BBookX: Creating Semi-Automated Textbooks to Support Student Learning and Decrease Student Costs

Bart Pursel[1], Crystal Ramsay[1], Nesirag Dave[1], Chen Liang[2], and C. Lee Giles[1]

[1] The Pennsylvania State University, University Park, PA 16802, USA
[2] Facebook, Menlo Park, CA 94025
{bkp10, cmg5, fcw5014, nud83, clg20}@psu.edu
liangc09@fb.com

Abstract. Open educational resources (OERs) are increasingly looked to as one approach for reducing costs and increasing access to educational materials. Unfortunately, developing OERs and operationalizing their use is fraught with difficulty. Users are challenged to search OER repositories for materials that are content-appropriate and high quality. Our team developed a new semi-automated text-authoring tool, BBookX [1, 2] to address these issues. We introduce BBookX, and discuss the utilization of a book generated using BBookX in an introductory information sciences and technology course. Survey results from students who used the book, as well as who engaged in creating their own books using BBookX, are presented. While BBookX has not been adopted for the use of creating open textbooks, the AI powering BBookX, along with faculty user testing, has led to similar derivative works in development to assist teachers with identifying relevant educational content and in creating assessments.

Keywords: Open Educational Resources, information retrieval, recommendation engines, personalized learning.

1 Introduction

Earning a college degree, particularly in the United States, is increasingly costly. One method to help defray the cost of a college degree is through the use of Open Educational Resources (OERs) designed to displace high-cost textbooks or other costly course-related resources. OERs can range from materials that represent an entire course such as textbooks, to small, modular materials such as a lesson, to individual pieces of content, such as an image or video. Hilton, Wiley, Stein, & Johnson [3] outline four different aspects of how people can use OER materials, including reusing, redistributing, revising, and remixing materials.

From a higher education perspective, the use of OERs is an alluring proposition. The obvious benefit is a reduced financial burden on students. Some OER initiatives report reducing instructional material costs by 90% for courses that adopt OERs [4]. OER initiatives can also be viewed as a prestige indicator when other universities adopt materials, and these initiatives are sometimes linked with recruiting efforts [5].

2 The Challenges of OER

Leveraging OER can be difficult. While OERs are free to the student or learner, the content is not free to produce. Instructors need to find time to identify, adapt, or create OER, then additional costs might be incurred in the technical infrastructure to store and distribute the OER. Additionally, the infrastructure must meet accessibility standards [6]. Transitioning to OERs presents a substantial time investment, as instructors locate, vet, and select OERs then invest time redesigning the course to best utilize these new materials [7]. Another challenge is assuring quality [8].

3 The Development of BBookX

The team began exploring how different AI approaches might help catalyze the adoption of OERs. We drew inspiration from SciGen [9], a search tool designed to take keywords or phrases from users, and generate artifacts in the form of computer science journal articles. This led to the conceptualization and prototyping of BBookX (https://bbookexp.psu.edu/), a recommendation engine designed to help a user generate customized books [1, 2]. The team leveraged Wikipedia as the first content repository for BBookX. Wikipedia is, arguably, the largest body of OER content available, and past research has found it to be nearly as accurate as Encyclopedia Britannica [10]. The design of BBookX begins with a searchable, local version of Wikipedia. This is preprocessed, such as removing stop words and punctuation, tokenization, and stemming. A full-text index is created for each Wikipedia document, and keyphrases from each document are extracted and indexed to compute similarity scores. The web-based interface of BBookX takes input from a user that describes keywords or phrases about a chapter the user wishes to create, then provides 10 possible matches to the user based on a similarity score that includes title similarity, content similarity, and keyphrase similarity. The user can then accept or reject each match, based on the relevance to the chapter he/she is creating. The acceptance/rejection is then leveraged to reformulate the subsequent query of Wikipedia, taking into account the user's actions to further refine each subsequent query. More details on the backend of BBookX can be found in Liang et al. 2015 [1].

