Opening the book on the price of student reading lists

Eamon Costello, Richard Bolger, Tiziana Soverino, Mark Brown and Gráinne Conole
Open Education, DCU, Glasnevin, Dublin
Ireland
eamon.costello@dcu.ie

Abstract: This paper summarises the findings of a study of textbooks costs reported elsewhere in a journal article and an associated open dataset. The cost of student textbooks is a huge concern in higher education, especially in North America. Less is known about the costs of textbooks in other parts of the world, including in Europe. We address this gap in the knowledge through a case study of one Irish higher education institution, focussing on the cost, accessibility, and licensing of textbooks. We report here on an investigation of textbook prices drawing from an official college course catalogue containing several thousand books. We explain how we sought to determine meta-data of these books including: formats (e-book, PDF) that they are available in, whether they are in the public domain and their commercial prices. We detail the methods we used to estimate textbook costs by using the Google Books API and how we made the code and dataset freely available.

Introduction

Uptake and use of iPads, tablets, e-book readers and smart phones has rapidly increased in recent years, which is having an impact on how learners learn. Open e-textbooks are a specific type of Open Educational Resource (OER) (Hilton, 2016). Many courses, particularly in the United States (US), have a dedicated textbook. Increasingly, paper-based textbooks are being replaced by e-textbooks, as evidenced by Seaman and Seaman (2018). E-textbook are more flexible than paper-based textbooks in a number of respects: the learner can search for keywords, highlight and annotate text, interact with multimedia resources and bookmark pages. A number of authors argue that e-textbooks are beneficial because of their flexibility, accessibility, interactivity, and extensibility (Daniel and Woody, 2013; Murray and Perez, 2011). However, Gu et al. (2015) argue that the promise of the potential of e-textbooks has not yet been realised, as issues around screen resolution and licensing restrictions remain.

Despite increased use of e-textbooks, the cost of textbooks remains a hot topic in Higher Education. It has been reported that by 2014 the average student spent $1,200 annually on textbooks (Baglione and Sullivan, 2016). Another study claimed that between 2006 and 2016 the costs of college textbooks increased over four times the cost of inflation (Senack and Donoghue, 2016). Despite this rise in textbook costs, a survey of more than 4,000 US faculty members and chairpersons ("The Babson Survey") found that 69% of all faculty reported the use of a required textbook in their course, and 98% of respondents still selected copyrighted printed textbooks (Seaman and Seaman, 2018). One response to the challenge of rising textbook costs is open textbooks. Open textbooks are a type of OER. OERs have been defined as "teaching, learning, and research resources that reside in the public domain or have been released under an intellectual property license that permits their free use and re-purposing by others. Open educational resources include full courses, course materials, modules, textbooks, streaming videos, tests, software, and any other tools, materials, or techniques used to support access to knowledge." (Hewlett Foundation, Open Educational Resources). OERs stem from the principle that access to education is a human right and that, as such, education should be accessible to all (2012 Paris OER Declaration). Hence an open textbook is made available under terms which grant legal rights to the public, not only to use, but also to adapt and redistribute it. Creative Commons licensing is the most prevalent and well-developed intellectual property licensing tool for this purpose. Open textbook projects aimed at promoting publishing and redistributing open textbooks, both in digital and print formats, have been growing. For example, the BCampus project in Canada began in 2012 with the aim of creating a collection of open textbooks aligned with the most popular subject areas in British Columbia (Burgess, 2017). The project has shown strong growth, with over 230 open digital textbooks now available and more than 40 institutions involved. A significant recent development in open textbooks occurred in March 2018, when the US Congress announced a $5 million investment in an open textbook initiative (Allen, 2018). In addition to helping change institutional culture, and challenge attitudes to traditional publishing models, one of the most oft-cited benefits of

open textbooks is cost savings. According to the College Board's Survey of Colleges, the average annual cost to US undergraduate students in 2017 for textbooks and materials was estimated at $1,250 (Ma, Baum, Pender and Welch, 2017). This figure is remarkably close to the aforementioned figure of $ 1,200 a year, reported by Baglione and Sullivan. However, there is little known about the monetary face value of books that students are expected to buy, beyond studies based on self-reported data. Students themselves in the USA have attempted to at least open the debate in this area by highlighting book price disparities (Vitez, 2018). Nonetheless, they only report on a very small number of books, and the College Board representing on-campus US textbook retailers have disputed their results for this reason, claiming that they have been selective in the book prices they have chosen.

Hence this study sought to address the gap that exists in knowledge about the true cost of textbooks in higher education. We have published the full results of this study elsewhere and this paper gives a summary of these findings (Costello et al., 2019b). This is in the context of a wider research project we are conducting on open textbooks in Ireland (Brown, Costello and Nic Giolla Mhichíl, 2018). Determining the cost of books is not straightforward as books can be new, used, rental or digital subscription. However, the cost of new books does set a baseline for other forms, particularly rental and used books. Our aim here is hence to start with new books, by analysing costs of all the required and recommended textbooks of one Higher Education Institution (HEI) in Ireland.

