Ogata, H. et al. (Eds.) (2015). Proceedings of the 23rd International Conference on Computers in Education. China: Asia-Pacific Society for Computers in Education

Visualization of e-Book Learning Logs

Sachio HIROKAWA[a*] Chengjiu YIN[b], Jingyun Wang[a], Misato Oi[a], Hiroaki OGATA[b]
[a]*Research Institute for Information Technology, Kyushu University, Japan*
[b]*Faculty of Arts and Science, Kyushu University, Japan*
*yinchengjiu@gmail.com

Abstract:. Learning environment with e-book enables learners to learn anytime and anywhere they like according to their own pace. There is a large expectation on e-book as personal learning tool. Understanding and grasping the learning status of students is crucial matter for teachers and for the learning system. Access log of e-books should be basis for analyzing the learning behavior. The authors are constructing an analysis system of learning logs kept in BookLooper system operated in Kyushu University. The present paper overviews the system and shows some "learning log graphs" which represent the learning process of students. The graphs tell which pages a student had difficulties and if the student grasps the thread of course as the teacher expected.

Keywords: Learning analytics, Visualization tool, Learning log, E-books, Educational Data Mining

1. Introduction

By 2020, Japanese government is scheduled to use digital textbooks for elementary, middle, and high schools with e-books[1]. “e-book” is defined as “Texts that are digital and accessed via electronic screens”, and e-book is as effective for learning as the traditional textbook. (Rockinson-Szapkiw et al., 2013). The good usability of e-book was also reported (Shepperd et al., 2008).

As a forerunner to this institutional effort, Kyushu University carried out the BYOPC (Bring Your Own PC) program in 2012, which encourage all students to use their own PCs in the University campus, and then provide 3 learning support systems Moodle, Mahara, Booklooper, which is called M2B. Currently the M2B has around 19,000 users.

Instead of traditional textbooks, traditional classrooms in Kyushu University use BookLooper. BookLooper is a document viewer system provided by a partner to this research, Kyocera Communication Systems. Instructors’ lecture materials, such as slides or other notes, can be posted to BookLooper, which can record students’ learning behaviors when they use e-books to read their learning contents (Yin et al., 2015).

Yin et al. (2015) proposed that the students’ learning status can be shown through visualization tool. A tool, which can visually show learning behaviors of students to both students and teachers, was present by Yin et al. (2015). This tool have 2 features: (1) it can infer the relation between knowledge through learning log data; (2) it can analyze the learning styles of learners and then discover if there are correlations with learning achievements. The paper also suggests that learners with the habit of digital backtrack reading study more efficiently.

Uosaki et al. (2105) analyzed the access logs and evaluated the effect of e-book system to increase up outside-class learning time. Worm et al. (2013) evaluated e-learning effects on simple recall and complex problem-solving. For simple recall both methods were equally effective. For problem-solving, they confirmed that the e-book group achieved a comparable knowledge level compared with the text book group. They evaluated the number of logged-ins and revealed that e-book group spent significantly more time. Their analysis is limited on the number and the time. On the other hand, the present paper analyzes the sequence of access patterns in detail. Littlewood et al.(2014) analyzed the change of e-book usage for 10 years from 2003 to 2013 at the Library of the University of Waikato. They showed the number of user sessions and pages viewed over time. They succeed to confirm the increase of eBook usage. However, the target of the analysis is not on individual users but on the whole statistics.

[1] http://www.mext.go.jp/

They did not consider the linkage of pages of e-book and how the user followed the pages. Ahmad et al.(2014) applied binary logistic regression to distinguish the e-book power user from the non-power user from their log data. They adapted as the predictor variables, minutes max, sessions, titles browsed, titles read, and unique titles and obtained a model with high success rate.

Clickstream has been an important issue in Web mining for more than a decade (Facca et al.2005). Mobasher et al.(2002) analyzed the users by their behavior on access logs of Web sites and their profiles to predict their next access or give them useful recommendation. The approach of the present paper is a visualization of clickstream of learning log. Behavior analysis of students' e-book usage is gaining hot attention. Pan et al.(2014) applied the eye-tracking to evaluate graphic design effects on e-book reading. They analysed the eye-tracking as well as clickstream which were kept as log.

