How Can We Help Our K-12 Teachers?: Using a Recommender to Make Personalized Book Suggestions

Maria Soledad Pera and Yiu-Kai Ng

Computer Science Department, Brigham Young University, Provo, Utah, 84602, U.S.A.

Emails: mpera@cs.byu.edu, ng@compsci.byu.edu

Abstract—K-12 teachers, especially the ones who teach reading and literacy, are expected to guide their students to read in order to learn. Teachers can promote good reading habits among K-12 readers by offering books that match their interests. Unfortunately, finding the right book for each individual or group of students is not an easy task due to the huge volume of books available these days that cover a diversity of topics at varied reading levels. To address this problem, we have developed BRéT, a book recommender for K-12 teachers. BRéT adopts a multi-dimensional strategy to suggest books that simultaneously match the interests, preferences, and reading abilities of K-12 students based on the content, topics, literary elements, and grade levels specified by a teacher. BRéT is novel, since it recommends books to K-12 teachers tailored to their individual students or groups of students, either for pleasure reading or fulfilling their current instructional activities. Unlike existing book-searching tools currently being used by teachers, which adopt a “one-size-fits-all” strategy, BRéT offers personalized suggestions. Conducted empirical studies using Mechanical Turk have verified the effectiveness of BRéT in making book recommendations.

Keywords—recommendation; books; K-12; teachers

I. INTRODUCTION

Reading, which is an important skill for everyone to develop, is the driving force behind a productive, successful life. Promoting good reading habits among K-12 students is essential, given the enormous influence of reading on students’ development as learners and members of the society. While parents are expected to establish their children’s foundation to becoming lifelong readers as learning to read begins “long before a child enters school,” teachers are responsible for guiding K-12 readers through their formative years. It is anticipated that K-12 teachers assist their students in achieving their reading goals by exposing them to a volume and range of motivating texts. Among the diversity of reading materials targeting K-12 readers, which include scholastic magazines and blogs, US children/teenagers spend more time reading books than other materials, a trend that is also observed worldwide. However, it is a real challenge for teachers to find relevant books to suggest to each of their (groups of) students, since it is unrealistic to expect a teacher to be cognizant of the huge amount of books, which are being published on a regular basis, to identify the ones for suggestions. Moreover, they are not available around the clock and cannot serve more than one student at any given instant in time.

A recommendation system that provides teachers with books of interest to their students is an ideal solution to the problems mentioned above. Unfortunately, as stated by Prieto et al. [15], majority of available recommendation strategies and techniques are centered in learners, e.g., readers, rather than assisting teachers in their search for students’ reading materials. Moreover, existing book recommenders [8], [14], [18] require user-defined information, such as tags, ratings, connections, and accessing patterns, to make suggestions for the respective individuals. Personal information of K-12 users, however, may not be publicly accessible due to the ethical obligation of everyone to respect the online privacy of children/teenagers. Furthermore, these recommenders do not explicitly consider the reading ability of a reader, which is a necessity when making recommendations for students who are at different reading levels. Due to their design deficiencies and constraints, existing recommenders fail to assist educators in their quest for books relevant to their K-12 students. We have developed BRéT, an unsupervised book recommender for K-12 teachers, which facilitates one of the tasks undertaken by teachers on a regular basis — identifying books that match the interests of their individual/groups of students, either for pleasure reading or as supplement to their current lesson plans.

To make book recommendations, BRéT only requests a K-12 teacher to provide a brief description of the contents, topics, and/or literary elements of interest to, besides the range of grade levels of, the students for whom books are recommended. While the contents, topics, and literary elements ensure that books suggested by BRéT adequately address (to a degree) the information needs of the teacher, which in turn capture the interests/preferences of his/her students, the

1http://www.ksl.com/?sid=15431484
2K-12, which is a term used in the educational system in the United States and Canada (among other countries), refers to the primary and secondary/high school years of public/private school grades prior to college. These grades are kindergarten (K) through 12th grades.
3http://www.literacymidsouth.org/resources/literacy-statistics/
4http://goo.gl/SYMFeR
5http://goo.gl/DdakGR
grade levels ensure that books recommended by BReT can be comprehended by students to avoid suggesting books that are either too easy/difficult to read which could diminish the students’ interest in reading [1]. BReT simulates the readers’ advisory (RA) service [17] offered by knowledgeable staff at school/public libraries which provides reading materials that match readers’ needs, interests, and reading levels [17]. The mission of BReT is the same as RA, which offers a “vital link” between reading materials and readers [17], i.e., students in our case, through librarians (i.e., teachers in our case). By automating the RA process using a multi-dimensional recommendation strategy, BReT suggests to K-12 teachers books tailored to their individual students or groups of students.

