UBR: A Book Search — Recommender Hybrid

Jason Hall, Maria Soledad Pera
People and Information Research Team
Computer Science Department
Boise State University
Boise, Idaho, 83725

Abstract

We present UBR, a Unit Book Recommender that combines search and recommendation into a single tool that eases the task of locating educational resources, specifically books. UBR relies on the extensive selection of books available at popular book websites and the expert knowledge archived on educational websites to generate a list of suitable book alternatives that can be readily obtained and used when teaching any unit of study. As a case study, UBR focuses on the Joy School teaching curriculum but has applications for other preschool or elementary schools that have predefined units of study.

1 Introduction

Teaching children is fundamental in shaping the future. Those who teach them have units of study (a collection of lessons focused on a specific topic) that help prepare the children for that future. These units of study are particularly helpful in non-traditional educational environments, such as home schooling, where educators are parents, who are not necessarily teachers (Greenwalt 2016; Mayberry and others 1995; Medlin 2000). As home schooling continues to grow (Ray 2015), addressing the challenge of finding suitable books to use when teaching becomes imperative. Some units of study or specific lesson plans provide books that relate to a given topic but what happens when the books on that list cannot be found?

Teachers spend a lot of time each day preparing their lessons, researching and gathering materials and, of course, teaching. In steering this process, many educational curriculums have specific Units of Study that guide teachers' materials and/or topics selections. However, some do not have a list of supplemental teaching materials, most specifically a list of books that would go well with each lesson. Others might include this list, but the items on it may be hard to find or even out of print. The teacher then loses time in trying to locate these materials instead of being able to focus on other of the aforementioned tasks. To illustrate this point, consider the following two examples.

- Teacher A has a week of lessons that focus on honesty. The lessons overview contains a list of recommended books for supplementing the lessons but Teacher A cannot find any of them.
- Teacher B has taught lessons that focus on honesty for 20 years and has gathered all of the suggested supplemental teaching resources but the resources are now outdated. Teacher B wants to find some new resources without spending a lot of time doing so.

To aid teachers in their quest for supplemental teaching resources, e.g., books, correlating with their lesson plans, we introduce **UBR**, a Unit Book Recommender. UBR achieves its design goal by extracting the main ideas from a Unit of Study that are then used to search popular sites archiving books, such as Amazon[1], for a set of candidate books to recommend. The retrieved set is then ranked based on key features from expert reviews found on educational-related sites, such as Common Sense Media[2] (CSM), and the top-k most suitable books are recommended to the teacher.

Although UBR is meant to help teachers regardless of the lesson plans they follow for teaching purposes, the scope of this research work focuses on books, as opposed to general educational materials, and we use as a case study the units of study defined by the Joy School educational program. Joy School was founded by Richard and Linda Eyre and is for preschool aged children (ValuesParenting.com 2015). This program was chosen because it has over 100,000 participants and a well-defined teaching curriculum. In fact, the idea for UBR came into being over conversations with Joy School teachers about the difficulty they had when trying to find the books listed for a given unit of study. This problem is what led to the birth of UBR.

UBR uses a simple strategy for extracting the main ideas from a unit of study and then uses those ideas to generate queries for finding age-appropriate books that are currently in stock at Amazon. These are the books that are treated as candidate books to be recommended. We chose Amazon since, as of today, it archives over 249,000 books rated for children between the ages of 3 and 5 inclusive (Amazon 2015). Amazon was also chosen because of its nationwide shipping options to solve the problem of availability. Candidate books are further examined based on information identified from existing expert reviews on CSM. We chose

[1] www.amazon.com/

[2] www.commonsensemedia.org

CSM because of their expert knowledge on media for children. The only limitation to using CSM is the number of books that have been expertly reviewed. As of the date of this paper, CSM archives over 3,700 books that have been reviewed by multiple parents and educators (CommonSenseMedia.org 2015) Lastly, candidate books are ranked based on a number of traits considered by UBR so that the top-k books that are deemed better suited to correlate with a teachers lesson plan are recommended.

By considering popular sites that archive books, UBR ensures that books recommended to teachers are in circulation. Furthermore, informing the recommendation process based on knowledge extracted from expert reviews translates into prioritizing relevant resources of educational value, as opposed to just popular ones.

In the remaining of this paper, we first briefly discuss works related to search and book recommendation, two areas of study most closely related with UBR (see Section 2). Thereafter, in Section 3, we introduce the design methodology of UBR. In Section 4, we detail the results of the initial assessment based on teachers' feedback, conducted in order to assess the correctness of UBR. Lastly, in Section 5, we offer concluding remarks and directions for future work.