The present paper reports on the visualization system of e-book learning log. The system is constructed on a special search engine for log data. The thread of pages that students followed are displayed as directed graphs. Case analysis are conducted on samples of log graphs based on 298,054 logs of 108 freshmen at Kyushu University on the course of "information science".

2. Access Log Graph System

2.1 Access Log Data

The authors constructed a visualization system of e-book learning logs for the class of "information science" in Kyushu University. The class consists of 108 freshmen. The log was collected from November 2, 2014 to January 21, 2015. The total number of logs is 298,054. 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. The system visualizes how students read the pages and how moved to one page to another page. The present paper concerns the page transition by students.

2.2 Search Engine of Access Log

Fig. 1 shows a part of access log. To improve the efficiency of the analysis, we constructed a search engine for the access log. We utilized GETA[2] system in our implementation. We considered each log data as a document that consists of items of student's ID, the section, the page, the action, the date and the time. Those items are indexed as the following index file (Fig.2). For example, the first line of the log (Fig.1) is indexed as the data #92733 that contains the keywords with "tag" (Fig.2). The at-mark symbol (@) represents an item. Each line below the item number represents an item that contained in the log. The words are augmented with a tag to distinguish the interpretation.

```
user id,section,operation,page,date,time,duration
399889bc2c3efb4cf1e2962fc4eadb4e,C-03,NEXT,29,2015-01-14,09:30:37,1
399889bc2c3efb4cf1e2962fc4eadb4e,C-03,NEXT,30,2015-01-14,09:30:38,1
399889bc2c3efb4cf1e2962fc4eadb4e,C-04,OPEN,0,2015-01-14,09:30:49,7
399889bc2c3efb4cf1e2962fc4eadb4e,C-04,NEXT,3,2015-01-14,09:30:56,10
399889bc2c3efb4cf1e2962fc4eadb4e,C-04,NEXT,5,2015-01-14,09:31:06,10
399889bc2c3efb4cf1e2962fc4eadb4e,C-04,NEXT,7,2015-01-14,09:31:16,18
399889bc2c3efb4cf1e2962fc4eadb4e,C-04,NEXT,9,2015-01-14,09:31:34,6
```

Figure 1 Sample of Access Log

```
@92733
1 x:92733
1 d:150114
1 h:09
1 a:1421195042
1 i:21
1 u:399889bc2c3efb4cf1e2962fc4eadb4e
1 t:C-03
1 o:prev
1 p:29
```

Figure 2. Index of an Access Log

```
Log_Graph_Generation (Input : query) {
  @LOG = search_log(query);
  foreach user in @log {
    @log = grep {$_.id == user} @LOG;
    for(i=0;$i<@log_i;$i++){
      from = log[i].sec/log[i].page;
      to   = log[i+1].sec/log[i+1].page;
      add edge from->to
    }
  }
}
```

Figure 3. Graph Generation Algorithm

2.3 Algorithm of Graph Generation

The first line and the second line of Fig.1 indicates that the student read the page 29 of the section C-03 and then moved to the next page 30. Note that the BookLooper system displays the consecutive two pages at the screen. It may happen that only the left of the right page was kept in the access log. It is not always the case that access by the same student is kept in consecutive position in the log. In fact, when many students are using the system, the next log of a students may be other student's log. So, we have to extract all the logs of the same student to track his/her learning behavior. Then we have to follow the pages he/her read according to the time.

The system generates a directed graph whose nodes represents the pages of a section and whose edges represent the student's move to one page to another page. Even if two students moved from a page X to another page Y, the nodes X and Y and the edge X->Y are not duplicated. The frequency of the visit to the page is shown in the node.

To analyze repeated paths that students visits, such nodes are emphasized by the red color with 5 level graduation according to the frequency. The repeated paths are identified as directed closed path in the graph. The nodes in non-closed paths are drawn with no color. The red line represents the frequency of students' visiting. There is 5 levels in total. The higher frequency is, the thicker the color is.