BReT is designed for solving the information overload problem while minimizing the time and efforts imposed on K-12 teachers in discovering unknown books of interest to their K-12 readers. It can handle any number of teachers’ requests anytime and anywhere simultaneously, which cannot be achieved by traditional RA. Given the clear correlation between the academic performance of students and their reading ability, BReT plays a significant role in recommending suitable reading materials for educators. While BReT has been developed with K-12 teachers in mind, it can also be used by parents to locate books appealing to their children/teenagers.

The current implementation of BReT tailors towards recommendation of books written in English and classification of their readability levels based on the K-12 grade level system. The design of BReT, however, can be adopted to make suggestions based on diverse grade-level scales and in languages other than English.

In the remaining of this paper, we first discuss existing approaches on book recommendation and currently widely-used technologies/tools employed by teachers to locate potential books of interests to their students. Thereafter, we present the design methodology of BReT and the results of the experiments conducted to validate its correctness. Lastly, we provide a concluding remark and directions for future work.

II. RELATED WORK

To the best of our knowledge, there is no existing book recommendation system developed specifically to help teachers in their quest for reading materials that meet the needs of their students. Instead, teachers rely on resources available at a number of existing websites, including, but not limited to, ARbookfind.com, Kidsread.com, Scholastic.com, and WorldCat.org. These popular sites offer teachers different tools to search for books in various domains to be suggested to their students. The sites, however, either (i) supply (read-alike) non-personalized booklists, (ii) require teachers to select a topic/subject area of interest from a predefined list, which limits the themes of books that can be obtained from the sites, (iii) offer reading choices grouped by age/grade ranges, which is a constraint, since students in the same grade or age group might not reach the same reading level, or (iv) allow teachers to create keyword-queries to specify their information needs that quite often yield an overwhelming volume of items to choose from, and thus imposes an additional burden on the teachers to sort though. Unlike resources made available through the aforementioned websites, BReT minimizes the time and efforts imposed on teachers in locating books relevant to the information needs of their individual students and at a reading level that the students can read and understand.

Even though book recommenders solely developed for teachers do not exist, a number of book recommendation systems that have been designed for general audience are available. The recommendation module offered by Amazon.com suggests books based on the purchase patterns of its users [11], whereas Yang et al. [20] analyze users’ access logs to infer their preferences and apply the traditional collaborative-filtering (CF) strategy to make book recommendations. Givon and Lavrenko [8] combine CF and social tags to capture the content of books for recommendation. Sieg et al. [18] rely on the standard user-based CF framework and incorporate semantic knowledge in the form of a domain ontology to capture the topics of interest to a user. PReF [14] examines users’ connections as part of the recommendation process, which may not be accessible when they involve K-12 readers due to privacy imposed on children. BReK12 [14], which is based on content and readability analysis, considers reading patterns of users and thus relies heavily on the availability of bookmarking information offered by social bookmarking site users.

The aforementioned recommenders are tailored to individual, direct users, as opposed to teachers, and require (i) historical data on the users in the form of ratings, which may not always be available, or (ii) an ontology, which can be labor-intensive and time-consuming to construct. In addition, neither of these recommenders (with the exception of BReK12) considers the readability level of their users as part of their recommendation strategies.

It is worth mentioning that even though BReT is not a recommendation system for direct learning, its design goal is to enhance reading selections for K-12 users by locating suitable books among the overwhelming number of choices available these days. BReT does offer “a promising approach to facilitate both learning and teaching tasks” [19]. (An in-depth description of existing recommenders in the educational domain can be found in [12].)