Hence we sought to find out about: the extent of textbooks that are required reading; the retail costs of new books; the availability of books in digital or e-book form; and the availability in the public domain of each book.

Methods

In this section we describe our approach, the dataset generated and the methods we used to analyse the data. We identified a suitable data source comprising the official course catalogue of a HEI in Ireland with more than 10,000 students. In the course catalogue, faculty give required and recommended textbook details for all courses. This information is freely accessible on the website of the HEI; the course catalogue is powered by a software system known as Akari (http://www.akarisoftware.com/). Akari is a proprietary software system used by several higher educational institutions in Ireland and outside to create and manage academic course catalogues. The course team gained access to a download of all books recorded in the database of the course catalogue. In this catalogue, fields are provided for lecturers to input information for students about books such as title, International Standard Book Number (ISBN), author and publisher. Following manual and automated data cleansing, 3,014 unique records of books were created. Due to the large number of books, at this stage we sought a programmatic solution for finding out more information about these books. We initially thought that ISBNs might prove the best way to accurately reconcile records of books. However, many ISBNs were incomplete or mistyped. Moreover, many instructors simply did not enter an ISBN, or simply entered "I will tell you in class" in the book title field. Given the capacity for errors in the data, we required a tool that could handle fuzzy search queries e.g. cases where a book title or author were misspelled. The tool we selected was the Google Books Application Programming Interface (API) (https://developers.google.com/books/docs/v1/reference/volumes). This API provides an interface to the Google Books database of circa 30 million books. The service, like the main Google search engine, is forgiving of queries that are mistyped or misspelled. Hence, we constructed a query based on a combination of author name, book title and publisher. Following experimentation, we determined that these three search terms together allowed us to find books with a high degree of accuracy whilst also accounting for possible spelling errors.

We then wrote a JavaScript program deployed in the Google Cloud Platform. This program parsed the file of the book search queries, passed them to the Google Books API as search requests and saved the results. The API returned results in JavaScript Object Notation (JSON) format. JSON is a modern web language for describing data. It is related to JavaScript and can be used to translate objects in the JavaScript programming language into textual strings. It is used as a replacement for XML as it is arguably more human readable and less verbose. We imported this JSON into a MongoDB database to filter and clean the data, before exporting to Excel for statistical analysis. MongoDB is a document store database that natively stores objects in the JSON. The Google Books API provides key metadata on books. As well as author, publisher, ISBN, edition, pages etc. as it also provides prices for many of its books. Google draws this information from its own e-book store, which contains over 3 million books and a network of resellers who sell print and digital versions of the books. In addition to price, Google Books also contains information on accessible versions of books, digital/e-pub versions, PDF versions and whether the book is in the public domain. We have published a release of this dataset and all of our code to the software repository GitHub.

We then used the Zenodo platform to generate a Digital Object Identifier (DOI) for the code (Costello and Bolger, 2019). One of the functions of the Zenodo platform is to allow for code to be properly cited and referenced. We published our code in this way for others interested in replicating this work in other contexts. In the next section we will provide an analysis of the results of our queries.

Results

After extracting and processing the data from the course catalogue and Google platforms, we obtained 3,030 unique course names and in these courses, we found over 15,414 books listed. From the course catalogue data, we found that 11,022 (71.5%) books were required readings and the remaining 4,392 (28.5%) were recommended. Upon cleaning and removing duplicates and missing data, we identified 3,014 books that could be queried using the Google Books API. Querying the API returned results for 2,940 books i.e. it returned results for 97% of the books, while only 74 books could not be found. The Google Books API returns information in JSON format. From the total of 2,940 individual books listed, their availability was as follows: 854 were available in both PDF and e-book format; only 0.18% (6) of books had a version available in the public domain according to Google Books.
With regard to costs, the Google Books API only returned prices for 596 (20%) of the books that we searched for. Within that sample, the cost ranged from $0.99 to over $452. The median price of a book was $40. Given an average 3.96 books per course, there is an average cost of $224.41 per course per student. With an average of 8.05 courses per student per year, the cost per year is $1,806.50 for each student should they buy new versions of each book.

Discussion and Conclusion

We demonstrated that it is possible to programmatically search and discover prices of large numbers of books. We used this data to attempt to estimate the full economic cost of books to students. We are still actively developing this technique and encourage others to use and even contribute to the code which we have published with the dataset. This proof of concept tool may allow stakeholders with an interest in book costs for students to quickly get real data on large numbers of books. Ultimately, we hope that this will help highlight the costs of many textbooks. Our findings also highlight relatively low levels of digital book availability. Very few books were found to be in the public domain. A limitation of this research is that there are issues around the coverage of Google Books and its index policies or algorithms. In a literature review of research articles about Google Books in 2017, Condit Fagan pointed out that the coverage of Google Books is “hit and miss.” (2017). In 2017, Google Books included about 30 million books, though Google did not release specific details on its database, as also emphasized by Condit Fagan (2017). It is known that content includes digitized collections from over 40 libraries, and that US and English-language books are overrepresented (Condit Fagan, 2017). Furthermore, Google Books is only returning results for books that are in the public domain and cannot tell us if books are made available through open licenses such as Creative Commons. Accepting such caveats, however, we have found the Google Books API to be a very useful tool for answering questions about large numbers of books in a systematic way and hope that our findings can help others.