3. Analysis Example -- access relationship map

The total of 108 freshmen attended the course of Information Science opened in October 2014. We analyze the students' learning log of the Information Science course and describe them by maps. Each map shows the reading orders in a period of time. Fig.4 shows a log/access relationship map which describes sequences of pages. The description of this map is based on the time when the learner read the pages. The words inside the circle represent the name of a document, the visited pages, and the times of the visiting. For example, "B-10/63 (6)" means that the learner read page 63 of a document called "B-10" 6 times. When a student study some document, she/he may come back to confirm some related knowledge in prior documents.

3.1 Analysis knowledge relationship Sample 1:

Fig.5 shows an example of a student's log/access relationship map. As shown in the figure, there is 5 nodes around Node"B-10/63 (6)".When the learner is reading page 5 (Linear Search) and page 27 (Suffix Array) of doucument B11, , she/he went back to page 63 (Complexity of Sort) of doucument B10 to check the related knowledge. Similarly, when the learner is reading page 63 of doucument B10, she/he jumps to page pape 1（What is Computation?） once and pape 25 (Bomplexity of Binary Search) of B11 twices.

[2] http://geta.ex.nii.ac.jp/

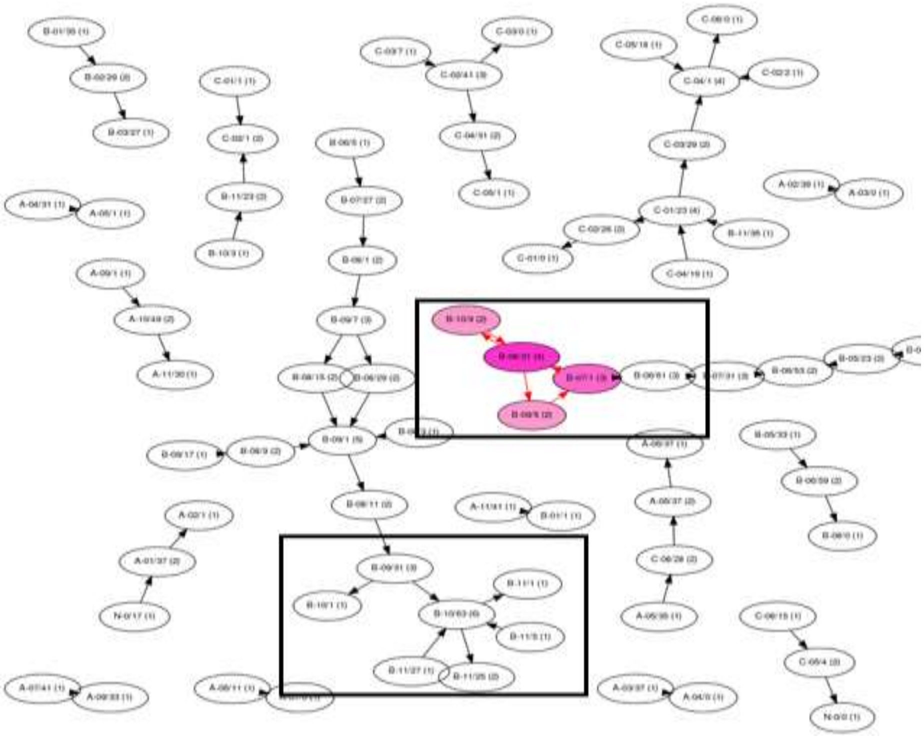

Figure 4. Visualized learning behavior

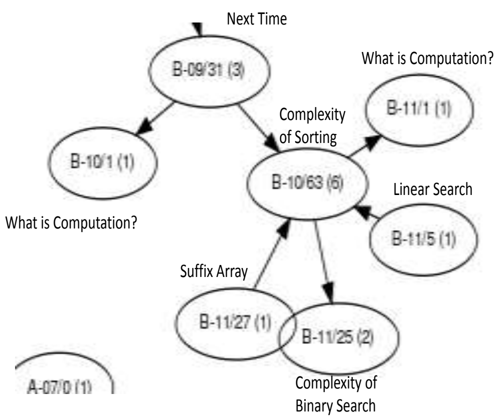

Figure 5. Relation of Contents of Slides in Figure 4 (lower sub graph)