III. OUR PROPOSED RECOMMENDER

In an attempt to locate suitable books for a (group of) student(s) $S$, regardless whether they are for pleasure reading or to fulfill a current instructional activity, a teacher $T$ provides BReT with a number of specifications: (i) a
range of grade levels, (ii) a brief content description, (iii) a topical description, and (iv) the literary elements of the desired books. Thereafter, BReT examines what types of book contents are favored by S, which topics S prefer(s), and why a book appeals to S. The analysis conducted by BReT yields a number of books that are recommended to T. As previously stated, Specification (i) ensures that books recommended by BReT can be comprehended by S, whereas Specifications (ii) to (iv) guarantee that books suggested by BReT adequately address (to a degree) the interests/preferences of S. BReT serves as an aid to teachers in their quest for books appealing to their K-12 students so that they are motivated to read. By emulating the RA process, BReT offers a unique approach for making book suggestions.

A. Selecting Candidate Books

BReT first determines the set of candidate books, denoted CBS, to be considered for recommendation with grade levels within the range defined by T to ensure that suggested books match the reading abilities of S, the student(s) of T. As “reading for understanding cannot take place unless the words in the quest for books are accurately and efficiently decoded” [13], BReT applies Equation 1 to determine the set of candidate books.

\[ CBS = \{ CB \mid CB \in Rep \land GL(CB) \in Range \} \]  

where CB is a candidate book at a book repository Rep,\(^6\) Range is the grade-level range deemed acceptable for S, and GL(CB) is the computed grade level of CB.

Popular readability prediction formulas/tools, such as Flesch-Kincaid, Lexile, and ATOS (discussed in detail in [2]), rely on sample text of a book to compute its grade level, which is a severe constraint, since sample text is not always freely accessible due to copyright law. BReT relies on TRoLL [14], a newly-developed readability analysis tool to determine the grade levels of books. TRoLL computes the readability level of any book on-the-fly, even in the absence of book excerpts, by using publicly accessible book metadata, such as audience levels, Library of Congress Subject Headings, and US curriculum subject areas. Experimental results presented in [14] show that TRoLL is not only highly accurate in predicting the grade levels of K-12 books, but it also outperforms other existing readability formulas/tools, such as Flesch-Kincaid and Accelerated Reader (AR).

B. Analyzing Candidate Recommendations

To determine the most promising books to be recommended, among the ones included in CBS, BReT analyzes the content, topical, and appeal-term descriptions of each candidate book CB in CBS.

1) Content Analysis: BReT analyzes the content description of CB, which is extracted from reputable book-related websites, such as Amazon.com and the Library of Congress,\(^7\) to determine the degree to which CB addresses written matters that are appealing to S based on the short description \(T_{Desc}\) provided by T. As shown in Equation 2, BReT computes the content similarity score of CB, i.e., \(CSim(CB)\), based on the “bag-of-words” representation of the description of CB \(T_{Desc, CB}\) respectively. \(CSim(CB)\) considers the word-correlation factor [9] \((wcf)\) of each word in \(T_{Desc}\) with respect to each word in the description of CB, and prioritizes candidate books with the highest degree of shared content. Word-correlation factors in the pre-computed word-correlation matrix reflect the degree of similarity between any two words according to their (i) frequencies of co-occurrence and (ii) relative distances in a collection of Wikipedia.org documents. BReT relies on word-correlation factors, as opposed to WordNet\(^8\)-based similarity measures [3], since it has been empirically verified that the former correlates with human assessments on word similarity more accurately than the latter.

\[ CSim(CB) = \frac{\sum_{i=1}^{n} \text{Min} \{ \sum_{j=1}^{m} \text{wcf}(T_{Desc, CB}, T_{Desc, CB_j}) \}}{n} \]  

where \(n \) (\(m \), respectively) is the number of distinct words in \(T_{Desc}\) (the description of CB, respectively), \(T_{Desc, CB}\) \(T_{Desc, CB_j}\), respectively) is a word in \(T_{Desc}\) (the description of CB, respectively), \(wcf(T_{Desc, CB}, T_{Desc, CB_j})\) is the correlation factor of \(T_{Desc, CB}\) and \(T_{Desc, CB_j}\) in the word-correlation matrix.