3.1 BBookX Utilization

Since being published on the web in May of 2015, BBookX has 1,218 registered users who created 1,263 books, involving 132,710 search queries. Admittedly many of these users are from the Pennsylvania State University, where one of the authors both uses a textbook created by BBookX, as well as leverages BBookX with students as part of an assignment. The author created the textbook in 2015, then used the BBookX-generated text for two semesters in fall 2016 and fall 2017. BBookX was used to create each chapter of the book, then the Wikipedia content of each chapter was migrated to Pressbooks, an ebook publishing platform. Once in Pressbooks, the instructor edited the book, deleting different portions of Wikipedia pages not relevant to the course, adding

introductions and conclusions to each chapter, and inserting periodic case studies and images important to understanding key concepts.

3.2 Field Test and Student Survey

The ebook created with BBookX was designed to support an introductory course in information sciences and technology. It consisted of 15 chapters of material, where students read a chapter per week. Chapters covered foundational concepts for the course, similar to the textbook used by other instructors of the course that comes from a publisher, and students were required to complete assessments that were partially based on the text. Survey data were collected about the text, including questions that targeted students' perceptions of the credibility and utility of Wikipedia-based resources. Questions were 5-point Likert-type, with 1 representing "Strongly Disagree", 5 representing "Strongly Agree", and with a midpoint of "Neither Agree nor Disagree." Student responses (n=257) indicated generally favorable reactions when asked questions about Wikipedia readings compared to a traditional textbook.

Table 1. Student perceptions of the course ebook compared to tradtional textbooks.

Question	Response (%)				
	1	2	3	4	5
Compared to a traditional book…					
I found IST 110 required readings more interesting.	3	15	27	46	9
I found IST 110 required readings more useful.	3	9	33	44	11
I found IST 110 required readings more relevant.	2	5	24	55	14
I found IST 110 required readings more up-to-date.	0	3	16	56	25

Just over half of the students (55%) responded "Agree" or "Strongly Agree" when asked whether Wikipedia readings were more interesting or more useful than a traditional text. The majority of students (81%) either "Agree" or "Strongly Agree" that the Wikipedia-based readings are more up-to-date compared to traditional textbooks. This is likely due to the fact that traditional book publishing models often take significant time, while Wikipedia updates are published moments after a user makes a change.

A second set of questions, using the same 5-point Likert-type scale, was used to explore how students leveraged the affordances provided by the format of an ebook built using Wikipedia content.

Table 2. Student perceptions of the affordances of the course ebook.

Question	Response (%)				
	1	2	3	4	5
I prefer Wikipedia readings to traditional textbooks because…					

	1	2	3	4	5
They allow me to quickly jump to other, related readings based on my own personal interests.	3	9	20	46	22
I didn't need to purchase a textbook.	1	5	11	29	54
I can easily access the readings on any device connected to the Internet.	2	4	11	46	37

The majority of students (>80%) prefer the course ebook to traditional texts because it is free and gives them the ability to access readings from any device. One nuance of the format of this specific ebook is that it maintains the link structure found within Wikipedia articles, so students have the ability to click an embedded link in the content of the ebook, and navigate out to a Wikipedia article that sparks an interest. Approximately two thirds of the class (68%) appeared to appreciate this feature of the book when compared to traditional textbooks.

The final set of questions used a 5-point Likert-type scale where 1 represents "Never", 2 representing "Rarely", 3 representing "Sometimes", 4 representing "Quite Often", and 5 representing "Very Often" dealt with how students interacted with the ebook.

Table 3. A summary of how students reported interacting with the course ebook.

Question	Response (%)				
	1	2	3	4	5
How often did you…					
Click on a link contained on a page of our textbook, and navigate to a new Wikipedia page that was not part of the required readings for the course?	6	26	49	15	4
Re-visit assigned readings more than once?	10	25	47	15	3
Read all of the assigned pages included in a chapter?	8	26	38	23	5
Read the assigned pages on a computer (laptop or desktop)?	4	7	24	32	33
Read the assigned pages on a mobile device (smart phone or tablet)?	35	27	22	10	6
Print the assigned pages to read offline?	74	11	10	4	1

While two thirds of the students indicated an appreciation of the ability of to jump directly into Wikipedia from the ebook, only 19% of students responded that they "Quite Often" or "Very Often" click on a link containted in the ebook to jump out to a Wikipedia page that was not part of the required course readings. Also worth noting is the method students indicate consuming the eBook. In terms of mobile devices, 16% of students indicated they either "Often" or "Very Often" used a mobile device to complete readings, while 5% of students indicated they printed the readings either "Often" or "Very Often".

