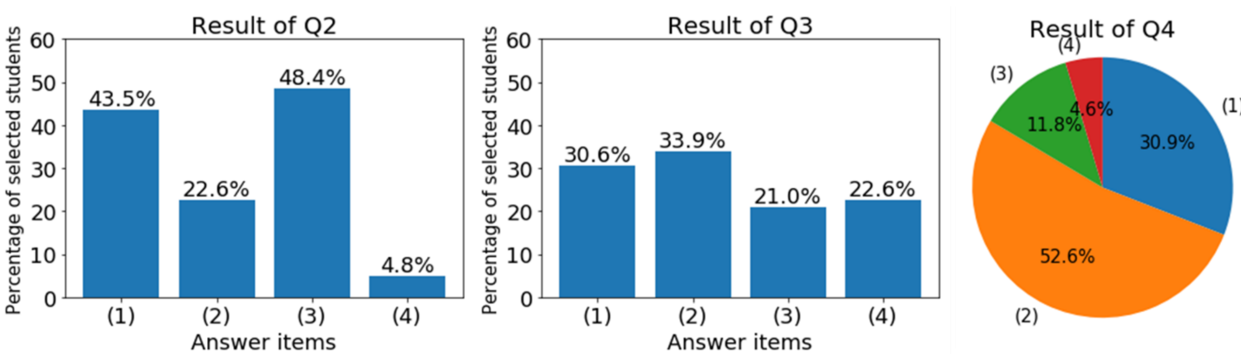

Figure 5: Result of the Questionnaire

Analysis of event logs

Event logs are automatically collected and stored in the database when students operate some functions of the e-Book system. We analyzed the event logs for the recommended STMs collected during the lecture.

(Fig. 6) shows the number of clicks on the recommended STMs for each page of the e-textbook. The number of clicks on each page shown is the total number for course 1 and course 2. In (Fig. 6), it can be seen that there were numerous clicks on or around page 32 of the e-textbook. The contents of page 32 and the front and back of the e-textbook were not explained. Instead, students were given time to browse the pages themselves. For this reason, it is assumed that the number of clicks increased because students learned based on the recommended STMs. In addition, since there are many clicks on pages that contain exercises or pages with difficult contents, it is assumed that recommending the information for each page is useful for supporting learning.

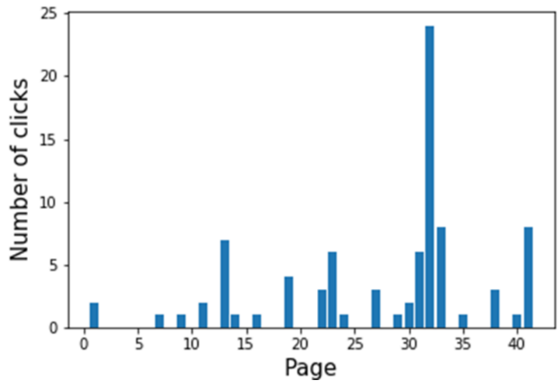

Figure 6: Number of clicks on recommended information about each page

(Fig. 7) shows the number of clicks every five minutes in each course. The duration of the lecture was 90 minutes from 13:00 to 14:30. (Fig. 7) shows that, in both courses, the use of the system increased considerably around 14:00 to 14:20. The reason for this is that the explanation by the teacher was completed around 14:00, the time for reviewing the e-textbook was given from around 14:00 to 14:20, and the post test was conducted from 14:20 in both classes. From (Fig. 7), it can be confirmed that the system can support the review process of the student after lectures, and it is assumed that recommending the information is effective.