The \(\text{Min} \) function in Equation 2 imposes a constraint on summing up the word-correlation factors of words in the description of CB and \(T_{Desc}\). Even if a word in the description of \(T_{Desc}\) (i) matches exactly one of the words in CB (ii) is similar to some of the remaining words in CB, which yields a value greater than 1.0, i.e., the word-correlation factor of an exact match, BReT limits the sum of their similarity measure to 1.0. This constraint ensures that if \(T_{Desc}\) contains a dominant word \(w\) in its description which is highly similar to a few words in CB, \(w\) alone cannot dictate the content resemblance value of \(T_{Desc}\) with respect to CB. Words in the brief overview of CB that are similar to most of the words in \(T_{Desc}\) should yield a higher \(CSim(CB)\) than words in the description of CB that are similar to only one dominant word in \(T_{Desc}\).

2) Topical Analysis: BReT examines the topical description of CB based on Library of Congress Subject Headings (LCSH) assigned to CB by professional catalogers. LCSH, a de facto universal controlled vocabulary,

---

\(^6\)Book repositories include, but are not limited to, reputable websites, such as OpenLibrary.org and WorldCat.org, which are two of the largest online library catalogs, school/public library catalogs, and book-related bookmarking sites, such as BiblioNasium.com, which is a website that encourages reading among children/teenagers.

\(^7\)catalog.loc.gov

\(^8\)From now on, unless stated otherwise, “word” means non-stop, stemmed word.

\(^9\)wordnet.princeton.edu
constitutes the largest general indexing vocabulary in English [21]. A subject heading, which is a term or phrase that denotes a concept, event, or name [16], is used by librarians to categorize and index books according to their themes. (“Computers” and “Juvenile Fiction” are examples of LCSH.) BReT explores the resemblance between the topical dimension of CB and the topical specifications provided by T (i.e., Ttopics) using Equation 3. TSim(CB) determines the overlap between LCSH assigned to CB and the ones specified in Ttopics using an enhanced version of the Lennon similarity measure [10] such that the larger the overlap, the more likely CB addresses topics of interest for S.

\[
TSim(CB) = \frac{|S_\cap|}{\min(|S_{CB} - S_\cap|, |S_{T\text{topics}} - S_\cap|) + |S_\cap|} 
\]

(3)

where \(S_{CB}\) (\(S_{T\text{topics}}\), respectively) is the set of LCSH assigned to CB (specified by T, respectively), \(S_\cap = S_{CB} \cap S_{T\text{topics}}\), \(|S_\cap|\) is the number of LCSH assigned to both CB and Ttopics, and \(|S_{CB} - S_\cap|\) (\(|S_{T\text{topics}} - S_\cap|\), respectively) is the number of LCSH assigned to CB, but not to Ttopics (Ttopics, but not to CB, respectively).

In Equation 3, the minimum of the differences between \(|S_{CB} - S_\cap|\) and \(|S_{T\text{topics}} - S_\cap|\) is chosen, since by using the smaller of the two differences, we can more accurately capture the topical similarity between CB and Ttopics. As a difference reflects the number of LCSH assigned to CB, but not to Ttopics or vice versa, a smaller difference signifies that proportionally a larger number of topics are shared between CB and Ttopics, which is a more accurate indication of the degree of topical resemblance of CB with respect to Ttopics. Furthermore, in Equation 3, we relax the exact-matching constraint imposed for determining \(S_\cap\) such that LCSH of CB and Ttopics that are analogous\(^{10}\) to, but do not exactly match, each other are still included in \(S_\cap\). This enhanced version of the Lennon measure accounts for T using keywords outside of the vocabulary defined by the Library of Congress to describe the topics of interest to S.

3) Examining Appeal-Term Descriptions: BReT examines the literary elements, i.e., appeal factors, of CB to determine its appeal similarity with respect to the appeal factors of interest to S as specified by T, denoted TAFactors.

Appeal factors are “the elements of a book—whether definable or just understood—that make readers enjoy the book” [17]. These factors capture general traits of a book that attract the attention of a reader and are analyzed to determine the reasons why the reader is interested in a given book. For example, some readers might enjoy the Harry Potter books (by J.K. Rowling) because of the friendships among students and the boarding school setting, whereas others like the fantasy aspect of the story. The most prominent appeal factors in RA-related literature include: (i) characterization, (ii) frame, (iii) pacing, (iv) storyline, (v) language and writing style, (vi) tone, and (vii) special topics [17]. The first six appeal factors are well-known literary elements of (non-fiction books [4], whereas the last one identifies subjects addressed in a book that can cause emotional stress to some readers but can be tolerated/enjoyed by others [5]. Each appeal factor is associated with a vocabulary, which is a set of keywords, called appeal terms, employed to describe the factor. These appeal terms are defined based on well-known RA literature [5], [17].