2 Related Work

There are many proposed solutions for search and recommendation. However, researchers nowadays argue in favor of systems that blend search and recommendation into a single tool (Chi 2015). UBR is designed with this principle in mind as it combines these two strategies along with an automatic query-generation methodology to locate books available to purchase that are not out of circulation into a single system. In the remaining of this section we briefly discuss research works related to UBR's areas of study.

A number of tools have been designed to aid teachers in their quest of materials. BReT, introduced in (Pera and Ng 2014b), can be employed by teachers to locate books of interest to each of their students. Other popular sites teachers can use to locate materials include Scholastic.com, Pbslearningmedia.org, Readingglue.com, or Instagrok.com. Unfortunately, BReT as well as these aforementioned sites, require teachers to provide detailed specifications related to the type of material they are looking for, from keyword descriptions to readability levels.

There is no lack of literature on book recommenders (Pera and Ng 2014a; Shapira et al. 2011). Rabbit, introduced in (Pera and Ng 2014a) relies on emulating the reader's advisory process to identify books to be recommended. Similar to Rabbit, UBR also depends upon appeal terms to qualify books. However, Rabbit as well as traditional book recommendation strategies (Shapira et al. 2011) are constrained by the existence of a user profile to initiate the recommendation process, while UBR focuses on a lesson plan.

UBR correlates most with the works introduced by Zhang et al. (2014) and Htait et al. (2014). The former focuses on social book search re-ranking with generalized content-based filtering. The latter discusses a book recommendation system which automatically generates queries that are submitted to Amazon to locate books that will be recommended to users. Their proposed strategy is based on the combination of book tags mentioned on the requests, other extracted information, including price, number of pages and publication date of books, as well as information about previous books read by users. Unlike its counterparts, UBR relies on pre-defined filters to retrieve books that are age relevant and in stock. Furthermore, UBR creates queries on-the-fly based on a Unit of Study, as opposed to the strategy in (Zhang et al. 2014) which depends on users triggering queries or the one in (Htait, Fournier, and Bellot 2014) which depends upon user descriptive requests.

3 The Proposed UBR

In its initial conception, UBR makes the following assumptions for it to function. First, it assumes that the format that Amazon uses for displaying their search results and product pages will not change. Second, that CSM will not change the way that they display their search results and review pages and third, that Joy School will not change the format of its header in their Units of Study. Given those assumptions, we present below a brief overview of the recommendation process of UBR (which we depict in Figure 1).

3.1 Extraction

UBR first examines a Unit of Study file provided by a teacher, i.e., the lesson plan for which the teacher needs supplemental material, and extracts the main ideas, i.e., concepts, from it. UBR does this by using Apache PDFBox,[3] an open source library, to read the file. In order to streamline the process, UBR scans each line as it is read looking for specific markers in the unit of study that indicate main ideas. Joy School, for example, uses the phrase "The Joy of ..." to indicate the main ideas. Once the main ideas have been found, UBR stops reading through the provided file and proceeds with the next step.

3.2 Query Generation

Once the main ideas are gathered, UBR creates search phrases that will be used to query Amazon and limit the results to (i) children's books for ages 3 to 5 inclusive and (ii) books that are currently in stock. The desired goal of this step is to retrieve candidate books to be recommended to a teacher. In order to maximize the chances of finding as many books as possible given the main ideas related to a given topic in a unit of study, the queries are formulated by (i) substituting the words"The Joy of" for "Books about", (ii) using the phrase in its entirety as one query, and (iii) splitting the original phrase apart, removing all stock words, and then appending "Books About" to each of the remaining words. This results in multiple queries that can be used for a unit of study, such as "Books about Honesty and Candor", "Books about Honesty", and "Books about Candor".

3.3 Candidate Books

Using the same approach as the one employed for reading through the provided lesson file, UBR analyses the stream

[3] https://pdfbox.apache.org/

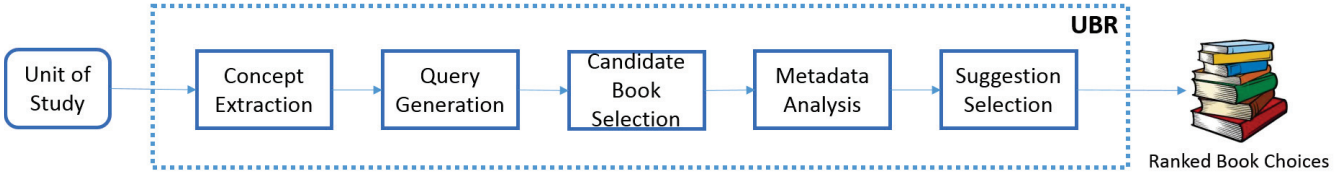

Figure 1: An overview of UBR's search/recommendation hybrid strategy

returned from Amazon until specific details about each book had been gathered, including total number of books found, ranking position of each candidate book (as defined by Amazon), Amazon URL for the specific book, URL for the Book Cover image, author, total number of pages, ISBN10, and ISBN13[4] for the book. The ranking position of each book is used to create an initial weight for ranking the results for recommendation. The initial weight was simply calculated by taking the total number of results and subtracting the current position from that.