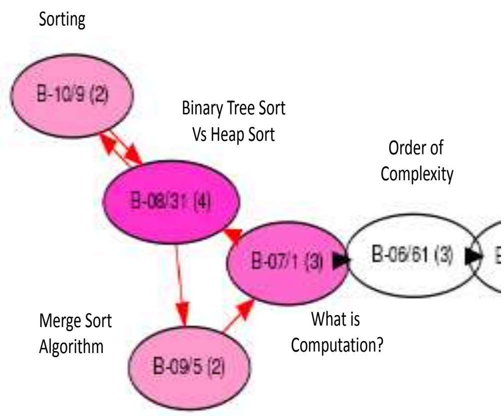

Figure 6. Repeated View in Figure 4 (upper subgraph)

3.2 Analysis knowledge relationship Sample 2:

Fig.6 descibes a situation that a student access a high frequency node repeatly. This figutre shows a log/access relationship map centered on page 31 (Binary Tree Sort vs Heap Sort) of document B-08. When a student is reading page 31 of B-08, she/he jumped to page 9 of document B-10.

Page 31 of B-08 was accessed, when the learner was reading page 1 of B-07, page 5 of B-09 and page 9 of B-10.

We confirm that those page have some knowledge is highly related. This suggest our access relationship map can successfully infer the relationship between knowledge.

4. Analysis Example -- Digital Backtrack Reading

The students received the paper tests twice at the middle and at the end of the semester. We compared the graphs of high achievement students who gained high score at the final test with that of low achievement students.

4.1 Samples of high achiement student

The students with high achievement normally repeatedly studied some key points related to the final test contents. Moreover, those key pages appear in the connected components in the graph and shown by colored nodes and edges. This kind observation would not be able to obtain by simple counting of views. Relation of pages is not clear without using visualization of the page linkage. In fact, the paths and the cycles in the graph displays the linkage so clearly that we can grasp immediately.

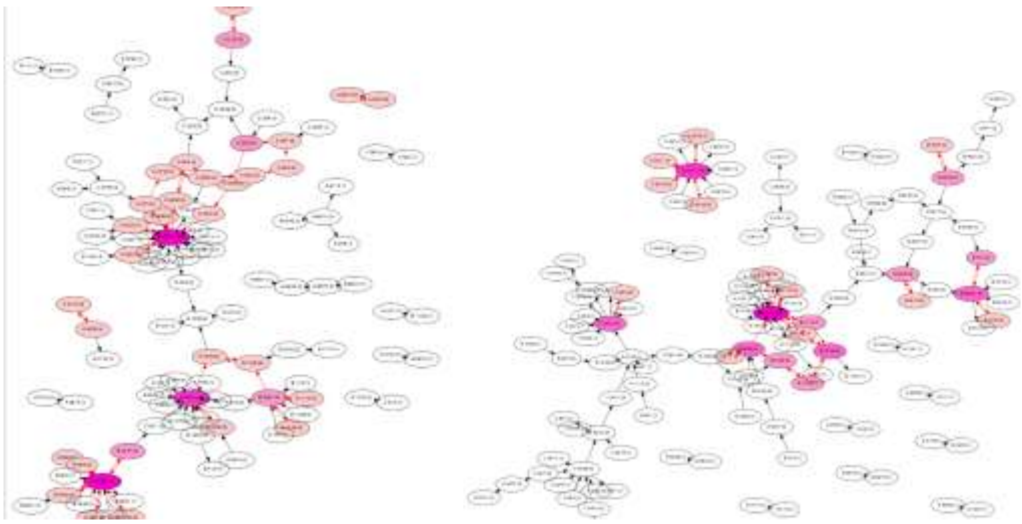

Figure 7. Two samples of high achievement students

4.2 Samples of low achiement students

There were two types of students with low achievements. One type is the students who study with some key points of the course that do not related to the test contents (Fig. 8 left). The other type is students who put very few time on study and hardly study key points (Fig.8 right).