In calculating AFSim(CB), the appeal-factor similarity score of CB with respect to TAFactors, BReT adopts the cosine measure as defined in Equation 4 and relies on the availability of the description of the appeal factors that apply to CB. Obtaining the appeal-term description of CB, however, is not a trivial task, given that, at present, appeal-term descriptions of books are only available through RA databases or determined by professionals. Unfortunately, accessing reputable RA databases, such as NovelList, requires a paid subscription, whereas professionals might not have read a particular book and thus cannot provide on-the-fly the corresponding appeal-term description for the book. BReT turns to ABET [14] to obtain the appeal-term description of each book in CBS.

ABET is a newly-developed, unsupervised tool that automatically extracts an appeal-term description of CB by analyzing (up to) 500 distinct reviews on CB, which can be retrieved from well-known book-related websites, such as Amazon.com. By analyzing reviews, ABET extracts diverse readers’ opinions on a book based on appeal terms that describe the corresponding appeal factors of the book. ABET, which performs linguistic and semantic analysis on sentences in reviews using Stanford Part-of-Speech Tagger and Dependency Parser (nlp.stanford.edu/software/lex-parser.shtml), employs a number of extraction rules on word pairs in sentences included in reviews that capture the semantic link between appeal terms and their corresponding appeal factors (based on typed dependency relations).\(^{11}\) These rules simply look for words in sentences that (directly or indirectly) describe the appeal factors of a book, which are often the subjects or objects of sentences.

The ABET-generated appeal-term description of CB includes not only the appeal terms extracted from reviews on CB, but also their frequency of occurrence. The latter captures the relative degree of significance of an appeal term in describing its corresponding factor based on reviewers’ varied opinions on appeal factors that apply to CB. (A sample appeal-term description for the book “The Hunger Games” by Suzanne Collins is shown in Figure 1.)

---

\(^{10}\)Two topical words are treated as analogous if their correlation factor is included in a reduced version of the aforementioned word-correlation matrix which contains 13% of the most frequently-occurring words in the collection of Wikipedia documents.

\(^{11}\)It is natural for ABET to turn to typed dependencies, since they capture the semantic connection, i.e., association, between words in sentences.
Figure 1: ABET-generated appeal-term description for “The Hunger Games” where the number indicates the frequency in which a term was used to describe its corresponding appeal factor in book reviews

\[ AFSim(CB) = \sum_{af \in AF} \frac{C_{af} \cdot T_{AFactors}^c}{|C_{af}| \cdot |T_{AFactors}^c|} \]  

(4)

where \( AF \) is the set of appeal factors in the appeal-term description for \( CB \) and \( T_{AFactors} \), \( |AF| \) is the size of \( AF \), \( C_{af} \) and \( T_{AFactors}^c \) are the \( n \)-dimensional vector representation of the appeal-term distribution of an appeal factor \( af \) for \( CB \) and \( T_{AFactors} \), respectively, \( n \) is the number of distinct appeal terms in the distribution of the corresponding appeal factor for \( CB \) and \( T_{AFactors} \), \( W_{CB_{af}} \) is the weight of the \( i \)th term in \( C_{af} \), and \( W_{T_{AFactors}^c} \) is “1” if \( T \) employs the term to describe the corresponding factor, and “0” otherwise, which is the weight of the \( i \)th term in \( T_{AFactors}^c \).

C. Rank Aggregation

Having determined the content, topical, and appeal-term similarity scores of \( CB \) with respect to the specifications provided by \( T \), BReT computes its overall ranking score by applying CombMNZ [6], a popular linear combination measure, on \( CSim(CB), TSim(CB), \) and \( AFSim(CB) \).

\[ \text{Rank}(CB) = \sum_{c=1}^{3} \text{score}^c \times |\text{score}^c > 0| \]  

(5)

where \( \text{score}^c \) is the value of one of the scores computed in Sections III-B1 - III-B3 and \( |\text{score}^c > 0| \) is the number of non-zero scores of \( CB \). CombMNZ combines multiple existing lists of rankings on an item into a joint ranking, a task known as rank aggregation or data fusion. The rank aggregation strategy adopted by BReT accounts for the fact that not all books in \( CBS \) are assigned a non-zero score for each of the measures computed in Sections III-B1 to III-B3, i.e., its content, topical, and appeal factors degree of resemblance with respect to the specifications provided by \( T \). The ranking considers the strength of each evidence regardless of their actual values, i.e., \( CSim, TSim, \) and \( AFSim \), as opposed to simply positioning higher in the ranking of candidate books with non-zero scores for all the measures. After a ranking score has been computed for each book in \( CBS \), the top-3 highest-ranked books are suggested to \( T \).