3.4 Book Metadata Analysis

Using the title of each candidate book to be recommended, UBR creates a query which is submitted to CSM and limits its results to books. Once again, UBR crawls the results and looks for a book match, i.e., if either the Amazon book title is contained in the CSM title or vice-versa and the authors match. UBR then examines a set of key features from CSM reviews that are used to qualitatively describe each candidate book. Among these factors, CSM pre-defines age, overall quality, educational value, and language. After aggregating the observations provided by a sample of Joy School teachers, it was determined that these factors were the most valuable, in terms of prioritizing suitable candidate books.

3.5 Recommendation Generation

Having identified candidate books and qualifying each of them based on a number of criteria, UBR adopts a linear weighted aggregation strategy[5] to combine scores computed for multiple evidences. As detailed in Section 3.4, the educational value, book quality, age suitability, similarity with respect to topic in a lesson for which recommendations are being generated are evidences that are weighted positively, whereas other information, such as the levels of violence described in the books, inappropriate language, and consumerism are weighted negatively.

UBR examines these book traits for recommendation purposes, since these are the attributes that educators tend to consider when selecting teaching materials (Pera and Ng 2014a). By using this weighted model, UBR generates a single ranking score that is then used to determine which books among the retrieved from Amazon, i.e., candidate books, are the ones that offer more educational value and thus should be recommended.

[4]ISBN is a unique book identifier.

[5]We empirically determine the weights of each evidence. Details on this study are omitted due to space limitations.

4 Initial Assessment

In this section we discuss the initial assessment conducted to verify the performance of UBR.

4.1 Dataset & Baseline Strategy

To the best of our knowledge, there is no currently-available benchmark dataset that can be used to measure the performance of UBR. To overcome this issue, and present an initial validation of our claims, we took advantage of 5 Joy School units (VP 2016). As previously discussed in Section 3.1, these Joy School units refer to a specific topic or theme, such as "The Joy of the Body". Each unit is comprised of various classes or sessions related to the topic of the unit, e.g., "the parts of the body" and is complemented with lists of supplementary materials, including books, which educators can use to inform and prepare their lessons.

For each of the topics specified in these units, we generated the corresponding recommendations using both UBR and Amazon. We treat the latter as the baseline strategy in which the set of books recommended for a given topic in a unit of study is generated by directly querying Amazon.

We asked 5 Joy School teachers to assess the relevance of the recommendations generated by both systems, using a blind-approach that prevented them from knowing which system generated each recommendation, to avoid any bias in the assessment. We treated their selected relevant book suggestions as the *gold standard* for evaluation purposes.

4.2 Metrics

To evaluate UBR and Amazon we considered two standard performance measures: Normalized Discounted Cumulative Gain ($NDCG$) and Mean Reciprocal Rank (MRR) (Croft, Metzler, and Strohmann 2010). The former considers the correctness of the recommendations and penalizes relevant recommendations positioned lower in the ranking. The penalization is based on a reduction, which is logarithmically applied to the position of each relevant item in a ranked list. The latter captures the average number of suggested items a user has to scan through to identify a relevant one.

4.3 Results

To illustrate the need for a tool such as UBR, we first examined the lists of books provided along with each of the units considered for evaluation purposes, as detailed in Section 4.1. Upon manual assessment, we observed that (i) the majority of the books are outdated and (ii) on average, 40% of the books are out of print.

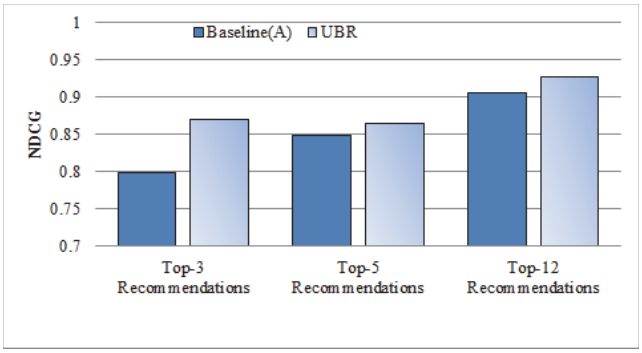

Figure 2: Assessment based on teachers feedback

We also conducted an initial assessment to determine the degree to which using UBR could ease the task of locating valuable resources to inform lesson plans using the dataset and performance metrics discussed in Section 4.1 . We empirically verified that UBR outperforms the baseline considered for evaluation purposes in terms of MRR (see Figure 2). In fact, across all lessons, the first book suggested by UBR is relevant. For UBR's counterpart, at least two books need to be examined before identifying a relevant one.