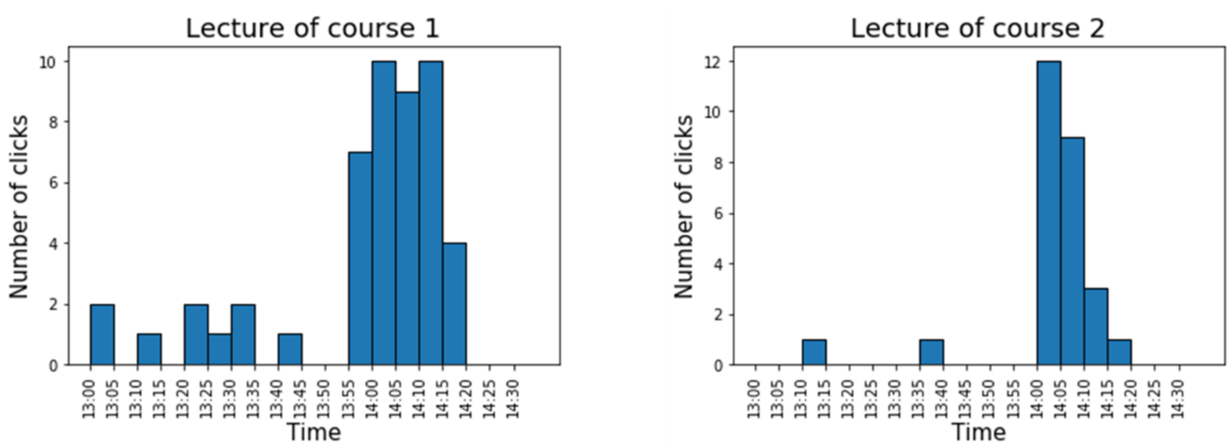

Figure 7: Number of clicks every five minutes in each lecture

Conclusion and future work

In this paper, we proposed a recommender system integrated in an e-Book viewer. The system extracts important keywords in each page in an e-textbook and recommends information related to extracted keywords. We conducted experiments using the developed recommender system in two courses of university education. Through the analytics results of student questionnaires, we obtained positive answers on the use of the system. Besides, we determined that the system was often used during self-learning time to explore related contents and/or to complement the understanding of lecture contents.

There are several aspects of future work. First, we are going to investigate how the developed recommendation system affects the grades of students. Second, it is necessary to verify the contents of the recommended information in detail. Since the difficulty level of recommended information is an important factor to support learning, we will analyze the contents of retrieved websites to discover more suitable information. Third, we will extend the recommender system for personalization. Since the demand for the recommender system will be different for each student, providing individual recommendation information is a promising solution. In our future work, we are going to improve the recommender system to support personalized and adaptive learning.

References

Bauman, K., & Tuzhilin, A. (2018). Recommending Remedial Learning Materials to Students by Filling Their Knowledge Gaps. *MIS Quarterly*, 42(1), 313-332.

Chen, G.D., Wei, F.H., Wang, C.Y. & Lee, J.H. (2007). Extending e-book with Contextual Knowledge Recommender for Reading Support on a Web-Based Learning System. *International Journal on E-Learning*, 6(4), 605-622.

Chen, W., Niu, Z., Zhao, X., & Li, Y. (2014). A hybrid recommendation algorithm adapted in e-learning environments. *World Wide Web*, 17(2), 271-284.

Erdt, M., Fernández, A., and Rensing, C. (2015). Evaluating Recommender Systems for Technology Enhanced Learning: A Quantitative Survey, *IEEE Transaction of Learning Technologies*, 8(4), 326-344.

Kudo, T., Yamamoto, K., & Matsumoto, Y. (2004). Applying conditional random fields to Japanese morphological analysis. In *Proceedings of the 2004 conference on empirical methods in natural language processing*.

Liang, G., Weining, K., and Junzhou, L. (2006). Courseware recommendation in e-learning system. In W. Liu, Q, Li, and W.H. Lau (Eds.) *International Conference on Web-based Learning 2006*, Lecture Notes in Computer Science 4181 (pp.10-24).

Mbipom, B., Massie, S., & Craw, S. (2018). An e-learning recommender that helps learners find the right materials. *The Eighth AAAI Symposium on Educational Advances in Artificial Intelligence 2018 (EAAI-18)*.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5), 513-523.

Tarus, J. K., Niu, Z., & Kalui, D. (2018). A hybrid recommender system for e-learning based on context awareness and sequential pattern mining. *Soft Computing*, 22(8), 2449-2461.

Wan, S., & Niu, Z. (2018). An e-learning recommendation approach based on the self-organization of learning resource. *Knowledge-Based Systems*, 160, 71-87.

Yamada, M., Kitamura, S., Matsukawa, H., Misono, T., Kitani, N., and Yamauchi, Y. (2014). Collaborative filtering for expansion of learner's background knowledge in online language learning: does "top-down" processing improve vocabulary proficiency?, *Educational Technology Research and Development*, 62(5), 529-553.

Acknowledgements

This work was partially supported by JST PRESTO Grant Number JPMJPR1505, JSPS KAKENHI Grand Number JP18H04125, Japan.