The prices that we derived in this research were for the cost of new books only. However, new book prices do provide a baseline for other prices e.g. a used book or a loan book price. These prices have large implications for students and libraries (Christie, Pollitz and Middleton, 2009). More research is needed to find out more about the actual costs of textbooks by asking students directly and hence better understanding the lived experience of textbooks in their academic lives (Costello et al., 2018).

References

Akari software, http://www.akarisoftware.com/

Allen, N. (2018). Congress Funds $5 Million Open Textbook Grant Program in 2018 Spending Bill, (March 20 News), SPARC Open, https://sparcopen.org/news/2018/open-textbooksfy18/.

Baglione, S. L. & Sullivan, K. (2016). Technology and Textbooks: The Future. *American Journal of Distance Education*, 30 (3), 145-155, https://doi.org/10.1080/08923647.2016.1186466.

Brown, M., Costello, E. & Nic Giolla Mhichíl, M. (2018). From Books to MOOCs and Back Again: An Irish Case Study of Open Digital Textbooks. In A. Volungeviciene & A. Szűcs (Eds.) *Exploring the Micro, Meso and Macro. Proceedings of the European Distance and E-Learning Network 2018 Annual Conference, Genova, 17-20 June, 2018* (pp. 206-14). Budapest: The European Distance and E-Learning Network.

Burgess, M. (2017). The BC Open Textbook Project. In S. R. Jhangiani & R. Biswas-Diener (Eds.). *Open: The philosophy and practices that are revolutionizing education and science* (pp. 227–36). London: Ubiquity Press.

Christie, A., Pollitz, J. H. & Middleton, C. (2009). Student strategies for coping with textbook costs and the role of library course reserves. *Portal: Libraries and the Academy*, 9 (4), 491-510, http://digital.library.wisc.edu/1793/38662.

Condit Fagan, J. (2017). An Evidence-Based Review of Academic Web Search Engines, 2014-2016: Implications for Librarians' Practice and Research Agenda. *Information Technology and Libraries*, 36 (2), 7-47, https://doi.org/10.6017/ital.v36i2.9718.

Costello, E., Bolger, R., Soverino, T. & Brown, M. (2019b). Determining textbook cost, formats, and licensing with Google Books API: a case study from an open textbook project. *Information Technology and Libraries,* 38 (1), https://doi.org/10.6017/ital.v38i1.10738

Costello, E. & Bolger, R. (2019). Textbooks authors, publishers, formats and costs in higher education. *BMC Research Notes* 12 (1), 12-56, https://doi.org/10.1186/s13104- 019-4099-1.

Costello, E., Brown, M., Brunton, J., Bolger, R. & Soverino, T. (2018). Textbook costs and accessibility: Could open textbooks play a role? In Prof. K. Ntalianis, Prof. A. Andreatos, & Prof. C. Sgouropoulou (Eds.), *Proceedings of the 17th European Conference on eLearning (ECEL)*, vol. 17 (pp. 99-106). Sonning Common: Academic Conferences and Publishing International.

Daniel, D. B. & Woody, W. D. (2013). E-Textbooks at what cost? Performance and use of electronic v. Print texts. *Computers &; Education*, 62, 18-23.

Google Books API (2018), https://developers.google.com/books/docs/v1/reference/volumes.

Gu, X., Wu, B. & Xu, X. (2015). Design, development and learning in e-textbooks: what we learned and where we are going. *Journal of Computers in Education*, 2(1), 25-41, https://link.springer.com/article/10.1007/s40692-014-0023-9.

Hewlett Foundation, Open Educational Resources, http://www.hewlett.org/programs/education-program/open-educational-resources.

Hilton, J. III (2016). Open educational resources and college textbook choices: a review of research on efficacy and perceptions. *Educational Technology Research and Development*, 64 (4), 573–590.

Ma, J., Baum, S., Pender, M. & Welch, M. (2017). *Trends in College Pricing*, Report, The College Board. Retrieved from https://trends.collegeboard.org/sites/default/files/2017- trends-in-college-pricing_0.pdf.

Murray, M. C., & Perez, J. (2011). E-Textbooks are coming: Are we ready. *Issues in Informing Science and Information Technology*, 8, 49–60.

Seaman, J. E. & Seaman, J. (2018). *Freeing the Textbook: Educational Resources in U.S. Higher Education, 2018,* Report, BABSON Survey Research Group. Retrieved from https://www.onlinelearningsurvey.com/reports/freeingthetextbook2018.pdf.

Senack, E. & Donoghue, R. (2016). *Covering the Cost: Why we can no longer afford to ignore high textbook prices,* Report, The Student PIRGs (February). Retrieved from www.studentpirgs.org/textbooks.

Vitez, K. (2018) *Open 101 An action plan for affordable textbooks*, Report, Student PIRGs. Retrieved from https://studentpirgs.org/campaigns/sp/make-textbooks-affordable.

2012 Paris OER Declaration, http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/WPFD2009/English_Declaration.htm.