In addition to using a course text generated with BBookX, the instructor created an assignment that required students to interact with BBookX. He required students to

build a 3-chapter textbook, illustrating the intersection of information sciences to each student's respective discipline (all students in the class were non-information sciences majors). After the assignment, students were given a clicker-style question in class stating "BBookX surfaced interesting pages of content, includings things I did not know before completing this homework." Responses (n=249) were on a 4-point likert scale, ranging from Strongly Agree (17%), Agree (56%), Disagree (23%), and Strongly Disagree (4%). This is a positive indicator that the recommendation engine powering BBookX is helping some percentage of students learn about new and related topics and concepts within their respective disciplines.

4   Derivative Works

Through testing BBookX with faculty during its development, we observed faculty discovering new information that they did not necessarily want to include in a book, but instead repurposed this information into lecture material, course assignments, discussion prompts, or other learning materials. This led to a different prototype currently being tested that we call Eureka, designed to help a user find new information, however small that information might be, that can then be re-applied in a learning setting. A second prototype called Inquizitive is also being tested by faculty. Inquizitive uses a similar recommendation engine approach to helping instructors identify relevant distractors that can be used for multiple choice questions. A user creates a multiple choice question and provides the correct answer, and Inquizitive recommends plausible distractors for user selection. Both Eureka and Inquizitive leverage Wikipedia and provide additional use cases for how recommendation engines can be used in educational settings.

5   Conclusion

The evolution of BBookX illustrates one pathway forward for how educators can leverage Wikipedia in combination with AI-driven recommendation engines to help personalize the teaching and learning experience. As more open textbooks are released, and we can begin to index these textbooks in a standard way and use them in conjunction with Wikipedia, the accuracy and efficacy of recommendation technologies to support teaching and learning will only improve. At this point our prototypes are rather nascent and designed to be used by experts such as instructors. Once accuracy improves, however, one can imagine how these technologies will, in conjunction with a teacher or expert, help personalize student learning and diversify instruction in various settings.

References

1. Liang, C., Wang, S., Wu, Z., Williams, K., Pursel, B., Brautigam, B., … Giles, C. L. (2015). *BBookx: An automatic book creation framework.* 121–124. Lausanne, Switzerland: ACM.

2. Pursel, B., Liang, C., Wang, S., Wu, Z., Williams, K., Brautigam, B., … Giles, C. L. (2016). BBookX: Design of an Automated Web-based Recommender System for the Creation of Open Learning Content. *Proceedings of the 25th International Conference Companion on World Wide Web*, 929–933. International World Wide Web Conferences Steering Committee

3. Hilton III, J., Wiley, D., Stein, J., & Johnson, A. (2010). The four 'R's of openness and ALMS analysis: frameworks for open educational resources. *Open Learning: The Journal of Open, Distance and e-Learning*, *25*(1), 37–44.Author, F., Author, S., Author, T.: Book title. 2nd edn. Publisher, Location (1999).

4. Butcher, N., & Hoosen, S. (2012). *Exploring the business case for Open Educational Resources*. Commonwealth of Learning (COL).

5. Annand, D. (2015). Developing a sustainable financial model in higher education for open educational resources. *The International Review of Research in Open and Distributed Learning*, *16*(5).

6. Daniel, J., & Uvalić-Trumbić, S. (2011). UNESCO-COL guidelines for open educational resources (OER) in higher education.

7. Browne, T., Holding, R., Howell, A., & Rodway-Dyer, S. (2010). The challenges of OER to Academic Practice. *Journal of Interactive Media in Education*.

8. Yuan, M., & Recker, M. (2015). Not all rubrics are equal: A review of rubrics for evaluating the quality of open educational resources. *The International Review of Research in Open and Distributed Learning*, *16*(5).

9. Stribling, J., Krohn, M., & Aguayo, D. (n.d.). Scigen-an automatic cs paper generator, 2005. *URL Http://Pdos. Csail. Mit. Edu/Scigen*.

10. Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, *438*(7070), 900–901. https://doi.org/10.1038/438900a