As shown in Figure 2, UBR is also on par or better than the baseline for each unit of study, in terms of NDCG. This means that regardless of the topic of the unit, UBR recommends materials that are not only relevant to the corresponding lesson, but also these resources are positioned higher in the ranking. This means that in general, teachers minimize time and effort in locating relevant resources of interest. More importantly, teachers do not have to create queries to locate suitable books, as UBR does it automatically, which reduces their burden.

5 Conclusions and Future Work

We presented UBR, a tool that recommends books to teachers, based on their lesson plan needs. This is accomplished using a hybrid search/recommendation strategy that examined online resources retrieved from popular book sites and expert reviews found on educational sites. Initial experiments conducted based on teachers' assessments demonstrated the promising performance of our tool, which examines book metadata and features to favor those books that exhibit traits that correlate with main ideas (automatically identified) from a Unit of Study.

The first step for future work is to conduct other studies to not only further validate the correctness and usability of UBR, but also compare its performance to that of state-of-the-art. In addition, we plan to extend UBR so that it can take advantage of other online resources, beyond Amazon and CSM, which are used in our case study, as well as content matching between lesson plan keywords and expert reviews, and more in-depth analysis based on type dependencies to identify book appeal terms mentioned on expert reviews (see (Pera and Ng 2014a) for further details on the use of appeal terms to describe books). These appeal terms are fundamental in determining the suitability (in terms of age and comprehension level of students) of books to be suggested. Currently, UBR is designed to work specifically with the Joy School curriculum. However, we envision enhancing UBR so that it is capable of accepting Units of Study from other schools curriculum besides Joy School. Another avenue of research we plan to explore is updating the design of UBR so that it can serve educators by facilitating the tailoring of book recommendations to suit the needs and interests of individual students, i.e., pursuing personalization of book recommendations. We also want to enhance the design of UBR so that it can recommend other educational media besides books, such as movies, games, and music. In this way, UBR will become not just a book recommender but a *multimedia* recommender that can aid teachers in their quest for materials that can engage students in their learning process.

References

Amazon. 2015. Amazon's children books. In *Retrieved from http://goo.gl/B5IVIs*.

Chi, E. H. 2015. Blurring of the boundary between interactive search and recommendation. In *IUI*, 2–2. ACM.

CommonSenseMedia.org. 2015. Common sense media book reviews. In *Retrieved from http://www.commonsensemedia.org/reviews/category/book*.

Croft, W. B.; Metzler, D.; and Strohmann, T. 2010. *Search engines*. Pearson Education.

Greenwalt, K. 2016. Home/schooling our children. In *Home/Schooling*. Springer. 95–108.

Htait, A.; Fournier, S.; and Bellot, P. 2014. Sbs 2016: Combining query expansion result and books information score for book recommendation. Available at http://ceur-ws.org/Vol-1609/16091053.pdf.

Mayberry, M., et al. 1995. *Home Schooling: Parents as Educators.* ERIC.

Medlin, R. G. 2000. Home schooling and the question of socialization. *Peabody Journal of Education* 75(1-2):107–123.

Pera, M. S., and Ng, Y.-K. 2014a. Automating readers' advisory to make book recommendations for k-12 readers. In *ACM RecSys*, 9–16. ACM.

Pera, M. S., and Ng, Y. K. 2014b. How can we help our k-12 teachers?: Using a recommender to make personalized book suggestions. In *IEEE/WIC/ACM WI-IAT*, 335–342. IEEE Computer Society.

Ray, B. D. 2015. Research facts on homeschooling. *National Home Education Research Institute*.

Shapira, B.; Ricci, F.; Kantor, P. B.; and Rokach, L. 2011. Recommender systems handbook.

ValuesParenting.com. 2015. Joy school. In *Retrieved from http://www.valuesparenting.com/joyschool/*.

VP. 2016. Joy school fall semester year 1. http://www.valuesparenting.com/members/joyschool/year1/joyschool_fall_semester.phtml.

Zhang, B.-W.; Yin, X.-C.; Cui, X.-P.; Qu, J.; Geng, B.; Zhou, F.; Song, L.; and Hao, H.-W. 2014. Social book search reranking with generalized content-based filtering. In *ACM CIKM*, 361–370. ACM.