Example 1: Consider the specifications provided by a teacher \( T \) and submitted to BReT as shown in Figure 2. Also consider the grade level, content, topic, and appeal term descriptions indicated in the top-2 recommendations made by BReT in response to the specifications as shown in Figure 3. We have manually verified the effectiveness of the multi-dimensional analysis adopted by BReT in generating book suggestions, i.e., “Planting a Rainbow” by Lois Ehlert and “Koala and the Flower” by Mary Murphy, which match the content, topical, appeal factors, and grade levels specified by \( T \). (See the words that appear in bold in Figure 3 which either exactly-match or are analogous to the ones specified by \( T \) as shown in Figure 2.)

IV. EXPERIMENTAL RESULTS

In this section, we first introduce our evaluation framework. Thereafter, we present the results of the empirical studies conducted to assess the performance of BReT.
A. Evaluation Framework

Due to the lack of benchmark datasets for assessing the performance of book recommenders designed with teachers in mind, we conducted three empirical studies (presented in Sections IV-B to IV-D) using Amazon’s Mechanical Turk, a “marketplace for work that requires human intelligence”, which allows individuals or businesses to programmatically access thousands of diverse, on-demand workers and has been used in the past to collect user feedback on various information retrieval designs.

The current implementation of BReT relies on close to 20,000 books available at BiblioNasium.com as its book repository. We turn to BiblioNasium, since it is a bookmarking site which archives books tailored towards K-12 readers and was set up exclusively to encourage children and teenagers to read. Note that besides BiblioNasium, any other book repository, such as OpenLibrary.org, can also be employed by BReT to make recommendations.

B. Matching Book Specifications

We conducted a survey on Mechanical Turk, in which we asked appraisers to identify, among a provided set of three books, the one that did not capture a pre-defined set of specified constraints $R$. Two of the books included in each survey are the top-2 suggestions recommended by BReT in response to $R$, whereas the remaining one was randomly selected from the BiblioNasium repository. The survey was conducted to quantify the degree to which books suggested by BReT meet the specifications in $R$. The lower the number of appraisers who have chosen BReT-recommended books is, the more useful BReT is in terms of providing book suggestions that adequately satisfy teachers’ specifications. We created five HITs (Human Intelligent Task) on Mechanical Turk, each one with a different set of specifications, and collected 75 responses during the month of January 2014. Based on the feedback collected through Mechanical Turk, we have observed that 98% of the independent appraisers did not select BReT-recommended books as the ones that do not match the given specifications in $R$. The lower number of appraisers who have chosen BReT-recommended books is, the more useful BReT is in terms of providing book suggestions that adequately satisfy teachers’ specifications.

C. Correctness of the Generated Recommendations

To further evaluate BReT, we created another five different book specifications which simulate the requirements provided by teachers. These specifications address the contents, topics, literary elements, and grade levels of books of interest. For each of the specifications $R$, we set up a HIT. Each HIT required appraisers to select, among the three books included in the HIT, the ones, if any, that satisfy the specifications in $R$. Books included in a HIT are the top-3 suggestions made by BReT in response to $R$.

We collected responses to the HITs from 117 independent appraisers during the month of January 2014. The responses provided by each appraiser are treated as the “gold standard”, i.e., the chosen books are treated as relevant to the specifications in the corresponding HIT. Based on the responses, we evaluated the performance of BReT using two metrics, Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (nDCG) [6]. While MRR computes the average ranking position of the first relevant book suggested by a recommendation system, nDCG determines the overall ranking performance of the recommender and penalizes relevant books ranked lower in the recommendation list. The penalization is based on a reduction, which is logarithmically applied to the position of each relevant book in a ranked list.