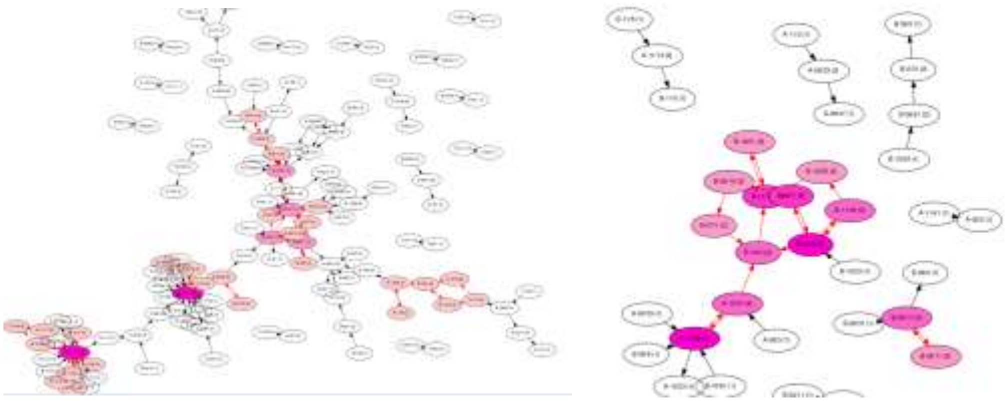

Figure 8. Two samples of low achievement students

5. Conclusion and Future work

The present paper showed a visualization system of e-book learning logs. Student learning behavior is displayed as a directed graph consisting of course pages and transitions between pages. The system overview is shown as well as samples of learning log graphs constructed from 298,054 logs of 108 freshmen at Kyushu University on the course of "information science". The graph generation system is constructed with a special search engine for log data.

By very simple analysis of those graphs, it is confirmed that students trace not only next/previous relationship of pages, but also follow the contents of course pages. Those pages are not always in the same section and do not have explicit links. We implemented the extraction of directed cycles among the graph by which we can find repeatedly studied pages. Those pages are highlighted with color and easy to recognize.

Some typical patterns were found to distinguish high achievement students from low achievement students. However, the quantitative analysis is further work as well as the improvement of the system for interactive use.

Acknowledgements

The research results have been achieved with support from the Research and Development on Fundamental and Utilization Technologies for Social Big Data (No. 178A03) and the Commissioned Research of National Institute of Information and Communications Technology, Japan.

References

C-J. Yin, F. Okubo, A. Shimada, M. OI, S. Hirokawa, and H. Ogata, “Identifying and Analyzing the Learning Behaviors of Students using e-Books”, Proc. International Conference of Computers on Education, 2015, in print.

Rockinson-Szapkiw, Amanda. J., Courduff, J, Carter, K., Bennett, D. (2013). Electronic versus traditional print textbooks: A comparison study on the influence of university students’ learning, Computers & Education, vol. 63, pp. 259–266

Shepperd, J. A., Grace, J. L., & Koch, E. J. (2008). Evaluating the electronic textbook: is it time to dispense with the paper text? Teaching of Psychology, vol. 35, no. 1, pp.2–5.

Uosaki N., Ogata H., Mouri K. (2015). How can we boost up outside-class learning time of English vocabulary Effectiveness of ubiquitous learning log system, International Journal of Mobile Learning and Organisation,Vol.9, No.2, pp.160-181

Worm B.S., Learning from Simple Ebooks (2013). Online Cases or Classroom Teaching When Acquiring Complex Knowledge. A Randomized Controlled Trial in Respiratory Physiology and Pulmonology, PLoS ONE, Vol.8, No.9

Littlewood H., Hinze A., Vanderschantz N., Timpany C., Cunningham S.J.(2014). A log analysis study of years of 10 Years of eBook consumption in academic library collections, Springer Lecture Notes in Computer Science 8839, pp.171-181

Ahmad P., Brogan M., Johnstone M.N. (2014). The E-book Power User in Academic and Research Libraries, Deep Log Analysis and User Customisation Australian Academic and Research Libraries, Vol.45, No.1,pp.35-47

Facca, F.M., Lanzi, P.L. (2005). Mining interesting knowledge from weblogs: A survey, Data and Knowledge Engineering 53 (3), pp. 225-241

Mobasher, B., Dai, H., Luo, T., Nakagawa, M. (2002). Discovery and evaluation of aggregate usage profiles for Web personalization, Data Mining and Knowledge Discovery, 6 (1), pp. 61-82

Pan T.-W., Hsu M.-C., Tsai M.-J. (2013). Effect of graphic design on E-book reading, A pilot eye-tracking study, Proc. ICCE 2013, pp. 298-305