The nDCG and MRR scores computed according to the collected responses are shown in Figure 5. Among the appraisers who provided their occupation, 41% were teachers, English instructors, or librarians. Given that the goal of BReT is to assist teachers in locating books pertaining to their specific requests, it is appropriate to quantify the performance of BReT reflected by the opinions of educators/librarians who are professionally trained to analyze books based on their contents, topics, literary elements,
and/or grade levels. As shown in Figure 5, the nDCG and MRR scores calculated according to teachers/English instructors/librarians’ responses yield a statistically significant improvement ($p < 0.05$) over the ones based on all the collected responses, which is determined using the Wilcoxon signed-ranked test. The results compiled using the opinion of experts in the education domain are of special importance in assessing the performance of BReT, given the lack of benchmark datasets to evaluate recommendation tools for teachers. The fact that appraisers who are experts appreciate the recommendations made by BReT even more than the general appraisers provides further evidence of the usefulness of BReT for the educational community.

D. Human Assessment on BReT

We conducted another performance evaluation of BReT to verify whether its recommendations are perceived as accurate by independent appraisers who initiated the recommendation process by specifying their own requirements instead of being given a set of specifications to work with. In accomplishing this task, we implemented BReT as an online application and created another HIT on Mechanical Turk which required appraisers to interact with BReT and answer a set of five pre-defined questions. These questions were set up to inquire the suitability of the BReT-generated recommendations in response to the content/topic/literary elements/grade levels specifications. (See Figure 6 for an example of the survey Mechanical Turk appraisers were asked to complete.)

We summarize the results (as shown in Figure 7) of the tens of responses collected between January 28 and January 31, 2014 as follows:

- An overwhelming number of appraisers who were involved in the study stated that the grade levels of the recommended books matched the range they have defined. This verifies that BReT is capable of suggesting books that conform to the reading level of the targeted audience.
- More than 80% of the appraisers indicated that the literary elements of books suggested by BReT closely matched the ones specified. This shows the accuracy of ABET (BReT, respectively) in generating appeal-term descriptions on the literary elements of books (in satisfying the literary elements requirements provided by the appraisers, respectively).
- Appraisers determined that BReT-generated recommendations generally matched the contents/topics submitted to BReT. This has been validated by the fact that eighty percent of independent appraisers picked 4 and 5 (out of 5) to quantify the degree to which the contents and topics of the BReT-suggested books satisfied their specifications (with 5 being “completely satisfied”).
- Close to 80% of independent appraisers found that at least two (out of the three) suggestions made by BReT were relevant to their specifications on requested books, which correlates with the results shown in Section IV-C.
V. CONCLUSION

We have introduced BReT, a book recommender for K-12 teachers. Using BReT, a teacher can specify the contents, topical information, literary elements, and grade levels of books of interests to his/her students. Hereafter, BReT suggests books that simultaneously match the preferences and reading abilities of the teacher’s individual or groups of students. Unlike existing search tools/online resources teachers currently rely on, which (i) provide non-personalized booklists grouped by either topics of interests or grade/age levels or (ii) retrieve an overwhelming number of reading choices for teachers to sort through, BReT suggests only a short list of books ranked according to their degree of relevance to a teacher’s request. Furthermore, BReT does not rely on large historical data on its users to make suggestions as existing state-of-the art book recommendation methodologies do.

The design goal of BReT is to assist K-12 teachers in their quest for books for their students, either for pleasure reading or knowledge acquisition, which can enrich the students’ choices on books and thus their reading experience to become lifelong readers. We have conducted empirical studies using a crowdsourcing platform which have verified not only the effectiveness of the design methodology of BReT, but also the correctness of its book suggestions.

As BReT is designed to help teachers find the “right material” for the “right audience,” for future work we plan to conduct a user-study to quantify the degree to which BReT can help teachers in their quest for materials that encourage reading habits among their K-12 students. We will follow the evaluation framework presented in [12], which is widely accepted for assessing the performance of recommendation systems within the domain of technology-enhanced learning. Furthermore, we would like to extend BReT to consider reading materials other than books, by (i) examining metadata on texts, such as HTML tags of web pages, age suitability of texts, and topical suitability of blogs, which goes beyond the traditional contents/topics/literary elements, and (ii) extending the functionality of the readability-prediction tool presented in Section III so that it can estimate the grade level of any reading material.

REFERENCES